On Notions of Distortion and an Almost Minimum Spanning Tree with Constant Average Distortion[*]

Yair Bartal [†]   Arnold Filtser [‡]   Ofer Neiman [§]

Abstract

This paper makes two main contributions: The first is the construction of a *near-minimum spanning tree* with *constant average distortion*. The second is a general equivalence theorem relating two refined notions of distortion: *scaling distortion* and *prioritized distortion*.

Minimum Spanning Trees of weighted graphs are fundamental objects in numerous applications. In particular in distributed networks, the minimum spanning tree of the network is often used to route messages between network nodes. Unfortunately, while being most efficient in the total cost of connecting all nodes, minimum spanning trees fail miserably in the desired property of approximately preserving distances between pairs. While known lower bounds exclude the possibility of the worst case distortion of a tree being small, it was shown in [ABN15] that there exists a spanning tree with constant average distortion. Yet, the weight of such a tree may be significantly larger than that of the MST. In this paper, we show that any weighted undirected graph admits a *spanning tree* whose weight is at most $(1 + \rho)$ times that of the MST, providing *constant average distortion $O(1/\rho)$*. Our result exhibits the best possible tradeoff of this type.

This result makes use of a *general equivalence theorem* relating two recently developed notions of distortion for metric embedding. The first is the notion of *scaling distortion*, which provides improved distortion for $1 - \epsilon$ fractions of the pairs, for all ϵ simultaneously. A stronger version called *coarse* scaling distortion, has improved distortion guarantees for the furthest pairs. The second notion is that of *prioritized distortion*, a property allowing to prioritize the nodes whose associated distortions will be improved. We show that prioritized distortion is essentially equivalent to coarse scaling distortion via a general transformation.

Our spanning tree result is achieved by first showing the existence of a low weight *spanner* with small prioritized distortion, which by our theorem implies scaling distortion, which in turn implies the constant average distortion bound.

The scaling-prioritized distortion equivalence theorem has further implications and is of independent interest. In particular, it allows us to resolve a few questions concerning prioritized embedding. Mainly, we obtain a strengthening of Bourgain's theorem on embedding arbitrary metrics into Euclidean space, possessing optimal prioritized distortion.

---

1 Introduction

One of the fundamental problems in graph theory is that of constructing a Minimum Spanning Tree (MST) of a given weighted graph $G = (V, E)$. This problem and its variants received much attention, and has found numerous applications. In many of these applications, one may desire not only minimizing the weight of the spanning tree, but also other desirable properties, at the price of losing a small factor in the weight of the tree compared to that of the MST. Define the *lightness* of T to be the total weight of T (the sum of its edge weights) divided by the weight of an MST. One well known example is that of a Shallow Light Tree (SLT) [KRY93, ABP92], which is a rooted spanning tree having near optimal $(1 + \rho)$ lightness, while approximately preserving all distances from the root to the other vertices.

It is natural to ask that the spanning tree will preserve well all pairwise distances in the graph. However, it is easy to see that no spanning tree can maintain such a requirement. In particular, even in the case of the unweighted cycle graph on n vertices, for every spanning tree there is a pair of neighboring vertices whose distance increases by a factor of $n - 1$. A natural relaxation of this demand is that the spanning tree approximates all pairwise distances *on average*. Formally, the distortion of the pair $u, v \in V$ in T is defined as $\frac{d_T(u,v)}{d_G(u,v)}$, and the *average distortion* is $\frac{1}{\binom{n}{2}} \sum_{\{u,v\} \in \binom{V}{2}} \frac{d_T(u,v)}{d_G(u,v)}$, where d_G (respectively d_T) is the shortest-path metric in G (resp. T).[1] There are graphs where the average distortion of the MST can be as large as $\Omega(n)$ (see for example the graph described in Lemma 13). In [ABN15], it was shown that for every weighted graph, it is possible to find a spanning tree which has constant average distortion. However, the question of whether it is possible to obtain a tree with non-trivial bounds on both its lightness and the average distortion remained open.

In this paper we resolve this question in the affirmative. Specifically, we devise a spanning tree of optimal $(1 + \rho)$ lightness that has $O(1/\rho)$ average distortion over all pairwise distances. We show that this result is tight by exhibiting a lower bound on the tradeoff between lightness and average distortion, that in order to get $1 + \rho$ lightness the average distortion must be $\Omega(1/\rho)$ (this holds for $1/n \leq \rho \leq 1$, and even if the spanning subgraph is not necessarily a tree).

Our main result of a light spanning tree with constant average distortion may be of interest for network applications. It is extremely common in the area of distributed computing that an MST is used for communication between the network nodes. This allows easy centralization of computing processes and an efficient way of broadcasting through the network, allowing communication to all nodes at a minimum cost. Yet, as already mentioned above, when communication is required between specific pairs of nodes, the cost of routing through the MST may be extremely high, even when their real distance is small. However, in practice it is the average distortion, rather than the worst-case distortion, that is often used as a practical measure of quality, as has been a major motivation behind the initial work of [KSW09, ABN11, ABN15]. As noted above, the MST still fails even in this relaxed measure. Our result overcomes this by promising small routing cost between nodes on average, while still possessing the low cost of broadcasting through the tree, thereby maintaining the standard advantages of the MST.

Our main result on a low average distortion embedding follows from analyzing the *scaling distortion* [KSW09, ABN11] of the embedding. This refined notion of distortion turns out to be closely related to another useful measure of *prioritized distortion* [EFN18]. The second main contribution of this paper is providing an equivalence theorem stating the relation between these useful notions.

---

[1]Distortion is sometimes referred to as stretch.

1.1 Scaling Distortion vs. Prioritized Distortion: A General Equivalence Theorem

Scaling distortion, first introduced in [KSW09][2], requires that for *every* $0 < \epsilon < 1$, the distortion of all but an ϵ-fraction of the pairs is bounded by the appropriate function of ϵ. In [ABN11] it was shown that one may obtain bounds on the average distortion, as well as on higher moments of the distortion function, from bounds on the scaling distortion. In [ABN11] several scaling distortion results were shown including $O(\log(1/\epsilon))$ scaling distortion embedding into Euclidean space, and in [ABN15] an $O(1/\sqrt{\epsilon})$ scaling embedding into trees, and spanning trees in particular.

Prioritized distortion, introduced recently in [EFN18], requires that for every given ranking $v_1, \ldots, v_n$ of the vertices of the graph, there is an embedding where the distortion of pairs including v_j is bounded as a function $\alpha(j) : [n] \to \mathbb{R}_+$ of the rank j. Several prioritized distortion results were given in [EFN18], including $\tilde{O}(\log j)$ [3] prioritized distortion embedding into Euclidean space. (See Section 2 for formal definitions of embeddings and various notions of distortion).

One of the main ingredients of our work is a *general reduction* relating the notions of prioritized distortion and scaling distortion. In fact, we show that prioritized distortion is essentially equivalent to a strong version of scaling distortion called *coarse* scaling distortion, in which for every point p and every $0 < \epsilon < 1$, the distances to the $1 - \epsilon$ fraction of the farthest points from p are preserved with the desired distortion. We prove that any embedding with a prioritized distortion α has coarse scaling distortion bounded by $O(\alpha(8/\epsilon))$. This result could be of independent interest; in particular, it shows that the results of [EFN18] have their scaling distortion counterparts (some of which were not known before). We further show a reduction in the opposite direction, informally, that given an embedding with coarse scaling distortion γ, there exists an embedding with prioritized distortion $\gamma(\mu(j))$, where μ is a function such that $\sum_i \mu(i) = 1$ (e.g. $\mu(j) = \frac{6}{(\pi \cdot j)^2}$). We note that this reduction heavily relied on the property of coarse scaling distortion embeddings and does not apply to non-coarse scaling embeddings. Yet, most existing scaling embeddings are indeed coarse. This result implies that all existing *coarse* scaling distortion results have priority distortion counterparts, thus improving few of the results of [EFN18]. In particular, by applying a theorem of [ABN11] we obtain prioritized embedding of arbitrary metric spaces into l_p in dimension $O(\log n)$ and prioritized distortion $O(\log j)$, which exhibits a strengthening of Bourgain's theorem [Bou85] (which asserts $O(\log n)$ worst-case distortion), and is best possible. It also implies better bounds for decomposable metrics (see [ABN11]), such as planar and doubling metrics, where we obtain an optimal $O(\sqrt{\log j})$ prioritized distortion.

We also show an equivalence between embeddings with coarse partial distortion and terminal embeddings, which can be used to extend and improve previous results. See Section 3.3 for details.

In the context of our main construction of a light spanning tree, the first direction of the above equivalence theorem allows us to devise prioritized distortion embeddings and use these to obtain scaling distortion embeddings which possess the desired constant average distortion.

1.2 Light Spanning tree of Constant Average Distortion.

Our main spanning tree construction provides a spanning tree with $1 + \rho$ lightness and scaling distortion $\tilde{O}(1/\sqrt{\epsilon})/\rho$, which is nearly tight as a function of ϵ [ABN15]. This result implies that the average distortion is $O(1/\rho)$.

