Terminal Embeddings

Michael Elkin, Arnold Filtser, and Ofer Neiman

Department of Computer Science, Ben-Gurion University of the Negev
Beer-Sheva, Israel
{elkinm,arnoldf,neimano}@cs.bgu.ac.il

Abstract

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

We also strengthen this notion, and consider embeddings that approximately preserve the distances between *all* pairs, but provide improved distortion for pairs containing a terminal. Surprisingly, we show that such embeddings exist in many settings, and have optimal distortion bounds both with respect to $X \times X$ and with respect to $K \times X$.

Moreover, our embeddings have implications to the areas of Approximation and Online Algorithms. In particular, [7] devised an $\tilde{O}(\sqrt{\log r})$-approximation algorithm for sparsest-cut instances with r demands. Building on their framework, we provide an $\tilde{O}(\sqrt{\log |K|})$-approximation for sparsest-cut instances in which each demand is incident on one of the vertices of K (aka, terminals). Since $|K| \leq r$, our bound generalizes that of [7].

1998 ACM Subject Classification G.2.2 Graph algorithms

Keywords and phrases embedding, distortion, terminals

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2015.242

1 Introduction

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

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

Some of the basic results in the field of metric embedding are: a theorem of [15], asserting that any metric space on n points embeds with distortion $O(\log n)$ into Euclidean space (which was shown to be tight by [35]), and probabilistic embedding into a distribution of ultrametrics (or trees) with expected distortion $O(\log n)$ [23], or expected congestion $O(\log n)$ [40] (which are also tight [12]).

In this paper we study a natural variant of embedding, in which the input consists of a finite metric space or a graph, and in addition, a subset of the points are designated as *terminals*. The objective is to embed the metric into a simpler metric (e.g., Euclidean metric), or into a simpler graph (e.g., a tree), while approximately preserving the distances between the terminals to *all other points*. We show that such embeddings, which we call *terminal embeddings*, can have improved parameters compared to embeddings that must preserve all

18th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'15) /
19th Int'l Workshop on Randomization and Computation (RANDOM'15).
Editors: Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim; pp. 242–264
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

pairwise distances. In particular, the distortion (and the dimension in embedding to normed spaces) depends only on the number of terminals, regardless of the cardinality of the metric space.

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

As a possible motivation for studying such embeddings, consider a scenario in which a certain network of clients and servers is given as a weighted graph (where edges correspond to links, weights to communication/travel time). It is conceivable that one only cares about distances between clients and servers, and that there are few servers. We would like to have a simple structure, such as a tree spanning the network, so that the client-server distances in the tree are approximately preserved.

We show that there exists a general phenomenon; essentially any known metric embedding into an ℓ_p space or a graph family can be transformed via a general transformation into a terminal embedding, while paying only a constant blow-up in the distortion. In particular, we obtain a terminal embedding of any finite metric into any ℓ_p space with terminal distortion $O(\log k)$, using only $O(\log k)$ dimensions. We also show that many of the embeddings into normed spaces, probabilistic embedding into ultrametrics (including capacity preserving ones), and into spanning trees, have their *strong* terminal embedding counterparts. Our results are tight in most settings.[1]

It is well known that embedding a graph into a single tree may cause (worst-case) distortion $\Omega(n)$ [38]. However, if one only cares about client-server distances, we show that it is possible to obtain distortion $2k-1$, where k is the number of servers, and that this is tight. Furthermore, we study possible tradeoffs between the distortion and the total weight of the obtained tree. This generalizes the notion of *shallow light trees* [32, 11, 21], which provides a tradeoff between the distortion with respect to a single designated server and the weight of the tree.

We then address probabilistic approximation of metric spaces and graphs by ultrametrics and spanning trees. This line of work started with the results of [4, 12], and culminated in the $O(\log n)$ expected distortion for ultrametrics by [23], and $\tilde{O}(\log n)$ for spanning trees by [3]. These embeddings found numerous algorithmic applications, in various settings, see [23, 19, 3] and the references therein for details. In their work on Ramsey partitions, [36] implicitly showed that there exists a probabilistic embedding into ultrametrics with expected terminal distortion $O(\log k)$ (see Section 2 for the formal definitions). Here we generalize this result by obtaining a *strong* terminal embedding with the same expected $O(\log k)$ distortion guarantee for all pairs containing a terminal, and $O(\log n)$ for all other pairs. We also show a similar result that extends the embedding of [3] into spanning trees, with $\tilde{O}(\log k)$ expected distortion for pairs containing a terminal, and $\tilde{O}(\log n)$ for all pairs. A slightly different notion, introduced by [39], is that of trees which approximate the congestion (rather than the distortion), and [40] showed a distribution over trees with expected congestion $O(\log n)$. We provide a strong terminal version of this result, and show expected congestion of $O(\log k)$ for all edges incident on a terminal, and $O(\log n)$ for the rest. In [1], it was shown that the

[1] All our terminal embeddings are tight, except for the probabilistic spanning trees, where they match the state-of-the-art [3], and except for our terminal spanners.

average distortion (taken over all pairs) in an embedding into a single tree can be bounded by $O(1)$ (in contrast to the $\Omega(\log n)$ lower bound for the average stretch over edges). Here we extend and simplify their result, and obtain $O(1)$ average terminal distortion, that is, the average is over pairs containing a terminal. We do this both in the ultrametric and in the spanning tree settings.

We also consider spanners, with a stretch requirement only for pairs containing a terminal. Our general transformation produces for any $t \geq 1$ a $(4t-1)$-terminal stretch spanner with $O(k^{1+1/t} + n)$ edges. The drawback is that this is a metric spanner, not a subgraph of the input graph. We alleviate this issue by constructing a graph spanner with the same stretch and $O(\sqrt{n} \cdot k^{1+1/t} + n)$ edges.[2] A result of [41] implicitly provides a terminal graph spanner with $(2t-1)$ stretch and $O(t \cdot n \cdot k^{1/t})$ edges. Our graph terminal spanner is sparser than that of [41] as long as $k \leq t \cdot n^{1/2(1+1/t)}$.

1.1 Algorithmic Applications

We overview a few of the applications of our results to approximation and online algorithms. Some of the most striking applications of metric embeddings are to various cut problems, such as the sparsest-cut, min-bisection, and also to several online problems. Our method provides improved guarantees when the input graph has a small set of "important" vertices. Specifically, these vertices can be considered as terminals, and we obtain approximation factors that depend on the cardinality of the terminal set, rather than on the input size. The exact meaning of importance is problem specific; e.g. in the cut problems, we require that the set of important vertices touches every demand pair, or every edge (that is, forms a vertex cover).

For instance, consider the (general) *sparsest-cut* problem [34, 10, 35]. We are given a graph $G = (V, E)$ with capacities on the edges $c : E \to \mathbb{R}_+$, and a collection of pairs $(s_1, t_1), \ldots, (s_r, t_r)$ along with their demands $D_1, \ldots, D_r$. The goal is to find a cut $S \subseteq V$ that minimizes the ratio between capacity and demand across the cut:

$$\phi(S) = \frac{\sum_{\{u,v\}\in E} c(u,v)|\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\sum_{i=1}^{r} D_i|\mathbf{1}_S(s_i) - \mathbf{1}_S(t_i)|} \ ,$$

where $\mathbf{1}_S(\cdot)$ is the indicator for membership in S. Following the breakthrough result of [9], which showed $O(\sqrt{\log n})$ approximation for the uniform demand case, [7] devised an $\tilde{O}(\sqrt{\log r})$ approximation for the general case. If there is a set of k important vertices, such that every demand pair contains an important vertex, we obtain an $\tilde{O}(\sqrt{\log k})$ approximation using the terminal embedding of negative-type metrics to ℓ_1. Observe that $k \leq r$, and so our result subsumes the result of [7]. Our bound is particularly useful for instances with many demand pairs but few distinct sources s_i (or few targets t_i).

We also consider other cut problems, and show a similar phenomenon: the $O(\log n)$ approximation for the *min-bisection* problem can be improved to an approximation of only $O(\log k)$, where k is the size of the minimum vertex cover of the input graph. For this result we employ our terminal variant of Räcke's result [40] on capacity-preserving probabilistic embedding into trees.

We then focus on one application of probabilistic embedding into ultrametrics [12, 23], and illustrate the usefulness of our terminal embedding result by the (online) *constrained file migration* problem [13]. Given a graph $G = (V, E)$ representing a network, each node

[2] Note that the number of edges is linear whenever $k \leq n^{1/(2(1+1/t))}$.

$v \in V$ has a memory capacity m_v, and there is a set of files that reside at the nodes, at most m_v files may be stored at node v at any given time. The cost of accessing a file is the distance in the graph to the node storing it (no copies are allowed). Files can also be migrated from one node to another. This costs D times the distance, for a given parameter $D \geq 1$. When a sequence of file requests from nodes arrives online, the goal is to minimize the cost of serving all requests. [12] showed a algorithm with $O(\log m \cdot \log n)$ competitive ratio for graphs on n nodes, where $m = \sum_{v \in V} m_v$ is the total memory available.[3] A setting which seems particularly natural is one where there is a small set of nodes who can store files (servers), and the rest of the nodes can only access files but not store them (clients). We employ our probabilistic terminal embedding into ultrametrics to provide a $O(\log m \cdot \log k)$ competitive ratio, for the case where there are k servers. (Note that this ratio is independent of n.)

1.2 Overview of Techniques

The weak variant of our terminal embedding into ℓ_2 maps every terminal x into its image $f(x)$ under an original black-box (e.g., Bourgain's) embedding of K into ℓ_2. This embedding is then appended with one additional coordinate. Terminals are assigned 0 value in this coordinate, while each non-terminal point y is mapped to $(f(x), d(x, y))$, where x is the closest terminal to y. It is not hard to see that this embedding guarantees terminal distortion $O(\gamma(k))$, where $\gamma(k)$ is the distortion of the original black-box embedding, i.e., $O(\log k)$ in the case of Bourgain's embedding. On the other hand, the new embedding employs only $\beta(k) + 1$ dimensions, where $\beta(k)$ is the dimension of the original blackbox embedding (i.e., $O(\log^2 k)$ in the case of Bourgain's embedding). [4] This idea easily generalizes to a number of quite general scenarios, and under mild assumptions (see Theorem 3) it can be modified to produce strong terminal embeddings.

This framework, however, does not apply in many important settings, such as embedding into subgraphs, and does not provide strong terminal guarantees in others. Therefore we devise embeddings tailored to each particular setting in a non-black-box manner. For instance, our probabilistic embedding into trees with strong terminal congestion requires an adaptation of a theorem of [6], about the equivalence of distance-preserving and capacity-preserving random tree embeddings, to the terminal setting. Perhaps the most technically involved is our probabilistic embedding into spanning trees with strong terminal distortion. This result requires a set of modifications to the recent algorithm of [3], which is based on a certain hierarchical decomposition of graphs. We adapt this algorithm by giving preference to the terminals in the decomposition (they are the first to be chosen as cluster centers), and the crux is assuring that the distortion of any pair containing a terminal is essentially not affected by choices made for non-terminals. Furthermore, one has to guarantee that each such pair can be separated in at most $O(\log k)$ levels of the hierarchy.

The basic technical idea that we use for constructing $(4t-1)$-terminal subgraph spanners with $O(\sqrt{n}k^{1+1/t} + n)$ edges is the following one. As was mentioned above, our general transformation constructs metric (i.e., non-subgraph) $(4t-1)$-terminal spanners with $O(n + k^{1+1/t})$ edges. The latter spanners employ some edges which do not belong to the original graph. We provide these edges as an input to a pairwise preserver. A pairwise preserver

[3] The original paper shows $O(\log m \cdot \log^2 n)$, the improved factor is obtained by using the embedding of [23].

[4] We can also get dimension $O(\log k)$ for terminal embeddings into ℓ_2 by replacing Bourgain's embedding with that of [2].

[18] is a sparse subgraph that preserves exactly all distances between a designated set of vertex pairs. We use these preservers to fill in the gaps in the non-subgraph terminal spanner constructed via our general transformation. As a result we obtain a subgraph terminal spanner which outperforms previously existing terminal spanners of [41] in a wide range of parameters.

1.3 Related Work

Already in the pioneering work of [35], an embedding that has to provide a distortion guarantee for a subset of the pairs is presented. Specifically, in the context of the sparsest-cut problem, [35] devised a non-expansive embedding of an arbitrary metric into ℓ_1, with distortion at most $O(\log r)$ for a set of r specified demand pairs.

Terminal distance oracles were studied by [41], who called them *source restricted* distance oracles. In their paper, [41] show $(2t-1)$-terminal stretch using $O(t \cdot n \cdot k^{1/t})$ space. Implicit in our companion paper [20] is a distance oracle with $(4t-1)$-terminal stretch, $O(t \cdot k^{1/t} + n)$ space and $O(1)$ query time. Terminal spanners with additive stretch for unweighted graphs were recently constructed in [31]. Specifically, they showed a spanner with $\tilde{O}(n^{5/4} \cdot k^{1/4})$ edges and additive stretch 2 for pairs containing a terminal. Another line of work introduced *distance preservers* [18]; these are spanners which preserve exactly distances for a given collection of pairs.

In the context of preserving distances just between the terminals, [26, 16, 22, 30] studied embeddings of a graph into a minor over the terminals, while approximately preserving distances. In their work on the requirement cut problem, among other results, [27] obtain for any metric with k specified terminals, a distribution over trees with expected expansion $O(\log k)$ for all pairs, and which is non-contractive for terminal pairs. (Note that this is different from our setting, as the extra guarantee holds for terminals only, not for pairs containing a terminal.)