We also devise a probabilistic embedding: a distribution over spanning trees, each tree in the support of the distribution has $1 + \rho$ lightness, and the expected scaling distortion is at most $\mathrm{polylog}(1/\epsilon)/\rho$. Thus

---

[2]Originally coined gracefully degrading embedding.

[3]By $\tilde{O}(f(n))$ we mean $O(f(n) \cdot \mathrm{polylog}(f(n)))$.

providing constant bounds on all fixed moments of the distortion (i.e., the l_q-distortion [ABN11] for fixed q). (See the end of Section 5 for a formal definition of probabilistic embedding.)

Our main technical contribution, en route to this result, may be of its own interest: We devise a *spanner* (a subgraph of G) with $1 + \rho$ lightness and low *prioritized distortion*. Here we show a light spanner construction with prioritized distortion at most $\tilde{O}(\log j)/\rho$. Using the equivalence theorem relating prioritized distortion and scaling distortion (discussed above), we obtain a spanner having scaling distortion $\tilde{O}(\log(1/\epsilon))/\rho$, and thus average distortion $O(1/\rho)$. Although we do not obtain a spanning tree here, this result has a few advantages, as we get constant bounds on all fixed moments of the distortion function (the ℓ_q-distortion). Moreover, the worst-case distortion is only logarithmic in n. We note that all of our results admit deterministic polynomial time algorithms.

Another technical contribution is a general, black-box reduction, that transform constructions of spanners with distortion t and lightness ℓ into spanners with distortion t/δ and lightness $1 + \delta\ell$ (here $0 < \delta < 1$). This reduction can be applied in numerous settings, and also for many different special families of graphs. In particular, this reduction allows us to construct prioritized spanners with lightness arbitrarily close to 1.

Outline and Techniques.   Our proof has the following high level approach; Given a graph and a ranking of its vertices, we first find a low weight spanner with prioritized distortion $\tilde{O}(\log j)/\rho$. We then apply the general reduction from prioritized distortion to scaling distortion to find a spanner with scaling distortion $\tilde{O}(\log(1/\epsilon))/\rho$. Finally, we use the result of [ABN15] to find a spanning tree of this spanner with scaling distortion $O(1/\sqrt{\epsilon})$. We then conclude that the scaling distortion of the composed embeddings [4] is roughly their product, which implies our main result of a spanning tree with lightness $1 + \rho$ and scaling distortion $\tilde{O}(1/\sqrt{\epsilon})/\rho$.

Similarly, we can apply the probabilistic embedding of [ABN15] to get a light counterpart, devising a distribution over spanning trees, each with lightness $1+\rho$, with (expected) scaling distortion $\text{polylog}(1/\epsilon)/\rho$.

The main technical part of the paper is finding a light prioritized spanner. In a recent result [CW18] (following [ENS15, CDNS92]), it was shown that any graph on n vertices admits a spanner with (worst-case) distortion $O(\log n)$ and with constant lightness. However, these constructions have no bound on the more refined notions of distortion. To obtain a prioritized distortion, we use a technique similar in spirit to [EFN18]: group the vertices into $\log \log n$ sets according to their priority, the set K_i will contain vertices with priority up to 2^{2^i}. We then build a low weight spanner for each of these sets. As prioritized distortion guarantees a bound for *every pair* containing a high ranking vertex, we must augment the spanner of K_i with shortest paths to all other vertices. Such a shortest path tree may have large weight, so we use an idea from [CDG06] and apply an SLT rooted at K_i, which balances between the weight and the distortion from K_i.

The main issue with the construction described above is that the weight of the spanner in each phase can be proportional to that of the MST, but we have $\log \log n$ of those. Obtaining constant lightness, completely independent of n, requires a subtler argument. We use the fact that the weight of the light spanners in each phase comes "mostly" from the MST, and then some additional weight. We ensure that all the spanners will have *the same* MST. Then we select the parameters carefully, so that the additional weights will be small enough to form converging sequences, without affecting the distortion by too much.

---

[4]Given two embeddings $f : X \to Y$, $g : Y \to Z$, the composition $g \circ f$ is defined from X to Z, sending every point $x \in X$ to $g(f(x)) \in Z$.

1.3 Related Work

Partial and scaling embeddings[5] have been studied in several papers [KSW09, ABC$^+$05, ABN11, CDG06, ABN15, ABNS14]. Some of the notable results are embedding arbitrary metrics into a distribution over trees [ABC$^+$05] or into Euclidean space [ABN11] with tight $O(\log(1/\epsilon))$ scaling distortion. The notion ℓ_q-distortion was introduced in [ABN11], they show that their scaling distortion results imply constant average distortion and $O(q)$ bound on the ℓ_q-distortion. This notion has been further studied in several papers, including [ABN15, ABNS14, CDG06], and most recently applied in the context of dimensionality reduction [BFN17]. In [ABN15], an embedding into a single spanning tree with tight $O(1/\sqrt{\epsilon})$ scaling distortion was shown, which implies, in particular, constant average distortion, but there is no guarantee on the weight of the tree. It follows from [ABC$^+$05] that this bound is tight even when embedding into arbitrary (non-spanning) trees.

Prioritized distortion embeddings were studied in [EFN18], for instance they give an embedding of arbitrary metrics into a distribution over trees with expected prioritized distortion $O(\log j)$, and into Euclidean space with prioritized distortion $\tilde{O}(\log j)$.

Probabilistic embedding into trees [Bar96, Bar98, Bar04, FRT04] and spanning trees [AKPW95, EEST08, ABN08, AN12] has been intensively studied, and found numerous applications to approximation and online algorithms, and to fast linear system solvers. While our distortion guarantee does not match the best known worst-case bounds, which are $O(\log n)$ for arbitrary trees [Bar04, FRT04] and $\tilde{O}(\log n)$ for spanning trees [ABN08, AN12], we give the first probabilistic embeddings into spanning trees with polylogarithmic scaling distortion in which all the spanning trees in the support of the distribution are light.

The paper [CDG06] considers partial and scaling embedding into spanners, and show a general transformation from worst-case distortion to partial and scaling distortion. In particular, they show a spanner with $O(n)$ edges and $O(\log(1/\epsilon))$ scaling distortion. For a fixed $\epsilon > 0$, they also obtain a spanner with $O(n)$ edges, $O(\log(1/\epsilon))$ partial distortion and lightness $O(\log(1/\epsilon))$.[6] Note that these results fall short of achieving both constant average distortion and constant lightness.

In a subsequent work, [FS16] used our general reduction for light spanners (Theorem 8), to show that the $O(\log n/\rho)$-greedy spanner has lightness $1 + \rho$.

2 Preliminaries

All the graphs $G = (V, E, w)$ we consider are undirected and weighted with nonnegative weights. Throughout the paper we denote by n the number of vertices and by m the number of edges in the given graph. We shall assume w.l.o.g. that all edge weights are different. If it is not the case, then one can break ties in an arbitrary (but consistent) way. Note that under this assumption, the MST T of G is unique. The weight of a graph G is $w(G) = \sum_{e \in E} w(e)$. Let d_G be the shortest path metric on G. For a subset $K \subseteq V$ and $v \in V$ let $d_G(v, K) = \min_{u \in K}\{d_G(u, v)\}$. For $r \geq 0$ let $B_G(v, r) = \{u \in V : d_G(u, v) \leq r\}$ (we often omit the subscript when clear from context).

For a graph $G = (V, E)$ on n vertices, a subgraph $H = (V, E')$ where $E' \subseteq E$ (with the induced weights) is called a *spanner* of G. We say that a pair $u, v \in V$ has *distortion* at most t if

$$d_H(v, u) \leq t \cdot d_G(v, u) \, ,$$

---

[5]A partial embedding (introduced by [KSW09] under the name *embedding with slack*) requires that for a *fixed* $0 < \epsilon < 1$, the distortion of all but an ϵ-fraction of the pairs is bounded by the appropriate function of ϵ.

[6]The original paper claims lightness $O(\log^2(1/\epsilon))$, but their proof in fact gives the improved bound.

(note that always $d_G(v, u) \leq d_H(v, u)$). If every pair $u, v \in V$ has distortion at most t, we say that the spanner H has distortion t. Let T be the (unique) MST of G, the *lightness* of H is the ratio between the weight of H and the weight of the MST, that is $\Psi(H) = \frac{w(H)}{w(T)}$. We sometimes abuse notation and identify a spanner or a spanning tree with its set of edges.

A metric space (X, d_X) is defined over a set of points X and a nonnegative distance function d_X, with positive values on distinct points, and obeying the triangle inequality. Every weighted graph G can be viewed as a metric space (V, d_G). For two metric spaces (X, d_X), (Y, d_Y) and a non-contractive embedding $f : X \to Y$,[7] the distortion of a pair $x, y \in X$ under f is defined as $\frac{d_Y(f(x), f(y))}{d_X(x, y)}$.

When considering a graph G and its subgraph H, we may view the metric of G as being embedded into H via the identity map, in which case the last definition of distortion given above coincides with the those given earlier. Hence, the following definitions may be interpreted in the graph case in the obvious way.

Prioritized Distortion. Let (X, d_X), (Y, d_Y) be metric spaces. Let $\pi = v_1, \ldots, v_n$ be a priority ranking (an ordering) of the points (vertices) of X, and let $\alpha : \mathbb{N} \to \mathbb{R}_+$ be some monotone non-decreasing function. We say that a non-contractive embedding $f : X \to Y$ has *prioritized distortion* α (w.r.t π), if for all $1 \leq j < i \leq n$, the pair v_j, v_i has distortion (under f) at most $\alpha(j)$.