Another line of research [37, 17, 22] studied cut and vertex sparsifiers. A *cut sparisifier* of a graph $G = (V, E)$ with respect to a subset K of terminals is a graph $H = (K, E_H)$ on just the set of terminals, so that for any subset $A \subset K$, the minimum value of a cut in G that separates A from $K \setminus A$ is approximately equal to the value of the cut $(A, K \setminus A)$ in H. Note that this notion is substantially different from the notion of terminal congestion-preserving embedding, which we study in the current paper.

In a companion paper [20], we study prioritized metric structures and embeddings. In that setting, along with the input metric (X, d), a priority ranking of the points of X is given, and the goal is to obtain a data structure (distance oracle, routing scheme) or an embedding with stretch/distortion that depends on the ranking of the points. This has some implications to the terminal setting, since the k terminals can be given as the first k points in the priority ranking. More concretely, implicit in [20] is an embedding into a single (non-subgraph) tree with strong terminal distortion $O(k)$, a probabilistic embedding into ultrametrics with expected strong terminal distortion $O(\log k)$, and embedding into ℓ_p space with strong terminal distortion $\tilde{O}(\log k)$. In the current paper we provide stronger and more general results: our single tree embedding has tight $2k-1$ stretch, the tree is a subgraph, and it can have low weight as well (at the expense of slightly increased stretch); we obtain probabilistic embedding into *spanning* trees. and in congestion-preserving trees; and our terminal embedding to ℓ_p space has a tight strong terminal distortion $(O(\log k), O(\log n))$ and low dimension. Furthermore, the results of this paper apply to numerous other settings (e.g., embeddings tailored for graphs excluding a fixed minor, negative-type metrics, spanners, etc.).

1.4 Organization

The general transformations are presented in Section 3, and the results on graph spanners appear in Section 4. The tradeoff between terminal distortion and lightness in a single tree embedding is shown in Section 5 (corresponding lower bounds in several settings are deferred to the full version). The probabilistic congestion preserving embedding into trees appears in Section 6 (see [20] for the distortion version). Finally, in Section 7 we describe some algorithmic applications of terminal embeddings.

In the full version of the paper we present our probabilistic embedding into spanning trees with strong terminal distortion, and an embedding into a single tree (ultrametric or a spanning tree) with constant average terminal distortion.

2 Preliminaries

Here we provide formal definitions for the notions of terminal distortion. Let (X, d_X) be a finite metric space, with $K \subseteq X$ a set of terminals. Throughout the paper we assume $|K| \leq |X|/2$.

▶ **Definition 1.** Let (X, d_X) be a metric space, and let $K \subseteq X$ be a subset of terminals. For a target metric (Y, d_Y), an embedding $f : X \to Y$ has *terminal distortion* α if there exists $c > 0$, such that for all $v \in K$ and $u \in X$,[5]

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

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

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

For a graph $G = (V, E)$ with a terminal set $K \subseteq V$, an *α-terminal (metric) spanner* is a graph H on V such that for all $v \in K$ and $u \in V$,

$$d_G(u, v) \leq d_H(u, v) \leq \alpha \cdot d_G(u, v) . \tag{1}$$

H is a *graph* spanner if it is a subgraph of G.

Denote by $\mathrm{diam}(X) = \max_{y,z \in X}\{d_X(y, z)\}$. For any $x \in X$ and $r \geq 0$ let $B_X(x, r) = \{y \in X \mid d_X(x, y) \leq r\}$ (we often omit the subscript when the metric is clear from context). For a point $x \in X$ and a subset $A \subseteq X$, $d_X(x, A) = \min_{a \in A}\{d_X(x, a)\}$. For $K \subseteq X$ we denote by (K, d_K) the metric space where d_K is the induced metric.

3 A General Transformation

In this section we present general transformation theorems that create terminal embeddings into normed spaces and graph families from standard ones. We say that a family of graphs $\mathcal{G}$ is *leaf-closed*, if it is closed under adding leaves. That is, for any $G \in \mathcal{G}$ and $v \in V(G)$, the graph G' obtained by adding a new vertex u and connecting u to v by an edge, belongs to $\mathcal{G}$. Note that many natural families of graphs are leaf-closed, e.g. trees, planar graphs, minor-free graphs, bounded tree-width graphs, bipartite graphs, general graphs, and many others.

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

▶ **Theorem 2.** *Let $\mathcal{X}$ be a family of metric spaces. Fix some $(X, d_X) \in \mathcal{X}$, and let $K \subseteq X$ be a set of terminals of size $|K| = k$, such that $(K, d_K) \in \mathcal{X}$. Then the following assertions hold:*

- *If there are functions $\alpha, \gamma : \mathbb{N} \to \mathbb{R}$, such that every $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ embeds into $\ell_p^{\gamma(m)}$ with distortion $\alpha(m)$, then there is an embedding of X into $\ell_p^{\gamma(k)+1}$ with* terminal *distortion $2^{(p-1)/p} \cdot ((2\alpha(k))^p + 1)^{1/p}$.*[6]
- *If $\mathcal{G}$ is a leaf-closed family of graphs, and any $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ embeds into $\mathcal{G}$ with distortion $\alpha(m)$ such that the target graph has at most $\gamma(m)$ edges, then there is an embedding of X into $\mathcal{G}$ with* terminal *distortion $2\alpha(k) + 1$ and the target graph has at most $\gamma(k) + n - k$ edges.*

▶ Remark. The second assertion holds under probabilistic embeddings as well.

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

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

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

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

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

For the second assertion, there is a non-contractive embedding f of K into $G \in \mathcal{G}$ with distortion at most $\alpha(k)$. As above, for each $x \in X \setminus K$ define k_x as the nearest point in K to x. And for each $x \in X$, add to G a new vertex $f(x)$ that is connected by an edge of length $d_G(x, k_x)$ to $f(k_x)$. The resulting graph $G' \in \mathcal{G}$, because it is a leaf-closed family. Fix any $x \in X$ and $t \in K$, then as above $d(t, k_x) \le 2d(t, x)$, and so

$$\begin{aligned}
d_{G'}(f(t), f(x)) &= d_G(f(t), f(k_x)) + d_{G'}(f(x), f(k_x)) \\
&\le \alpha(k) \cdot d(t, k_x) + d(x, k_x) \\
&\le d(t, x) \cdot (2\alpha(k) + 1) \ .
\end{aligned}$$

Also note that

$$d_{G'}(f(t), f(x)) = d_G(f(t), f(k_x)) + d(x, k_x) \ge d(t, k_x) + d(x, k_x) \ge d(t, x) \ ,$$

[6] Note that for any $p, \alpha \ge 1$ we have that $2^{(p-1)/p} \cdot ((2\alpha)^p + 1)^{1/p} \le 4\alpha$.

so the terminal distortion is indeed $2\alpha(k)+1$. Since f embeds into a graph with $\gamma(k)$ edges, and we added $n-k$ new edges, the total number of edges is bounded accordingly, which concludes the proof. ◀

Next, we study strong terminal embeddings into normed spaces. Fix any metric (X,d), a set of terminals $K \subseteq X$ and $1 \le p \le \infty$. Let $f : K \to \ell_p$ be a non-expansive embedding. We say that f is *Lipschitz extendable*, if there exists a non-expansive $\hat{f} : X \to \ell_p$ which is an extension of f (that is, the restriction of $\hat{f}$ to K is exactly f). It is not hard to verify that any Fréchet embedding[7] is Lipschitz extendable. For example, the embeddings of [15, 33, 8] are Fréchet.

▶ **Theorem 3.** *Let $\mathcal{X}$ be a family of metric spaces. Fix some $(X,d_X) \in \mathcal{X}$, and let $K \subseteq X$ be a set of terminals of size $|K| = k$, such that $(K, d_K) \in \mathcal{X}$. If any $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ embeds into $\ell_p^{\gamma(m)}$ with distortion $\alpha(m)$ by a* Lipschitz extendable map*, then there is a (non-expansive) embedding of X into $\ell_p^{\gamma(n)+\gamma(k)+1}$ with* strong *terminal distortion $O(\alpha(k), \alpha(n))$.*

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

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

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

so F has expansion at most $3^{1/p}$ for all pairs (which can easily be made 1 without affecting the distortion by more than a constant factor). Also note that

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

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

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

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

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

[7] In our context, it will be convenient to call an embedding $f : K \to \ell_p^t$ *Fréchet*, if there are sets $A_1, \ldots, A_t \subseteq X$ such that for all $i \in [t]$, $f_i(x) = \frac{d(x,A_i)}{t^{1/p}}$.

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

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

◀

▶ Remark. The results of Theorems 2 and 3 hold also if $\mathcal{X}$ is a family of graphs, rather than of metrics, provided that the embedding for this family has the promised guarantees even for graphs with Steiner nodes. (E.g., if $\hat{Z} \in \mathcal{X}$ is a graph and Z is a set of vertices of size m, then there exists a (Lipschitz extendable) embedding of (Z, d_Z) to $\ell_p^{\gamma(m)}$ with distortion $\alpha(m)$, where d_Z is the shortest path metric on $\hat{Z}$ induced on Z.) We note that many embeddings of graph families satisfy this condition, e.g. the embedding of [33] to planar and minor-free graphs.[8]

Here are some of the implications of Theorems 2 and 3.

▶ **Corollary 4.** *Let (X, d) be a metric space on n points, and $K \subseteq X$ a set of terminals of size $|K| = k$. Then for any $1 \leq p \leq \infty$,*

1. *(X, d) can be embedded to $\ell_p^{O(\log k)}$ with terminal distortion $O(\log k)$.*
2. *If (X, d) is an ℓ_2 metric, it can be embedded to $\ell_2^{O(\log k)}$ with terminal distortion $O(1)$.*
3. *For any $t \geq 1$ there exists a $(4t-1)$-terminal (metric) spanner of X with at most $O(k^{1+1/t}) + n$ edges.*
4. *If (X, d) is an ℓ_2 metric, for any $t \geq 1$ there exists a $O(t)$-terminal spanner of X with at most $O(k^{1+1/t^2}) + n$ edges.*
5. *(X, d) can be embedded to $\ell_p^{O(\log n + \log^2 k)}$ with strong terminal distortion $(O(\log k), O(\log n))$.*
6. *If (X, d) is a shortest-paths metric of a graph that excludes a fixed minor (e.g., a planar metric), it can be embedded to ℓ_p with strong terminal distortion $(O((\log k)^{\min\{1/2,1/p\}}), O((\log n)^{\min\{1/2,1/p\}}))$.*
7. *If (X, d) is a negative type metric, it can be embedded to ℓ_2 with strong terminal distortion $(\tilde{O}(\sqrt{\log k}), \tilde{O}(\sqrt{\log n}))$.*

The first two items use the first assertion of Theorem 2, the next two use its second assertion, and the last three apply Theorem 3. The corollary follows from known embedding results: (1) and (5) are from [15], with improved dimension due to [2], (2) is from [29], (3) is from [5] and (4) from [28], (6) from [33], and (7) from [8, 7].

[8] We remark that this requirement is needed for those graph families for which the following question is open: given a graph Z in the family with terminals K, is there another graph in the family over the vertex set K, that preserves the shortest-path distances with respect to Z (up to some constant). This question is open, e.g., for planar metrics.

4 Graph Terminal Spanners

While Theorem 2 provides a general approach to obtain terminal spanners, it cannot provide spanners which are subgraphs of the input graph. We devise a construction of such terminal spanners in this section, while somewhat increasing the number of edges. Specifically, we show the following.

▶ **Theorem 5.** *For any parameter $t \geq 1$, a graph $G = (V, E)$ on n vertices, and a set of terminals $K \subseteq V$ of size k, there exists a $(4t-1)$-terminal* graph *spanner with at most $O(n + \sqrt{n} \cdot k^{1+1/t})$ edges.*

▶ Remark. Note that the number of edges is linear in n whenever $k \leq n^{1/(2(1+1/t))}$.

We shall use the following result:

▶ **Theorem 6** ([18]). *Given a weighted graph $G = (V, E)$ on n vertices and a set $P \subseteq \binom{V}{2}$ of size p, then there exists a subgraph G' with $O(n + \sqrt{n} \cdot p)$ edges, such that for all $\{u, v\} \in P$, $d_G(u, v) = d_{G'}(u, v)$.*

Proof of Theorem 5. The construction of the subgraph spanner with terminal stretch will be as follows. Consider the metric induced on the terminals K by the shortest path metric on G. Create a $(2t-1)$ (metric) spanner H' of this metric, using [5], and let $P \subseteq \binom{K}{2}$ be the set of edges of H'. Note that $p = |P| \leq O(k^{1+1/t})$. Now, apply Theorem 6 on the graph G with the set of pairs P, and obtain a graph G' that for every $\{u, v\} \in P$, has $d_{G'}(u, v) = d_G(u, v)$. This implies that G' is a $(2t-1)$-spanner for each pair of vertices $u, w \in K$. Moreover, G' has at most $O(n + \sqrt{n} \cdot p)$ edges. Finally, create H out of G' by adding a shortest path tree in G with the set K as its root. This will guarantee that the spanner H will have for each non-terminal, a shortest path to its closest terminal in G. This concludes the construction of H, and now we turn to bounding the stretch. Since H is a subgraph clearly it is non-contracting. Fix any $v \in K$ and $u \in V$, let k_u be the closest terminal to u, then $d_G(k_u, v) \leq d_G(k_u, u) + d_G(u, v) \leq 2d_G(u, v)$, and thus

$$d_H(u, v) \leq d_H(u, k_u) + d_{G'}(k_u, v) \leq d_G(u, v) + (2t-1)d_G(k_u, v) \leq (4t-1)d_G(u, v) .$$

Finally observe that the total number of edges in H is at most $O(n + \sqrt{n} \cdot p) = O(n + \sqrt{n} \cdot k^{1+1/t})$. ◀

5 Light Terminal Trees for General Graphs

In this section we find a single spanning tree of a given graph, that has both light weight, and approximately preserves distances from a set of specified terminals. Theorem 2 can provide a tree with terminal distortion $2k-1$ (using that any graph has a tree with distortion $n-1$), but that tree may not be a subgraph and may have large weight.