Scaling Distortion. Let (X, d_X), (Y, d_Y) be metric spaces, with $|X| = n$. For $v \in X$ and $\epsilon \in (0, 1)$ let $R(v, \epsilon) = \min \{r : |B(v, r)| \geq \epsilon n\}$. A point $u \in X$ is called ϵ-*far* from v if $d_X(u, v) \geq R(v, \epsilon)$. Given a function $\gamma : (0, 1) \to \mathbb{R}_+$, we say that a non-contractive embedding $f : X \to Y$ has *scaling distortion* γ, if for every $\epsilon \in (0, 1)$, there are at least $(1 - \epsilon)\binom{|X|}{2}$ pairs that have distortion at most $\gamma(\epsilon)$. We say that f has *coarse* scaling distortion γ, if every pair $v, u \in X$ such that both u, v are $\epsilon/2$-far from each other, has distortion at most $\gamma(\epsilon)$. [8]

Moments of Distortion. Let (X, d_X), (Y, d_Y) be metric spaces. For $1 \leq q \leq \infty$, define the ℓ_q-*distortion* of a non-contractive embedding $f : X \to Y$ as:

$$\text{dist}_q(f) = \mathbb{E}\left[\left(\frac{d_Y(f(u), f(v))}{d_G(u, v)}\right)^q\right]^{1/q},$$

where the expectation is taken according to the uniform distribution over $\binom{X}{2}$. The classic notion of *distortion* is expressed by the ℓ_∞-distortion and the *average distortion* is expressed by the ℓ_1-distortion. The following was proved in [ABN15].

Lemma 1. *([ABN15]) Let (X, d_X), (Y, d_Y) be metric spaces. If a non-contractive embedding $f : X \to Y$ has scaling distortion γ then*

$$\text{dist}_q(f) \leq \left(2 \int_{\frac{1}{2}\binom{n}{2}^{-1}}^{1} \gamma(x)^q dx\right)^{1/q}.$$

---

[7]An embedding f is non-contractive if for every $x, y \in X$, $d_Y(f(x), f(y)) \geq d_X(x, y)$.

[8]It can be verified that coarse scaling distortion γ implies scaling distortion γ. Moreover, in this paper we prove that the definition that requires only one of u, v to be $\frac{\epsilon}{2}$-far from the other, is almost equivalent. See Remark 1.

3 Prioritized Distortion vs. Coarse Scaling Distortion

In this section we study the relationship between the notions of prioritized and scaling distortion. We show that there is a reduction that allows to transform embeddings with prioritized distortion into embeddings with coarse scaling distortion, and vice versa.

3.1 Coarse Scaling Distortion implies Prioritized Distortion

The following theorem shows that coarse scaling distortion implies prioritized distortion, implying some new prioritized distortion embedding results, and in particular a prioritized version of Bourgain's theorem.

Theorem 1. *Let $\mu : \mathbb{N} \to \mathbb{R}^+$ be a non-increasing function such that $\sum_{i \geq 1} \mu(i) = 1$. Let $\mathcal{Y}$ be a family of finite metric spaces, and assume that for every finite metric space (Z, d_Z) there exists a non-contractive embedding $f_Z : Z \to Y_Z$, where $(Y_Z, d_{Y_Z}) \in \mathcal{Y}$, with (monotone non-increasing) coarse scaling distortion γ. Then, given a finite metric space (X, d_X) and a priority ranking $x_1, \ldots, x_n$ of the points of X, there exists an embedding $f : X \to Y$, for some $(Y, d_Y) \in \mathcal{Y}$, with prioritized distortion $\gamma(\mu(i))$.*

Proof. Given the metric space (X, d_X) and a priority ranking $x_1, \ldots, x_n$ of the points of X, let $\delta = \min_{i \neq j} d_X(x_i, x_j)/2$. We define a new metric space (Z, d_Z) as follows. For every $1 \leq i \leq n$, every point x_i is replaced by a set X_i of $|X_i| = \lceil \mu(i)n \rceil$ points, and let $Z = \bigcup_{i=1}^{n} X_i$. For every $u \in X_i$ and $v \in X_j$ define $d_Z(u, v) = d_X(x_i, x_j)$ when $i \neq j$, and $d_Z(u, v) = \delta$ otherwise. Observe that $|Z| = \sum_{i=1}^{n} |X_i| \leq \sum_{i=1}^{n} (\mu(i)n + 1) \leq 2n$.

We now use the embedding $f_Z : Z \to Y_Z$ with coarse scaling distortion γ, to define an embedding $f : X \to Y_Z$, by letting for every $1 \leq i \leq n$, $f(x_i) = f_Z(u_i)$ for some (arbitrary) point $u_i \in X_i$. By construction of Z, for every $j > i$, we have that $X_i \subseteq B(u_i, d_Z(u_i, u_j)) \cap B(u_j, d_Z(u_i, u_j))$. As $|X_i| \geq \mu(i)n \geq \frac{\mu(i)}{2}|Z|$, it holds that u_i, u_j are $\epsilon/2$-far from each other for $\epsilon = \mu(i)$. This implies that $\frac{d_{Y_Z}(f(x_i), f(x_j))}{d_X(x_i, x_j)} = \frac{d_{Y_Z}(f_Z(u_i), f_Z(u_j))}{d_Z(u_i, u_j)} \leq \gamma(\mu(i))$. $\qquad\square$

It follows from a result of [EFN18] that the convergence condition on μ in the above theorem is necessary (more details below). We note that this reduction can also be applied to cases where the coarse scaling embedding is only known for a class of metric spaces (rather than all metrics), as long as the transformation needed for the proof can be made so that the resulting new space is still in the class. This holds for most natural classes, such as metrics arising from trees, planar graph, graphs excluding a fixed minor, bounded degree graphs, doubling metrics, etc. [9]

3.1.1 Implications

The reduction implies that all existing *coarse* scaling distortion embeddings and distance oracles have priority distortion counterparts, thus improving few of the results of [EFN18].

Embeddings. By applying a theorem of [ABN11] we get the following.

---

[9]For the graph classes mention above (as well as for doubling metrics), a small change in the construction is needed. From each original vertex x_i, we will grow a path X_i of $\lceil \mu(i)n \rceil$ vertices, where all the path edges have weight $\frac{\delta \cdot \alpha}{2n}$ for arbitrarily small $\alpha > 0$. We will also need to choose u_i to be the single leaf in the added path (rather then simply arbitrarily chosen vertex). The same proof will guarantee a $(1 + \alpha)\gamma(\mu(i))$ prioritized distortion.

Corollary 2. *For every $1 \leq p \leq \infty$ and every finite metric space (X, d_X) and a priority ranking of X, there exists an embedding with prioritized distortion $O(\log j)$ into $l_p^{O(\log |X|)}$.*

Another consequence of the results of [ABN11] is better bounds for decomposable metrics[10]:

Corollary 3. *For every $1 \leq p \leq \infty$ and every finite τ-decomposable metric space (X, d_X) and a priority ranking of X, there exists an embedding with prioritized distortion $O(\tau^{1-1/p}(\log j)^{1/p})$ into $l_p^{O(\log^2 |X|)}$.*

Spanners. Applying Theorem 1 on [CDG06, Corollary 3] we get a linear size prioritized spanner.

Corollary 4. *Given a graph $G = (V, E)$ and any priority ranking $v_1, v_2, \ldots, v_n$ of V, there exists a spanner H with $O(n)$ edges and prioritized distortion $O(\log j)$.*

We remark that in Theorem 6 we show directly a spanner with $O(n \log \log n)$ edges and prioritized distortion $\tilde{O}(\log j)$ (which could easily be made $O(\log j)$). While not being of linear size, that spanner is very light. We currently do not know how to achieve both lightness and linear size spanner with prioritized distortion $O(\log j)$.

Distance Oracles. In [EFN18], among other possible tradeoffs, it was shown how to construct distance oracles with $O(n \log \log n)$ space and prioritized distortion $O(\log n / \log(n/j))$ with $O(1)$ query time. (Alternatively, they had $O(n \log^* n)$ space and $O(2^{\log n / \log(n/j)})$ prioritized stretch with $O(1)$ query time.) The space requirement in [EFN18] was never truly linear in n. Chechick [Che15] showed that for every metric space (X, d_X), one can construct a distance oracle with $O(\log n)$-stretch, $O(1)$-query time and $O(n)$ space. A black box reduction from [CDG06], will provide us with distance oracle with $O(\log \frac{1}{\epsilon})$ coarse scaling distortion, $O(1)$-query time and $O(n)$ space. We conclude with a linear size prioritized distance oracle.

Corollary 5. *For every metric space (X, d_x) and every priority ranking, there exist a distance oracle requiring $O(n)$ space, that answer distance queries in $O(1)$ time and $O(\log j)$ priority distortion.*

We remark that the prioritized distortion $O(\log n / \log(n/j))$ of [EFN18] is superior to our $O(\log j)$.

3.2   Prioritized Distortion implies Coarse Scaling Distortion

Here we prove the direction that is used for our main result of a light constant average distortion spanning tree, specifically, that prioritized distortion implies scaling distortion.

Theorem 2. *Let (X, d_X), (Y, d_Y) be metric spaces, then there exists a priority ranking $\pi = x_1, \ldots, x_n$ of the points of X such that the following holds: If there exists an embedding $f : X \to Y$ with (monotone non-decreasing) prioritized distortion α (with respect to π), then f has coarse scaling distortion $O(\alpha(8/\epsilon))$.*

The basic idea of the proof is to choose the priorities so that for every ϵ, every $v \in X$ has a representative v' of sufficiently high priority within distance $\approx R(v, \epsilon)$. Then for any $u \in X$ which is ϵ-far from v, we can use the low distortion guarantee of v' with both v and u via the triangle inequality. To this end, we employ the notion of a *density net* due to [CDG06], who showed that a greedy construction provides such a net.

Definition 1 (Density Net). *Given a metric space (X, d) and a parameter $0 < \epsilon < 1$, an ϵ-density-net is a set $N \subseteq X$ such that: **1)** for all $v \in X$ there exists $u \in N$ with $d(v, u) \leq 2R(v, \epsilon)$ and **2)** $|N| \leq \frac{1}{\epsilon}$.*

---

[10]Roughly, a metric space is called τ-decomposable if it allows probabilistic partitions with padding parameter τ; e.g. Planar metrics and doubling metrics. An exact definition appears in [ABN11].

Proof of Theorem 2. We begin by describing π, the desired priority ranking of X. For every integer $1 \leq i \leq \lceil \log n \rceil$ let $\epsilon_i = 2^{-i}$, and let $N_i \subseteq X$ be an ϵ_i-density-net in X. Set π to be a priority ranking of X satisfying that every point $v \in N_i$ has priority at most $\left| \bigcup_{j=1}^{i} N_j \right| \leq \sum_{j=1}^{i} |N_j|$. As for any j, $|N_j| \leq \frac{1}{\epsilon_j} = 2^j$, each point in N_i has priority at most $\sum_{j=1}^{i} \frac{1}{\epsilon_j} \leq \sum_{j=1}^{i} 2^j < 2^{i+1}$.

Let $f : X \to Y$ be some non-contractive embedding with priority distortion α with respect to π. Fix some $\epsilon \in (0, 1)$ and a pair $v, u \in V$ so that u is ϵ-far from v. Let i be the minimal integer such that $\epsilon_i \leq \epsilon$ (note that we may assume $1 \leq i \leq \lceil \log n \rceil$, because there is nothing to prove for $\epsilon < 1/n$). By Definition 1 we can take $v' \in N_i$ such that $d(v, v') \leq 2R(v, \epsilon_i)$. As u is ϵ-far from v, it holds that

$$d_X(v, v') \leq 2R(v, \epsilon_i) \leq 2R(v, \epsilon) \leq 2d_X(v, u) . \tag{1}$$

In particular, by the triangle inequality,

$$d_X(u, v') \leq d_X(u, v) + d_X(v, v') \overset{(1)}{\leq} 3d_X(u, v) . \tag{2}$$

The priority of v' is at most 2^{i+1}, hence

$$
\begin{aligned}
d_Y(f(v), f(u)) & \\
& \leq \quad d_Y(f(v), f(v')) + d_Y(f(v'), f(u)) \\
& \leq \quad \alpha(2^{i+1}) \cdot d_X(v, v') + \alpha(2^{i+1}) \cdot d_X(v', u) \\
& \overset{(1)\wedge(2)}{\leq} \quad 5\alpha(2/\epsilon_i) \cdot d_X(v, u) .
\end{aligned}
$$

By the minimality of i it follows that $1/\epsilon_i \leq 2/\epsilon$, and since α is monotone

$$d_Y(f(v), f(u)) \leq 5\alpha(2/\epsilon_i) \cdot d_X(v, u) \leq 5\alpha(4/\epsilon) \cdot d_X(v, u) ,$$

as required. Since we desire distortion guarantee for pairs that are $\epsilon/2$-far, the distortion becomes $O(\alpha(8/\epsilon))$.

$\square$

Remark 1. *The proof of Theorem 2 provides an even stronger conclusion, that any pair $u, v \in X$ such that one is $\epsilon/2$-far from the other, has the claimed distortion bound. While in the original definition of coarse scaling both points are required to be $\epsilon/2$-far from each other, it is often the case that we achieve the stronger property. Yet, in some of the cases in previous work the weaker definition seemed to be of importance. Combining Theorem 2 and Theorem 1, we infer that essentially any coarse scaling embedding can have such a one-sided guarantee, with a slightly worse dependence on ϵ, as claimed in the following corollary.*

Corollary 6. *Let $\mu : \mathbb{N} \to \mathbb{R}^+$ be a non-increasing function such that $\sum_{i \geq 1} \mu(i) = 1$. Let $\mathcal{Y}$ be a family of finite metric spaces, and assume that for every finite metric space (Z, d_Z) there exists a non-contractive embedding $f_Z : Z \to Y_Z$, where $(Y_Z, d_{Y_Z}) \in \mathcal{Y}$, with (monotone non-increasing) coarse scaling distortion $\gamma(\epsilon)$. Then given any finite metric space X, there exists an embedding $f : X \to Y$, for some $(Y, d_Y) \in \mathcal{Y}$, with (monotone non-decreasing) one-sided coarse scaling distortion $O(\gamma(\mu(8/\epsilon)))$.*

Proof. By the assumption, there exists $(Y, d_Y) \in \mathcal{Y}$ so that X embeds to Y with coarse scaling distortion $\gamma(\epsilon)$. According to Theorem 1, there is an embedding f with prioritized distortion $\gamma(\mu(i))$ (w.r.t to any fixed priority ranking π). We pick π to be the ordering required by Theorem 2, and conclude that f has one-sided coarse scaling distortion $O(\gamma(\mu(8/\epsilon)))$. $\square$

3.3 Coarse Partial Distortion and Terminal Distortion

As a special case of the reductions Theorem 1 and Theorem 2, we can prove an equivalence between coarse partial distortion to terminal distortion.

Definition 2 (Coarse partial distortion). *Let $(X, d_X), (Y, d_Y)$ be metric spaces, and let $\epsilon \in (0, 1)$, $\gamma \geq 1$. A non-contractive embedding $f : X \to Y$ has $(1 - \epsilon)$-coarse partial scaling distortion γ, if every pair $v, u \in X$ such that both u, v are $\epsilon/2$-far from each other, has distortion at most γ.*

Note that the embedding f has coarse scaling distortion γ if and only if for every $\epsilon \in (0, 1)$, f has $(1 - \epsilon)$-coarse partial distortion $\gamma(\epsilon)$.

Definition 3 (Terminal distortion). *Let $(X, d_X), (Y, d_Y)$ be metric spaces, and $K \subseteq X$ a subset of terminals. A non-contractive embedding $f : X \to Y$ has terminal distortion α w.r.t. K, if every pair $(v, u) \in K \times X$ has distortion at most α.*

Note that for a priority ranking $\pi = x_1, \ldots, x_n$, the embedding f has priority distortion α w.r.t π if and only if for every k, f has terminal distortion $\alpha(k)$ w.r.t. $K = \{x_1, \ldots, x_k\}$. It is important to note that Definition 3 differs from the original definition of terminal distortion in [EFN17], which did not require f to be non-contractive on *all pairs*. We elaborate on this issue in Subsection 3.3.1.

Theorem 3. *Let $\mu : \mathbb{N} \to \mathbb{R}^+$ be a non-increasing function such that $\sum_{i \geq 1} \mu(i) = 1$. Let $k \in \mathbb{N}$. Let $\mathcal{Y}$ be a family of finite metric spaces, and assume that for every finite metric space (Z, d_Z) there exists an embedding $\phi : Z \to Y_Z$, where $(Y_Z, d_{Y_Z}) \in \mathcal{Y}$, with coarse $(1 - 1/(2k))$-partial distortion γ. Then, given a finite metric space (X, d_X) and a set of terminals $K \subsetneq X$ of size $|K| = k$, there exists an embedding $f : X \to Y$, for some $(Y, d_Y) \in \mathcal{Y}$, with terminal distortion γ.*

Proof. Simply follow the proof of Theorem 1, using $\mu(x) = \frac{1}{2k}$ for $x \in K$, and $\mu(x) = \frac{1}{2(|X|-k)}$ for $x \in X \setminus K$. As every $x \in K$ has $\frac{n}{2k}$ copies, and the new metric Z contains at most $2n$ points, x is $\frac{1}{4k}$-far from any other $y \in X$ (in the metric space Z). Also this y is $\frac{1}{4k}$-far from x (since $|B_Z(y, d(x, y))| \geq n/(2k)$). Thus the embedding with coarse $(1 - 1/(2k))$-partial distortion for Z has distortion at most γ for such a pair x, y. □

Theorem 4. *Let (X, d_X), (Y, d_Y) be metric spaces, and $k \in \mathbb{N}$ a parameter. There exists a subset $K \subseteq X$ of size k, such that the following holds: If there exists an embedding $f : X \to Y$ with terminal distortion α, then f has coarse $(1 - \frac{2}{k})$-partial scaling distortion 5α.*