For a weighted graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$, given a subgraph H of G, let $w(H) = \sum_{e \in E(H)} w(e)$, and define the *lightness* of H to be $\Psi(H) = \frac{w(H)}{w(MST(G))}$, where $w(MST)$ is the weight of a minimum spanning tree of G. The result of this section is summarized as follows.

▶ **Theorem 7.** *For any parameter $\alpha \geq 1$, given a weighted graph $G = (V, E, w)$, and a subset of terminals $K \subseteq V$ of size k, there exists a spanning tree T of G with terminal distortion $k \cdot \alpha + (k-1)\alpha^2$ and lightness $2\alpha + 1 + \frac{2}{\alpha - 1}$.*

When substituting $\alpha = 1$ in Theorem 7 we obtain a single tree with terminal distortion exactly $2k-1$, which is optimal. More specifically, for small $\epsilon > 0$, we get terminal distortion $2k-1+\epsilon$ and lightness $3+\frac{6k}{\epsilon}$. Also, note that the bound $2\alpha+1+\frac{2}{\alpha-1}$ is minimized by setting $\alpha = 2$, so there is no point in using the theorem with $\alpha > 2$.

Next we describe the algorithm for constructing a spanning tree that satisfies the assertion of Theorem 7.

A spanning tree T is an (α, β)-SLT with respect to a root $v \in V$, if for all $u \in V$, $d_T(v,u) \le \alpha \cdot d_G(v,u)$, and T has lightness β. A small modification of an SLT-constructing algorithm produces for any subset $K \subset V$, a forest F, such that every component of F contains exactly one vertex of K.[9] The forest F has distortion α with respect to K, and *lightness* $1+\frac{2}{\alpha-1}$. (Such a forest F is said to have distortion α with respect to K, if for every vertex $u \in V$, $d_F(K,u) \le \alpha \cdot d_G(K,u)$.)

The algorithm starts by building the aforementioned SLT-forest F from the terminal set K. No two terminals belong to the same connected component of F. Denote $K = \{v_1, \dots, v_k\}$, let V_i be the unique connected component of F containing v_i, and let $T_i \subseteq F$ be the edges of the forest F induced by V_i. It follows that for every $u \in V_i$, $d_F(K,u) = d_{T_i}(v_i,u) \le \alpha \cdot d_G(K,u)$. Let $G' = (K, E', w')$ be the super-graph in which two terminals share an edge between them if and only if there is an edge between the components V_i to V_j in G. Formally, $E' = \{\{v_i, v_j\} : \exists u_i \in V_i, u_j \in V_j \text{ such that } \{u_i, u_j\} \in E\}$. The weight $w'(v_i, v_j)$ is defined to be the length of the shortest path between v_i and v_j which uses *exactly one edge* that does not belong to F. (In other words, among all the paths between v_i and v_j in G which use exactly one edge that does not belong to F, let P be the shortest one. Then $w'(v_i, v_j) = w(P)$.) Note also that $w'(v_i, v_j)$ is given by $w'(v_i, v_j) = \min_{e \in E}\left\{d_{F\cup\{e\}}(v_i, v_j)\right\}$. We call the edge $e_{i,j} = \{u_i, u_j\}$ that implements this minimum $(w'(v_i, v_j) = d_{F\cup\{e_{i,j}\}}(v_i, v_j))$ the *representative edge* of $\{v_i, v_j\}$. (W.l.o.g the shortest paths, and thus the representative edges, are unique.) Observe that $\{v_i, v_j\} \in E'$ implies that $w'(v_i, v_j) < \infty$. Let T' be the MST of G'. Define $R = \left\{e_{i,j} \mid e_{i,j} \text{ is the representative edge of } e'_{i,j} = (v_i, v_j) \in T'\right\}$. Finally, set $T = F \cup R = \bigcup_{i=1}^{k} T_i \cup R$. Obviously, T is a spanning tree of G.

5.1 Proof of Theorem 7

As an embedding of a graph into its spanning tree is non-contractive, the tree T will have terminal distortion α if for all $v \in K$, $u \in V$, $d_T(v,u) \le \alpha \cdot d_G(v,u)$. We shall assume w.l.o.g that all edge weights are different, and every two different paths have different lengths. If it is not the case, then one can break ties in an arbitrary (but consistent) way.

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

▶ **Lemma 8.** *[The bottleneck lemma:] For every $v_i, v_j \in K$, there exists a path $P : v_i = z_0, z_1, ..., z_r = v_j$ in G' such that for every $s = 0, 1, \dots, r-1$, it holds that $\{z_s, z_{s+1}\} \in E'$ and $w'(z_s, z_{s+1}) \le \alpha \cdot d_G(v_i, v_j)$.*

Proof. Let $P_{i,j} : v_i = u_0, u_1, ..., u_s = v_j$ be the shortest path from v_i to v_j in G, i.e., $w(P_{i,j}) = d_G(v_i, v_j)$. For each $0 \le a \le s$, denote by $V^{(a)}$ the connected component of F that contains u_a, and let $v^{(a)}$ be the unique terminal in that component. Consider the path

[9] To obtain such a forest F, one should add a new vertex v_K to the graph and connect it to each of the vertices of K with edges of weight zero. Then we compute an (α, β)-SLT with respect to v_K in the modified graph. Finally, we remove v_K from the SLT. The resulting forest is F.

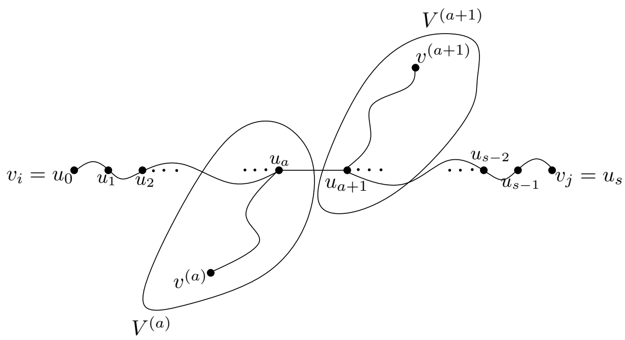

Figure 1 An illustration for the bottleneck lemma: v_i and v_j are terminals. The edge $\{u_a, u_{a+1}\}$ belongs to the shortest path from v_i to v_j in G. We conclude that for terminals $v^{(a)}, v^{(a+1)}$ such that $u_a \in V^{(a)}$ and $u_{a+1} \in V^{(a+1)}$ it holds that $w'\left(v^{(a)}, v^{(a+1)}\right)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

$P = v^{(0)}, v^{(1)}, ..., v^{(s)}$. (This path is not necessarily simple. In particular, it might contain self-loops.) For every index $a < s$, (see Figure 1 for illustration)

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

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

In particular, one can remove cycles from P and obtain a simple path with the desired properties. We get a simple path P' such that for every edge v, v' on this path, we have $w'(v, v') \leq \alpha \cdot d_G(v_i, v_j)$, as required. ◀

The following is a simple corollary.