Proof. Following the lines of the proof of Theorem 2 let K be a $\frac{1}{k}$-density net. Fix a pair $v, u \in V$ so that u is $\frac{1}{2} \cdot \frac{2}{k} = \frac{1}{k}$-far from v. Let $v' \in N$ such that $d_X(v, v') \leq 2R(v, \frac{1}{k}) \leq 2d_X(v, u)$. It holds that,

$$
\begin{aligned}
d_Y(f(v), f(u)) &\leq d_Y(f(v), f(v')) + d_Y(f(v'), f(u)) \\
&\leq \alpha \cdot d_X(v, v') + \alpha \cdot d_X(v', u) \\
&\leq 2\alpha \cdot d_X(v, u) + 3\alpha \cdot d_X(v, u) = 5\alpha \cdot d_X(v, u) .
\end{aligned}
$$
□

Among other implications, Theorem 4 implies the following:

Corollary 7. *For every parameters $0 < \epsilon, \delta < 1$, every n-vertex weighted graph G contains a spanning tree with $(1 - \epsilon)$-coarse partial distortion $O(\frac{1}{\epsilon\delta})$ and $1 + \delta$ lightness.*

Proof. In [EFN17] it was shown that for every weighted graph $G = (V, E, w)$ and terminal set $K \subseteq V$ of size k, there is a spanning tree T with terminal distortion $O(k)$ and constant lightness. Using Theorem 8 (proven below), we get that G contains a spanning tree with terminal distortion $O(k/\delta)$ and lightness $1 + \delta$. Now, Theorem 4 (with $k = \frac{2}{\epsilon}$) implies the corollary. □

3.3.1 Weak Terminal Distortion

Our definition of terminal distortion has a one-sided guarantee on *all pairs*, e.g., the embedding must not contract any distance. This definition differs from the original definition of terminal distortion which appears in [EFN17], where the non-contractive requirement was missing (formally, by [EFN17], f has terminal distortion α iff there is some constant $c \in \mathbb{R}$ such that $\forall (v, u) \in K \times X, d_X(u, v) \leq c \cdot d_Y(f(u), f(v)) \leq \alpha \cdot d_X(u, v)$). We will refer to the original definition from [EFN17] as *weak terminal distortion*.

This two definitions are indeed different. For example, in [EFN17] it was shown that given n points containing k terminals in $\mathbb{R}^n$, they can be embedded into $\mathbb{R}^{O(\log k)}$ with weak terminal distortion $O(1)$ (under the ℓ_2-norm). However, any non-contracting embedding with constant distortion requires $\Omega(\log n)$ dimensions, so this is impossible under our Definition 3. As a result of the difference between these definitions, there are some results in [EFN17] on which the reduction of Theorem 4 cannot be used.

Nevertheless, if the target space is ℓ_p, we devise a transformation from weak terminal distortion into terminal distortion, while increasing the dimension *additively* by $O(\log n)$. The first step is Theorem 5, in which we extend a standard embedding into a terminal one, in a different manner than [EFN17]. This theorem has other implications: in particular, we generalize and improve the dimension in a result of [ABC$^+$05, ABN11] on embedding into ℓ_p with coarse partial distortion.

Theorem 5. *Let (X, d_X) be metric space of size n, and $K \subseteq X$ be a subset of terminals. Suppose that there exists an embedding $f : K \to \ell_p^\beta$ with distortion α, then there is an embedding $\hat{f} : X \to \ell_p^{\beta + O(\log n)}$ with terminal distortion $O(\alpha)$.*

Proof. Assume, as we may, that f is non-contractive. That is, for every $v, u \in K$, $d_X(u, v) \leq \|f(u) - f(v)\|_p \leq \alpha \cdot d_X(u, v)$.

Fix $m = O(\log n)$. Let $g : X \to \{\pm 1\}^m$ such that for every $v, u \in X$, there are at least $\frac{m}{4}$ coordinates i where $g_i(v) \neq g_i(u)$ (a random g will work with high probability, as can be verified by Chernoff inequality). For every vertex $u \in X$, let k_u be the closest terminal to u. The embedding $\hat{f}$ is defined as follows. For $u \in X$,

$$\hat{f}(u) = f(k_u) \oplus \frac{d_X(u, k_u)}{m^{1/p}} \cdot g(u) \ .$$

First, we will show that $\hat{f}$ has expansion at most $O(\alpha)$ on terminal pairs. Fix some $v \in K$ and $u \in X$.

$$
\begin{aligned}
\|\hat{f}(v) - \hat{f}(u)\|_p^p &= \|f(v) - f(k_u)\|_p^p + \frac{1}{m} \sum_{i=1}^m |g_i(u) \cdot d_X(u, k_u)|^p \\
&\leq \alpha^p \cdot d_X(v, k_u)^p + d_X(u, k_u)^p \\
&\leq (\alpha^p + 1) \cdot (d_X(v, u) + d_X(u, k_u))^p \\
&\leq (\alpha^p + 1) \cdot (2 d_X(v, u))^p \ .
\end{aligned}
$$

Thus, $\|\hat{f}(v) - \hat{f}(u)\|_p \leq 2(\alpha^p + 1)^{1/p} \cdot d_X(v, u) \leq 2(\alpha + 1) \cdot d_X(v, u)$.

Next, we bound the contraction for all pairs. Fix some $v, u \in X$. If $d_X(u, v)/2 \geq d_X(v, k_v) + d_X(u, k_u)$, then

$$
\begin{aligned}
\|\hat{f}(v) - \hat{f}(u)\|_p &\geq \|f(k_v) - f(k_u)\|_p \geq d_X(k_v, k_u) \\
&\geq d_X(v, u) - d_X(v, k_v) - d_X(u, k_u) \geq d_X(v, u)/2 \ .
\end{aligned}
$$

11

Otherwise,

$$\|\hat{f}(v) - \hat{f}(u)\|_p^p \geq \frac{1}{m} \sum_{i=1}^{m} |g_i(v) \cdot d_X(v, k_v) - g_i(u) \cdot d_X(u, k_u)|^p$$

$$= \frac{1}{m} \cdot \frac{m}{4} \cdot |d_X(v, k_v) + d_X(u, k_u)|^p \geq \frac{1}{4} \cdot (d_X(v, u)/2)^p .$$

To ensure the embedding does not contract, our final embedding will be $2^{1+\frac{2}{p}} \cdot \hat{f}$. $\qquad\square$

We now show the transformation from weak terminal distortion to (non-contracting) terminal distortion. (By Theorem 4 this can provide embeddings with coarse partial distortion as well.) Suppose an embedding $f : X \to Y$ has weak terminal distortion α. In particular, its restriction to K has distortion α. Using Theorem 5 we conclude:

Corollary 8. *Let (X, d_X) be metric space of size n, and $K \subseteq X$ be a subset of terminals. Suppose that there exists an embedding $f : X \to \ell_p^\beta$ with weak terminal distortion α, then there exist embedding $\hat{f} : X \to \ell_p^{\beta + O(\log n)}$ with terminal distortion $O(\alpha)$.*

Fix some $p \geq 1$. Let $\mathcal{X}$ be a subset-closed family of finite metric spaces such that for any $n \geq 1$ and any n-point metric space $X \in \mathcal{X}$ there exists an embedding $f_X : X \to \ell_p$ with distortion $\alpha(n)$ and dimension $\beta(n)$. In [ABC+05, ABN11] it was shown that, assuming all f_X are strongly non-expansive, [11] there is a universal constant C and an embedding from X into ℓ_p with $(1 - \epsilon)$-coarse partial distortion $O(\alpha(C/\epsilon))$ and dimension $\beta(C/\epsilon) \cdot O(\log n)$. By combining Theorem 5 with Theorem 4 we considerably improve the dimension, and remove the strongly non-expansive requirement.

Corollary 9. *Fix some $p \geq 1$. Let $\mathcal{X}$ be a subset-closed family of finite metric spaces such that for any $n \geq 1$ and any n-point metric space $X \in \mathcal{X}$ there exists an embedding $f_X : X \to \ell_p$ with distortion $\alpha(n)$ and dimension $\beta(n)$. Then there is an embedding from X into ℓ_p with $(1 - \epsilon)$-coarse partial distortion $O(\alpha(2/\epsilon))$ and dimension $\beta(2/\epsilon) + O(\log n)$.*

4 Light Spanner with Prioritized Distortion

In this section we prove that every graph admits a light spanner with bounded prioritized distortion. (The runtime analysis of our algorithms is appears in Subsection 4.2.)

Theorem 6 (Prioritized Spanner). *Given a graph $G = (V, E)$, a parameter $0 < \rho < 1$ and any priority ranking $v_1, v_2, \ldots, v_n$ of V, there exists a spanner H with lightness $1 + \rho$ and prioritized distortion $\tilde{O}(\log j)/\rho$.*

Combining Theorem 6 and Theorem 2 we obtain the following.

Theorem 7. *For any parameter $0 < \rho < 1$, any graph contains a spanner with coarse scaling distortion $\tilde{O}(\log(1/\epsilon))/\rho$ and lightness $1 + \rho$. Moreover, the spanner can be computed in $O(n \cdot (m + n \log n))$ time.*

---

[11] Embedding $f : X \to \ell_p$ is strongly non-expansive if $f = (\eta_1 f_1, \ldots, \eta_m f_m)$ where $\sum_{i=1}^{m} \eta_i = 1$, and each f_i is non-expansive embedding into $\mathbb{R}$.

By Lemma 1 it follows that this spanner has ℓ_q-distortion $\tilde{O}(q)/\rho$ for any $1 \le q < \infty$.

We can also obtain a spanner with both scaling distortion and prioritized distortion simultaneously, where the priority is with respect to an arbitrary ranking $\pi = v_1, \ldots, v_n$. To achieve this, one may define a ranking which interleaves π with the ranking generated in the proof of Theorem 2.

We now turn to proving Theorem 6. The proof is based on the following main technical lemma:

Lemma 10. *Given a graph $G = (V, E)$, a subset $K \subseteq V$ of size k, and a parameter $0 < \delta < 1$, there exists a spanner H that 1) contains the MST of G, 2) has lightness $1 + \delta$, and 3) every pair in $K \times V$ has distortion $O((\log k)/\delta)$.*

Before proving this lemma, let us first apply it to prove Theorem 6.

Proof of Theorem 6. For every $1 \le i \le \lceil \log \log n \rceil$ let $K_i = \left\{ v_j \ : \ j \le 2^{2^i} \right\}$. Let H_i be the spanner given by Lemma 10 with respect to the set K_i and the parameter $\delta_i = \rho/i^2$. Hence H_i has $1 + \rho/i^2$ lightness and $O\left(\frac{\log |K_i|}{\delta_i} \right) = O(2^i \cdot i/\rho)$ distortion for pairs in $K_i \times V$. Let $H = \bigcup_i H_i$ be the union of all these spanners (that is, the graph containing every edge of every one of these spanners). As each H_i contains the unique MST of G, it holds that

$$\Psi(H) \ \le \ 1 + \sum_{i \ge 1} \rho/i^2 = 1 + O(\rho) \ .$$

To see the prioritized distortion, let $v_j, v_r \in V$ be such that $j < r$, and let $1 \le i \le \lceil \log \log n \rceil$ be the minimal index such that $v_j \in K_i$. Note that $2^{2^{i-1}} \le j$, and in particular $2^{i-1} \le \log j$ (with the exception of $j = 1$, but that case holds by the virtue of $j = 2$, say). This implies that

$$\begin{aligned} d_H(v_j, v_r) \ &\le \ d_{H_i}(v_j, v_r) \ \le \ O(2^i \cdot i^2/\rho) \cdot d_G(v_j, v_r) \\ &\le \ \tilde{O}(\log j)/\rho \cdot d_G(v_j, v_r) \ . \end{aligned}$$

as required. □

4.1 Proof of Lemma 10

The construction of the spanner that fulfills the properties promised in Lemma 10 is as follows. First, we use the spanner of [CW18] to get a spanner with lightness $O(1)$ and distortion $O(\log k)$ over pairs in $K \times K$. Then, by combining this spanner with the SLT by [KRY93], we expand the $O(\log k)$ distortion guarantee to all pairs in $K \times V$, while the lightness is still $O(1)$. Finally, we use a general reduction (Theorem 8), that reduces the weight of a spanner while increasing its distortion. By applying the reduction, we get a spanner with $1 + \rho$ lightness while paying additional factor of $1/\rho$ in the distortion.

We begin by describing the general reduction.

Theorem 8. *Let $G = (V, E)$ be a graph, $0 < \delta < 1$ a parameter and $t : \binom{V}{2} \to \mathbb{R}_+$ some function. Suppose that for every weight function $w : E \to \mathbb{R}_+$ there exists a spanner H with lightness ℓ such that every pair $u, v \in V$ suffers distortion at most $t(u, v)$. Then for every weight function w there exists a spanner H with lightness $1 + \delta\ell$ and such that every pair u, v suffers distortion at most $t(u, v)/\delta$. Moreover, H contains the MST of G with respect to w.*

Proof. Fix some weight function w and let $G = (V, E, w)$ be the graph associated with this weight function, and let T be the MST of G. Set $w' : E \to \mathbb{R}_+$ to be a new weight function

$$w'(e) = \begin{cases} w(e) & e \in T \\ w(e)/\delta & e \notin T \end{cases} \,,$$

that is, we multiply the weight of all non-MST edges by $1/\delta$. Let $G' = (V, E, w')$ be the graph G associated with the new weight function w'. Note that T is also the MST of G' (since the weight of any spanning tree is higher in G' than in G except for T itself). By our assumption there exists a spanner $H' = (V, E_{H'}, w')$ of G' with distortion bounded by t and lightness ℓ. Set $H = (V, E_{H'} \cup T, w)$ as a spanner of G. The edge set of H consists of the edges of H' together with the MST edges, all with the original weight function w.

As the weight of the non-MST edges are larger in G' by $1/\delta$ factor compared to their weight in G, we have

$$\begin{aligned} w(E_H) &=& w(T) + w\left(E_{H'} \setminus T\right) = w(T) + \delta \cdot w'\left(E_{H'} \setminus T\right) \leq w(T) + \delta \cdot w'\left(E_{H'}\right) \\ &\leq& w(T) + \delta\ell \cdot w'\left(T\right) = (1 + \delta\ell) \cdot w(T) \,, \end{aligned}$$

concluding that the lightness of H is at most $1 + \delta\ell$.

To bound the distortion, consider an arbitrary pair of vertices $u, v \in V$. Let $P_{u,v}$ be the shortest path from u to v in G. As for each edge $e \in P_{u,v}$, $w'(e) \leq w(e)/\delta$ we have that

$$d_{G'}(u, v) \leq \sum_{e \in P_{u,v}} w'(e) \leq \sum_{e \in P_{u,v}} \frac{1}{\delta} \cdot w(e) = \frac{1}{\delta} \cdot d_G(u, v) \,,$$

Therefore:

$$d_H(u, v) \leq d_{H'}(u, v) \leq t(u, v) \cdot d_{G'}(u, v) \leq \frac{t(u, v)}{\delta} d_G(u, v) \,,$$

as required. $\qquad\qquad\square$

In a recent work, Chechik and Wulff-Nilsen achieved the following result:

Theorem 9 ([CW18])**.** *For every weighted graph $G = (V, E, w)$ and parameters $k \geq 1$ and $0 < \epsilon \leq 1$, there exist a polynomial time algorithm that constructs a spanner with distortion $(2t-1)(1+\epsilon)$ and lightness $n^{1/t} \cdot \mathrm{poly}(\frac{1}{\epsilon})$.*

Note that for an n-vertex graph with parameters $t = \log n, \epsilon = 1$, they get a spanner with distortion $O(\log n)$ and constant lightness. However, their construction does not seem to provide lightness arbitrarily close to 1.

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

We will use [CW18] spanner to construct a spanner with $O(1)$ lightness and distortion $O(\log k)$ over pairs in $K \times K$. Let $G_k = (K, \binom{K}{2}, w_k)$ be the complete graph over the terminal set K with weights $w_k(u, v) = d_G(u, v)$ (for $u, v \in K$) that are given by the shortest path metric in G. Let T_k be the MST of G_k. Note that the MST T of G is a Steiner tree of G_k, hence $w_k(T_k) \leq 2 \cdot w_k(SMT(G_k)) \leq 2 \cdot w(T)$.

Using Theorem 9, let $H_k = (K, E_k, w_k)$ be a spanner of G_k with weight $O(w_k(T_k)) = O(w(T))$ (constant lightness) and distortion $O(\log k)$. For a pair of vertices $u, v \in K$, let P_{uv} denote the shortest path

between u and v in G. Let $H' = (V, E', w)$ be a subgraph of G with the set of edges $E' = \cup_{\{u,v\}\in E_k} P_{uv}$ (i.e. for every edge $\{u, v\}$ in H_k, we take the shortest path from u to v in G). It holds that,

$$w(H') \le \sum_{\{u,v\}\in E_k} w(P_{uv}) = \sum_{e\in E_k} w_k(e) = O(w(T)) \ .$$

Moreover, for every pair $u, v \in K$,

$$d_{H'}(u, v) \le d_{H_k}(u, v) \le O(\log k) \cdot d_{G_k}(u, v) = O(\log k) \cdot d_G(u, v) \ . \tag{3}$$

Now we extend H' so that every pair in $K \times V$ will suffer distortion at most $O(\log k)$. To this end, we use the following lemma regarding shallow light trees (SLT), which is implicitly proved in [KRY93, ABP92].