▶ **Corollary 9.** *For $\{v_i, v_j\} \in T'$, we have $w'(v_i, v_j) = d_T(v_i, v_j) \leq \alpha \cdot d_G(v_i, v_j)$.*

Proof. By Lemma 8, $w'(v_i, v_j) \leq \alpha \cdot d_G(v_i, v_j)$. (Indeed, otherwise the edge $\{v_i, v_j\}$ is strictly the heaviest edge in a cycle in G', contradiction to the assumption that it belongs to the MST of G'.) Since $\{v_i, v_j\} \in E'$ and the representative edge of $\{v_i, v_j\}$ was taken into T, it follows that $w'(v_i, v_j) = d_T(v_i, v_j)$. ◀

We conclude the following lemma, which bounds the stretch of terminal pairs.

▶ **Lemma 10.** *For $v_i, v_j \in K$, we have $d_T(v_i, v_j) \leq d_{T'}(v_i, v_j) \leq \alpha \cdot (k-1) \cdot d_G(v_i, v_j)$.*

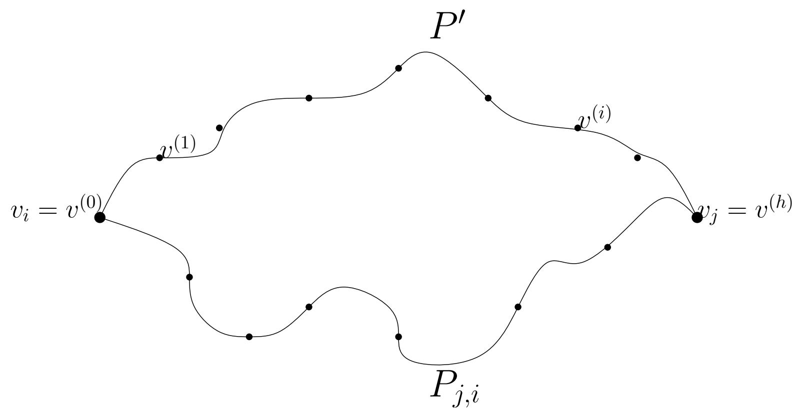

Figure 2 The two paths P' and $P_{j,i}$ considered in the proof of Lemma 10. The path P' is contained in T', while all edges of $P_{j,i}$ have weight at most $\alpha \cdot d_G(v_i, v_j)$.

Proof. Let $P' : v_i = v^{(0)}v^{(1)} \dots v^{(h)} = v_j$ be the (unique) path in T' between v_i and v_j. Since T' is a spanning tree of the k-vertex graph G', it follows that $h \le k-1$. Observe also that for every index $a \in [h-1]$, by Corollary 9 the edge $w'\left(v^{(a)}, v^{(a+1)}\right) = d_T\left(v^{(a)}, v^{(a+1)}\right)$. Also, we next argue that $w'\left(v^{(a)}, v^{(a+1)}\right) \le \alpha \cdot d_G(v_i, v_j)$. Indeed, suppose for contradiction that $w'\left(v^{(a)}, v^{(a+1)}\right) > \alpha \cdot d_G(v_i, v_j)$. Let $P_{j,i}$ be a path between v_j and v_i in G' such that all its edges have weight at most $\alpha \cdot d_G(v_i, v_j)$. The existence of this path is guaranteed by Lemma 8. In particular, since $w'\left(v^{(a)}, v^{(a+1)}\right) > \alpha \cdot d_G(v_i, v_j)$, it follows that $\left\{v^{(a)}, v^{(a+1)}\right\} \notin P_{j,i}$. Consider the cycle $P' \circ P_{j,i}$ in G'. It is not necessarily a simple cycle, but since $\left\{v^{(a)}, v^{(a+1)}\right\} \notin P_{j,i}$, the edge $\left\{v^{(a)}, v^{(a+1)}\right\}$ belongs to a simple cycle C contained in $P' \circ P_{j,i}$. The heaviest edge of C clearly does not belong to $P_{j,i}$, because the edge $\left\{v^{(a)}, v^{(a+1)}\right\}$ is heavier than each of them. Hence the heaviest edge belongs to P', but $P' \subseteq T'$. This is a contradiction to the assumption that T' is an MST of G'. (See Figure 2 for an illustration). Hence $d_T\left(v^{(a)}, v^{(a+1)}\right) = w'\left(v^{(a)}, v^{(a+1)}\right) \le \alpha \cdot d_G(v_i, v_j)$. Finally,

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

◀

Next, we analyze the terminal distortion of T.

▶ **Lemma 11.** *The terminal distortion of T is at most $k \cdot \alpha + (k-1)\,\alpha^2$.*

Proof. For each terminal $v_i \in K$ and any vertex $u \in V_j$, it holds that

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

The last inequality is because $d_G(v_j, u) \le d_F(v_j, u) = d_F(K, u) \le \alpha \cdot d_G(K, u) \le \alpha \cdot d_G(v_i, u)$. ◀

We now turn to analyze the lightness of T. A tree $T=(K,E',w')$ is called a *Steiner tree* for a graph $G=(V,E,w)$ if (1) $V\subseteq K$, (2) for any edge $e\in E\cap E'$, the edge has the same weight in both G and T, i.e. $w(e)=w'(e)$, and (3) for any pair of vertices $u,v\in V$ it holds that $d_T(u,v)\geq d_G(u,v)$. The *minimum Steiner tree* T of G, denoted $SMT(G)$, is a Steiner tree of G with minimum weight. It is well-known that for any graph G, $w(SMT(G))\geq\frac{1}{2}MST(G)$. (See, e.g., [25], Section 10.) The next lemma bounds the lightness of the tree T.

▶ **Lemma 12.** *The lightness of T is bounded by $\Psi(T)\leq 2\alpha+1+\frac{2}{\alpha-1}$.*

Proof. The main challenge is to bound $w(R)$. (Recall that R is the set of the *representative edges* of T'.) Consider an edge $\{v_i,v_j\}\in T'$, and let $\{u_i,u_j\}$ be its representative edge. Then $d_G(u_i,u_j)\leq w'(v_i,v_j)$. Also, since $\{v_i,v_j\}\in T'\subseteq E'$, it follows that $w'(v_i,v_j)=d_{G'}(v_i,v_j)$. Hence $d_G(u_i,u_j)\leq d_{G'}(v_i,v_j)$. Therefore, $w(R)\leq w'(T')$. Next we provide an upper bound for $w'(T')$. Define the graph $\widetilde{G}$ as the complete graph on the vertex set K, with weights $\tilde{w}$ induced by d_G (the shortest path distances in G). Let $\widetilde{T}$ be the MST of $\widetilde{G}$. We build a new tree $\hat{T}$ by the following process:

1. Let $\hat{T}\leftarrow\widetilde{T}$;
2. For each $\{v_i,v_j\}=\tilde{e}\in\tilde{T}$:
   a. Let $P_{\tilde{e}}$ be a path from v_i to v_j which consists of edges in E', such that for each edge e in $P_{\tilde{e}}$, $w'(e)\leq\alpha\cdot d_G(v_i,v_j)=\alpha\cdot\tilde{w}(\tilde{e})$; (By Lemma 8, such a path exists);
   b. Let $e'\in P_{\tilde{e}}$ be an edge such that $(\hat{T}\setminus\{\tilde{e}\})\cup\{e'\}$ is connected;
   c. Set $\hat{T}\leftarrow(\hat{T}\setminus\{\tilde{e}\})\cup\{e'\}$;

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

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

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

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

◀

6 Probabilistic Embedding into Trees with Terminal Congestion

In this section we focus on embeddings into trees that approximate capacities of cuts, rather than distances between vertices. This framework was introduced by Räcke [39] (for a single tree), and in [40] he showed how to obtain capacity preserving probabilistic embedding from a distance preserving one, such as the ones given by [23]. Later, [6] showed a complete equivalence between these notions in random tree embeddings. Here we show our terminal variant of these results. Informally, we construct a distribution over *capacity-dominating* trees (each cut in each tree is at least as large as the corresponding cut in the original graph), and for each edge, its expected congestion is bounded accordingly, with an improved bound for edges *containing a terminal*.

Recall that an ultrametric (U,d) is a metric space satisfying a strong form of the triangle inequality, that is, for all $x,y,z\in U$, $d(x,z)\le\max\{d(x,y),d(y,z)\}$. The following definition is known to be an equivalent one (see [14]).

▶ **Definition 13.** An ultrametric U is a metric space (U,d) whose elements are the leaves of a rooted labeled tree T. Each $z\in T$ is associated with a label $\Phi(z)\ge 0$ such that if $q\in T$ is a descendant of z then $\Phi(q)\le\Phi(z)$ and $\Phi(q)=0$ iff q is a leaf. The distance between leaves $z,q\in U$ is defined as $d_T(z,q)=\Phi(\mathrm{lca}(z,q))$ where $\mathrm{lca}(z,q)$ is the least common ancestor of z and q in T.

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

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

The notion of strong terminal distortion translates to this setting in the obvious manner.

We will need the following theorem, implicit in our companion paper [20].

▶ **Theorem 14.** *[20] Given a metric space (X,d) of size $|X|=n$ and a subset of terminals $K\subseteq X$ of size $|K|=k$, there exists a distribution over embeddings of X into ultrametrics with strong terminal distortion $(O(\log k),O(\log n))$.*

We next elaborate on the notions of capacity and congestion, and their relation to distance and distortion, following the notation of [6]. Given a graph $G=(V,E)$, let $\mathcal{P}$ be a collection of multisets of edges in G. A map $M:E\to\mathcal{P}$, where $M(e)$ is a path between the endpoints of e, is called a *path mapping* (the path is not necessarily simple). Denote by $M_{e'}(e)$ the number of appearances of e' in $M(e)$.

The path mapping relevant to the rest of this section is constructed as follows: given a tree $T=(V,E_T)$ (not necessarily a subgraph), for each edge $e'\in E_T$ let $P_G(e')$ be a shortest path between the endpoints of e' in G (breaking ties arbitrarily), and similarly for $e\in E$, let $P_T(e)$ be the unique path between the endpoints of e in T. Then for an edge $e\in E$, where $P_T(e)=e'_1e'_2\dots e'_r$, the path $M(e)$ is defined as $M(e)=P_G(e'_1)\circ P_G(e'_2)\circ\cdots\circ P_G(e'_r)$ (where $\circ$ denotes concatenation). In what follows fix a tree T, and let M be the path mapping of T.

Fix a weight function $w:E\to\mathbb{R}_+$, and a capacity function $c:E\to\mathbb{R}_+$. For an edge $e\in E$, $\mathrm{dist}_T(e)=\sum_{e'\in E}M_{e'}(e)\cdot w(e')$ is the weight of the path $M(e)$, and $\mathrm{load}_T(e)=\sum_{e'\in E}M_e(e')\cdot c(e')$ is the sum (with multiplicities) of the capacities of all the edges whose path is using e. Define $\mathrm{distortion}_T(e)=\frac{\mathrm{dist}_{T(e)}}{w(e)}$ to be the distortion of e in T, and $\mathrm{cong}_T(e)=\frac{\mathrm{load}_{T(e)}}{c(e)}$ is the congestion of e. Note that if T is a subgraph of G, then $\mathrm{dist}_T(e)$ is the length of the unique path between the endpoints of e, while $\mathrm{load}_T(e)$ is the total capacity of all the edges of E that are in the cut obtained by deleting e from T (for $e\notin T$, $\mathrm{load}_T(e)=0$).

▶ **Definition 15.** Let $K\subseteq V$ be a set of terminals of size k, and let $E_K\subseteq E$ be the set of edges that contain a terminal. We say that a distribution $\mathcal{D}$ over trees has strong terminal congestion (α,β) if for every $e\in E_K$.

$$\mathrm{cong}_{\mathcal{D}}(e):=\mathbb{E}_{T\sim\mathcal{D}}[\mathrm{cong}_T(e)]\le\alpha\ ,$$

and for any $e\in E$, $\mathrm{cong}_{\mathcal{D}}(e)\le\beta$.

A tight connection between distance preserving and capacity preserving mappings was shown in [6]. We generalize their theorem to the terminal setting in the following manner.

▶ **Theorem 16.** *The following statements are equivalent for a graph G:*

- *For every possible* weight *assignment G admits a probabilistic embedding into trees with strong terminal* distortion (α, β).
- *For every possible* capacity *assignment G admits a probabilistic embedding into trees with strong terminal* congestion (α, β).