Lemma 11. *Given a graph $G = (V, E)$, a parameter $\alpha > 1$, and a subset $K \subseteq V$, there exists a spanner S [12] of G with lightness $1 + \frac{2}{\alpha-1}$, and for any vertex $u \in V$, $d_S(u, K) \le \alpha \cdot d_G(u, K)$.*

Let S be the spanner of Lemma 11 with respect to the set K and parameter $\alpha = 2$. Define H" as the union of H' and S. As both H' and S have constant lightness, so does H". It remains to bound the distortion of an arbitrary pair $v \in K$ and $u \in V$. Let $k_u \in K$ be the closest vertex to u among the vertices in K with respect to the distances in the spanner S. By the assertion of Lemma 11,

$$d_S(u, k_u) = d_S(u, K) \le 2 \cdot d_G(u, K) \le 2 \cdot d_G(u, v) \ . \tag{4}$$

Using the triangle inequality,

$$d_G(v, k_u) \le d_G(v, u) + d_G(u, k_u) \le d_G(v, u) + d_S(u, k_u) \overset{(4)}{\le} 3 \cdot d_G(v, u) \ . \tag{5}$$

Since both $v, k_u \in K$ it follows that

$$d_{H'}(v, k_u) \overset{(3)}{\le} O(\log k) \cdot d_G(v, k_u) \overset{(5)}{\le} O(\log k) \cdot d_G(v, u) \ . \tag{6}$$

We conclude that

$$d_{H"}(v, u) \le d_{H'}(v, k_u) + d_S(k_u, u) \overset{(4)\wedge(6)}{\le} O(\log k) \cdot d_G(v, u) \ .$$

We showed a polynomial time algorithm, that given a weighted graph $G = (V, E, w)$ and a subset $K \subseteq V$ of size k, constructs a spanner H with lightness $O(1)$, and such that every pair in $K \times V$ has distortion at most $O(\log k)$. Now Theorem 8 implies Lemma 10.

4.2 Efficient Implementations

First we describe an $O(k \cdot (m + n \log n))$-time algorithm for the terminal spanner of Lemma 10. We start by constructing G_k, the full graph on K. For this goal we need to compute all distances in the terminal set K. This can be done by computing k shortest path trees rooted in each terminal vertex in K, using Dijkstra's algorithm it will take $O(k \cdot (m + n \log n))$ time. Our next step is to compute a light spanner for G_k. Instead of using [CW18] in the construction of the spanner H_k, we will use a more efficient construction by Alstrup et. al. [ADF+17], who provide us with a $O(\log k)$ stretch spanner with $O(1)$ lightness and $O(k)$ edges in

---

[12]In fact, S is a spanning forest of G.

$O(k^2)$ time.[13] Next, for every edge $\{u, v\} \in H_k$ we find the shortest path between u and v in G and add it to H'. This is done using the shortest path tree rooted in u computed earlier in $O(n)$ time per edge. Thus H' is constructed from H_k in total $O(k \cdot n)$ time. Our next step is to compute an SLT rooted in K using the $O(m + n \log n)$ time algorithm of [KRY93]. Finally, it is straightforward that the spanner reduction lemma takes linear time (once the MST is computed). We conclude that the total running time of the spanner construction of Lemma 10 is $O(k \cdot (m + n \log n))$.

Next, in order to compute the prioritized spanner of Theorem 6 we simply construct terminal spanners for the sets K_i, $1 \leq i \leq \lceil \log \log n \rceil$. The total time required is $\sum_{i=1}^{\lceil \log \log n \rceil} O\left(|K_i| \cdot (m + n \log n)\right) = O\left(n \cdot (m + n \log n)\right)$.

Finally, we analyze the running time of Theorem 7. Theorem 7 is achieved by combining Theorem 6 with Theorem 2. First we need to compute the priority ranking of Theorem 6, which is based on density nets. Specifically we need to compute for every $1 \leq i \leq \lceil \log n \rceil$, a 2^{-i} density net. [CDG06] did not analyzed explicitly the running time required for computing a density net, but it is not hard to see that the running time of their algorithm is $O(n \cdot (m + n \log n))$ (the time it takes to compute all-pairs-shortest-paths).

5   A Light Tree with Constant Average Distortion

Here we prove our main theorem on finding a light spanning tree with constant average distortion. Later on we show a probabilistic embedding into a distribution of light spanning trees with improved bound on higher moments of the distortion.

Theorem 10. *For any parameter $0 < \rho < 1$, any graph contains a spanning tree with scaling distortion $\tilde{O}(\sqrt{1/\epsilon})/\rho$ and lightness $1 + \rho$. Moreover, this tree can be found in $\tilde{O}(m \cdot n)$ time.*

It follows from Lemma 1 that the average distortion of the spanning tree obtained is $O(1/\rho)$. Moreover, the ℓ_q-distortion is $O(1/\rho)$ for any fixed $1 \leq q < 2$, $\tilde{O}\left(\log^{1.5} n\right)/\rho$ for $q = 2$, and $\tilde{O}(n^{1-2/q})/\rho$ for any fixed $2 < q < \infty$.

We will need the following simple lemma, that asserts the scaling distortion of a composition of two maps is essentially the product of the scaling distortions of these maps.[14]

Lemma 12. *Let (X, d_X), (Y, d_Y) and (Z, d_Z) be metric spaces. Let $f : X \to Y$ (respectively, $g : Y \to Z$) be a non-contractive onto embedding with scaling distortion α (resp., β). Then $g \circ f$ has scaling distortion $\alpha(\epsilon/2) \cdot \beta(\epsilon/2)$.*

Proof. Let $n = |X|$. Let $\text{dist}_f(v, u) = \frac{d_Y(f(v), f(u))}{d_X(v, u)}$ be the distortion of the pair $u, v \in X$ under f, and similarly let $\text{dist}_g(v, u) = \frac{d_Z(g(f(v)), g(f(u)))}{d_Y(f(v), f(u))}$. Fix some $\epsilon \in (0, 1)$. We would like to show that at most $\epsilon \cdot \binom{n}{2}$ pairs suffer distortion greater than $\alpha(\epsilon/2) \cdot \beta(\epsilon/2)$ by $g \circ f$. Let $A = \left\{ \{v, u\} \in \binom{X}{2} : \text{dist}_f(v, u) > \alpha(\epsilon/2) \right\}$ and $B = \left\{ \{v, u\} \in \binom{X}{2} : \text{dist}_g(v, u) > \beta(\epsilon/2) \right\}$. By the bound on the scaling distortions of f and g, it

---

[13]See Theorem 4 in [ADF+17]. In fact, for n-vertex graph with m edges with stretch parameter $O(\log n)$, their running time is $O(m + n^{1+\epsilon'})$ for arbitrarily small constant ϵ'.

[14]Note that this is not true for the average distortion – one may compose two maps with constant average distortion and obtain a map with $\Omega(n)$ average distortion.

holds that $|A \cup B| \leq |A| + |B| \leq \epsilon \cdot \binom{n}{2}$. Note that if $\{v, u\} \notin A \cup B$ then

$$
\begin{aligned}
\frac{d_Z\left(g(f(v)), g(f(u))\right)}{d_X\left(v, u\right)} &= \operatorname{dist}_f(v, u) \cdot \operatorname{dist}_g(v, u) \\
&\leq \alpha(\epsilon/2) \cdot \beta(\epsilon/2),
\end{aligned}
$$

which concludes the proof. $\qquad\square$

We will also need the following result, that was proved in [ABN15]. (The bound on the running time was not explicitly stated in [ABN15]. Their algorithm uses the star-decomposition of [EEST08] whose running time is $\tilde{O}(m)$, and a certain iterative algorithm that chooses a radius for each of the $\tilde{O}(n)$ cones, which can be trivially implemented in $\tilde{O}(n)$ time.)

Theorem 11 ([ABN15]). *Any graph contains a spanning tree with scaling distortion $O(\sqrt{1/\epsilon})$, and this tree can be found in $\tilde{O}(m \cdot n)$ time.*

Now we can prove the main result.

Proof. (of Theorem 10) Let H be the spanner given by Theorem 7. Let T be a spanning tree of H constructed according to Theorem 11. By Lemma 12, T has scaling distortion $O(\sqrt{1/\epsilon}) \cdot \tilde{O}(\log(1/\epsilon))/\rho = \tilde{O}(\sqrt{1/\epsilon})/\rho$ with respect to the distances in G. The lightness follows as $\Psi(T) \leq \Psi(H) \leq 1 + \rho$.
$\qquad\square$

Random Tree Embedding. We also derive a result on probabilistic embedding into light spanning trees with scaling distortion. That is, the embedding construct a distribution over spanning tree so that each tree in the support of the distribution is light. In such probabilistic embeddings [Bar96] into a family $\mathcal{Y}$, each embedding $f = f_Y : X \to Y$ (for some $(Y, d_Y) \in \mathcal{Y}$) in the support of the distribution is non-contractive, and the distortion of the pair $u, v \in X$ is defined as $\mathbb{E}_Y\left[\frac{d_Y(f(u), f(v))}{d_X(u, v)}\right]$. The prioritized and scaling distortions are defined accordingly. We make use of the following result from [ABN15].[15]

Theorem 12. *([ABN15]) Every weighted graph G embeds into a distribution over spanning trees with coarse scaling distortion $\tilde{O}(\log^2(1/\epsilon))$.*

We note that the distortion bound on the composition of maps in Lemma 12 also holds whenever g is a random embedding, and we measure the scaling expected distortion. Thus, following the same lines as in the proof of Theorem 10, (while using Theorem 12 instead of Theorem 11), we obtain the following.

Theorem 13. *For any parameter $0 < \rho < 1$ and any weighted graph G, there is an embedding of G into a distribution over spanning trees with scaling distortion $\tilde{O}(\log^3(1/\epsilon))/\rho$, such that every tree T in the support has lightness $1 + \rho$.*

It follows from Lemma 1 that the ℓ_q-distortion is $O(1/\rho)$, for every fixed $q \geq 1$.

---

[15]The fact the embedding yields coarse scaling distortion is implicit in their proof.

6 Lower Bound on the Trade-off between Lightness and Average Distortion

In this section, we give an example of a graph for which any spanner with lightness $1 + \rho$ has average distortion $\Omega(1/\rho)$ (of course this bound holds for the ℓ_q-distortion as well). This shows that our results are tight [16].

Lemma 13. *For any $n \geq 32$ and $\rho \in [1/n, 1/32]$, there is a graph G on $n + 1$ vertices such that any spanner H of G with lightness at most $1 + \rho$ has average distortion at least $\Omega(1/\rho)$.*

Proof. We define the graph $G = (V, E)$ as follows. Denote $V = \{v_0, v_1, \ldots, v_n\}$, $E = \binom{V}{2}$, and the weight function w is defined as follows.

$$w(\{v_i, v_j\}) = \begin{cases} 1 & \text{if } |i - j| = 1 \\ 2 & \text{otherwise} . \end{cases}$$

I.e., G is a complete graph of size $n + 1$, where the edges $\{v_i, v_{i+1}\}$ have unit weight and induce a path of length n, and all non-path edges have weight 2. Clearly, the path is the MST of G of weight n. Let $k = \lceil \rho n \rceil$. Let H be some spanner of G with lightness at most $1 + \rho \leq \frac{n+k}{n}$, in particular, $w(H) \leq n + k$. Clearly H has at least n edges (to be connected). Let q be the number of edges of weight 2 contained in H. Then $w(H) \geq (n - q) \cdot 1 + q \cdot 2 = n + q$. Therefore $q \leq k$.

Let S be the set of vertices which are incident on an edge of weight 2 in H. Then $|S| \leq 2q \leq 2k$. Let $\delta = \frac{1}{32\rho}$. For any $v \in S$, let $N_v \subseteq V$ be the set of vertices that are connected to v via a path of length at most δ in H, such that this path consists of weight 1 edges only. Necessarily, for any $v \in S$, $|N_v| \leq 2\delta + 1$. Let $N = \bigcup_{v \in S} N_v$, it holds that $|N| \leq 2k \cdot (2\delta + 1) \leq 4\rho n(\frac{1}{16\rho} + 1) \leq \frac{n}{4} + \frac{n}{8} = \frac{3n}{8}$. Let $\bar{N} = V \setminus N$.

Consider $u \in \bar{N}$. By definition of N every weight 2 edge is further than δ steps away from u in H. It follows that there are at most $2\delta + 1$ vertices within distance at most δ from u (in H). Let $F_u = \{v \in V : d_H(u, v) > \delta\}$. It follows that $|F_u| \geq n - 2\delta - 1$. Note that for any $v \in F_u$, the distortion of the pair $\{u, v\}$ is at least $\frac{\delta}{2}$. Hence, we obtain that

$$\sum_{\{v,u\} \in \binom{V}{2}} \frac{d_H(v, u)}{d_G(v, u)} \geq \frac{1}{2} \sum_{u \in \bar{N}} \sum_{v \in F_u} \frac{d_H(v, u)}{d_G(v, u)}$$

$$\geq \frac{5n}{16} \cdot (n - 2\delta - 1) \cdot \frac{\delta}{2}$$

$$\geq \frac{5n}{16} \cdot \frac{7n}{8} \cdot \frac{1}{64\rho} .$$

Finally, the average distortion is bounded as follows.

$$\text{dist}_1(H) = \frac{1}{\binom{n+1}{2}} \sum_{\{v,u\} \in \binom{V}{2}} \frac{d_H(v, u)}{d_G(v, u)}$$

$$\geq \frac{n}{n + 1} \cdot \frac{35}{64} \cdot \frac{1}{64\rho}$$

$$\geq \frac{1}{128\rho} . \qquad \square$$

---

[16]We also mention that in general the average distortion of a spanner cannot be arbitrarily close to 1, unless the spanner is extremely dense. E.g., when G is a complete graph, any spanner with lightness at most $n/4$ will have average distortion at least $3/2$.

7 Conclusions

In this paper we constructed a spanning tree with $1 + \rho$ lightness and $O(\frac{1}{\rho})$ average distortion for any parameter $\rho \in (0, 1)$. We also proved that this tradeoff is best possible, up to constant factors. The main technical contribution is a new equivalence theorem between prioritized and coarse scaling distortions. We used this equivalence theorem in order to derive several other results in metric embeddings. In particular, we show that every finite metric space embeds into ℓ_p space with prioritized distortion $O(\log j)$. We hope that this equivalence theorem will lead to additional results and applications.

8 Acknowledgments

We are grateful to Michael Elkin and Shiri Chechik for fruitful discussions.

References

[ABC+05] Ittai Abraham, Yair Bartal, Hubert T.-H. Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 83–100, 2005. 5, 11, 12

[ABN08] Ittai Abraham, Yair Bartal, and Ofer Neiman. Nearly tight low stretch spanning trees. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '08, pages 781–790, Washington, DC, USA, 2008. IEEE Computer Society. 5

[ABN11] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026 – 3126, 2011. 2, 3, 4, 5, 7, 8, 11, 12

[ABN15] 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. 1, 2, 3, 4, 5, 6, 17

[ABNS14] Ittai Abraham, Yair Bartal, Ofer Neiman, and Leonard J. Schulman. Volume in general metric spaces. *Discrete & Computational Geometry*, 52(2):366–389, 2014. 5

[ABP92] B. Awerbuch, A. Baratz, and D. Peleg. Efficient broadcast and light-weight spanners. Technical Report CS92-22, The Weizmann Institute of Science, Rehovot, Israel., 1992. 2, 15

[ADF+17] Stephen Alstrup, Søren Dahlgaard, Arnold Filtser, Morten Stöckel, and Christian Wulff-Nilsen. Constructing light spanners deterministically in near-linear time. *CoRR*, abs/1709.01960, 2017. 15, 16

[AKPW95] Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995. 5

[AN12] Ittai Abraham and Ofer Neiman. Using petal-decompositions to build a low stretch spanning tree. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 395–406, New York, NY, USA, 2012. ACM. 5

[Bar96] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pages 184–, Washington, DC, USA, 1996. IEEE Computer Society. 5, 17

[Bar98] 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. ACM. 5

[Bar04]   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. 5

[BFN16]   Yair Bartal, Arnold Filtser, and Ofer Neiman. On notions of distortion and an almost minimum spanning tree with constant average distortion. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 873–882, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics. 1

[BFN17]   Yair Bartal, Nova Fandina, and Ofer Neiman. On moment analysis of metric embedding and its application in dimensionality reduction. *Manuscript*, 2017. 5

[Bou85]   Jean Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1-2):46–52, 1985. 3

[CDG06]   T.-H. Hubert Chan, Michael Dinitz, and Anupam Gupta. Spanners with slack. In *Proceedings of the 14th Conference on Annual European Symposium - Volume 14*, ESA'06, pages 196–207, London, UK, UK, 2006. Springer-Verlag. 4, 5, 8, 16

[CDNS92]   Barun Chandra, Gautam Das, Giri Narasimhan, and José Soares. New sparseness results on graph spanners. In *Proceedings of the Eighth Annual Symposium on Computational Geometry*, SCG '92, pages 192–201, New York, NY, USA, 1992. ACM. 4

[Che15]   Shiri Chechik. Approximate distance oracles with improved bounds. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 1–10, 2015. 8

[CW18]   Shiri Chechik and Christian Wulff-Nilsen. Near-optimal light spanners. *ACM Trans. Algorithms*, 14(3):33:1–33:15, 2018. 4, 13, 14, 15

[EEST08]   Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. *SIAM Journal on Computing*, 38(2):608–628, 2008. 5, 17

[EFN17]   Michael Elkin, Arnold Filtser, and Ofer Neiman. Terminal embeddings. *Theor. Comput. Sci.*, 697:1–36, 2017. 10, 11

[EFN18]   Michael Elkin, Arnold Filtser, and Ofer Neiman. Prioritized metric structures and embedding. *SIAM J. Comput.*, 47(3):829–858, 2018. 2, 3, 4, 5, 7, 8

[ENS15]   Michael Elkin, Ofer Neiman, and Shay Solomon. Light spanners. *SIAM J. Discrete Math.*, 29(3):1312–1321, 2015. 4

[FRT04]   Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, 2004. 5

[FS16]   Arnold Filtser and Shay Solomon. The greedy spanner is existentially optimal. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing, PODC 2016, Chicago, IL, USA, July 25-28, 2016*, pages 9–17, 2016. 5

[GP68]   E. N. Gilbert and H. O. Pollak. Steiner minimal trees. *SIAM Journal on Applied Mathematics*, 16(1):1–29, Jan 1968. 14

[KRY93]   Samir Khuller, Balaji Raghavachari, and Neal E. Young. Balancing minimum spanning and shortest path trees. In *SODA*, pages 243–250, 1993. 2, 13, 15, 16

[KSW09]   Jon Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. *Journal of the ACM*, 56(6):32:1–32:37, September 2009. 2, 3, 5