An immediate corollary of Theorem 16, achieved by applying Theorem 14, is:[10]

▶ **Corollary 17.** *For any graph $G = (V, E)$ on n vertices, a set $K \subseteq V$ of k terminals, and any capacity function, there exists a distribution over trees with strong terminal congestion $(O(\log k), O(\log n))$.*

Proof of Theorem 16. Assuming the first item holds we prove the second. Let $\kappa(e) = \begin{cases} 1/\alpha & e \in E_K \\ 1/\beta & \text{otherwise} \end{cases}$. Given any capacity function $c : E \to \mathbb{R}_+$, we would like to show that there exists a distribution $\mathcal{D}'$ such that for any $e \in E$, $\mathbb{E}_{T\sim\mathcal{D}'}[\kappa(e) \cdot \text{cong}_T(e)] \le 1$. By applying the minimax principle (as in [4]), it suffices to show that for any coefficients $\{\lambda_e\}_{e\in E}$ with $\lambda_e \ge 0$ and $\sum_{e\in E} \lambda_e = 1$, there exists a single tree T such that

$$\sum_{e\in E} \lambda_e \cdot \kappa(e) \cdot \text{cong}_T(e) \le 1 \ . \tag{3}$$

To this end, define the weights $w(e) = \kappa(e) \cdot \frac{\lambda_e}{c(e)}$, and by the first assertion there exists a distribution $\mathcal{D}$ over trees such that for any $e \in E$,

$$\mathbb{E}_{T\sim\mathcal{D}}[\kappa(e) \cdot \text{distortion}_T(e)] \le 1 \ .$$

By applying the minimax again, there exists a single tree T such that

$$\sum_{e\in E} \lambda_e \cdot \kappa(e) \cdot \text{distortion}_T(e) \le 1 \ .$$

Now,

$$\begin{aligned} 1 &\ge \sum_{e\in E} \lambda_e \cdot \kappa(e) \cdot \text{distortion}_T(e) \\ &= \sum_{e\in E} \lambda_e \cdot \kappa(e) \cdot \frac{\sum_{e'\in E} M_{e'}(e) \cdot w(e')}{w(e)} \\ &= \sum_{e\in E} \lambda_e \cdot \kappa(e) \cdot \frac{\sum_{e'\in E} M_{e'}(e) \cdot \kappa(e') \cdot \lambda_{e'}/c(e')}{\kappa(e) \cdot \lambda_e/c(e)} \\ &= \sum_{e'\in E} \lambda_{e'} \cdot \kappa(e') \cdot \frac{\sum_{e\in E} M_{e'}(e) \cdot c(e)}{c(e')} \\ &= \sum_{e'\in E} \lambda_{e'} \cdot \kappa(e') \cdot \text{cong}_T(e') \ , \end{aligned}$$

which concludes the proof of (3). The second direction is symmetric. ◀

[10] Even though the embedding of Theorem 14 is into ultrametrics, which contain Steiner vertices, these can be removed while increasing the distortion of each pair by at most a factor of 8 [26].

Capacity Domination Property. As [40, 6] showed, under the natural capacity assignment, any tree T supported by the distribution of Theorem 16 has the following property: Any multi-commodity flow in G can be routed in T with no larger congestion. We would like to show this explicitly, using the language of cuts, as this will be useful for the algorithmic applications.

Fix some tree $T = (V, E_T)$, and for any edge $e' \in E_T$ let $S_{T,e'} \subseteq V$ be the cut obtained by deleting e' from T. Define the capacities $C_T : E_T \to \mathbb{R}_+$ by

$$C_T(e') = \sum_{e \in E(S_{T,e'}, \bar{S}_{T,e'})} c(e) \ ,$$

where $E(S, \bar{S})$ denotes the set of edges in the graph crossing the cut S. (Observe that for spanning trees, $C_T(e) = \mathrm{load}_T(e)$.)

▶ **Lemma 18.** *For any graph $G = (V, E)$ and tree $T = (V, E_T)$ with capacities as defined above, for any set $S \subseteq V$ it holds that*

$$\sum_{e \in E\left(S, \bar{S}\right)} c(e) \le \sum_{e' \in E_T\left(S, \bar{S}\right)} C_T(e') \le \sum_{e \in E\left(S, \bar{S}\right)} \mathrm{load}_T(e) \ . \tag{4}$$

Proof. We begin with the left inequality. For any graph edge $e \in E(S, \bar{S})$, there exists a tree edge $e' \in E_T(S, \bar{S})$ such that $e' \in P_T(e)$, because the path $P_T(e)$ must cross the cut. Since removing e' from T separates the endpoints of e, $C_T(e')$ will contain the term $c(e)$. We conclude that

$$\sum_{e \in E\left(S, \bar{S}\right)} c(e) \le \sum_{e' \in E_T\left(S, \bar{S}\right)} C_T(e') \ .$$

For the right hand side, consider an edge $e \in E$, and note that for any tree edge $e' \in E_T$ such that $e \in P_G(e')$, every edge $e'' \in E(S_{T,e'}, \bar{S}_{T,e'})$ will have $e \in M(e'')$ and thus contribute to $\mathrm{load}_T(e)$ (perhaps multiple times, due to different e'). This implies that

$$\mathrm{load}_T(e) = \sum_{e' \in E_T \ : \ e \in P_G(e')} C_T(e') \ . \tag{5}$$

Next, observe that any tree edge $e' \in E_T(S, \bar{S})$ must have at least one graph edge $e \in E(S, \bar{S})$ such that $e \in P_G(e')$. This suggests that

$$\begin{aligned}
\sum_{e \in E(S, \bar{S})} \mathrm{load}_T(e) &\overset{(5)}{=} \sum_{e \in E(S, \bar{S})} \sum_{e' \in E_T : e \in P_G(e')} C_T(e') \\
&= \sum_{e' \in E_T} |E(S, \bar{S}) \cap P_G(e')| \cdot C_T(e') \\
&\ge \sum_{e' \in E_T(S, \bar{S})} C_T(e') \ .
\end{aligned}$$

◀

7 Applications

In this section we illustrate several algorithmic applications of our techniques. Some of our applications are suitable for graphs with a small vertex cover. Recall that for a graph $G = (V, E)$, a set $A \subseteq V$ is a vertex cover of G, if for any edge $e \in E$, at least one of its endpoints is in A. A polynomial time 2-approximation algorithm to this problem is folklore.

7.1 Sparsest-Cut

In the sparsest-cut problem we are given a graph $G = (V, E)$ with capacities on the edges $c : E \to \mathbb{R}_+$, and a collection of pairs $(s_1, t_1), \dots, (s_r, t_r)$ along with their demands $D_1, \dots, D_r$. The goal is to find a cut $S \subseteq V$ that minimizes the ratio between capacity and demand across the cut:

$$\phi(S) = \frac{\sum_{\{u,v\}\in E} c(u,v)|\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\sum_{i=1}^{r} D_i|\mathbf{1}_S(s_i) - \mathbf{1}_S(t_i)|} ,$$

where $\mathbf{1}_S(\cdot)$ is the indicator for membership in S. Arora et. al. [7] present an $\tilde{O}\left(\sqrt{\log r}\right)$ approximation algorithm to this problem. Our contribution is the following.

▶ **Theorem 19.** *If there exists a set $K \subseteq V$ of size k such that any demand pair contains a vertex of K, then there exists a $\tilde{O}\left(\sqrt{\log k}\right)$ approximation algorithm for the sparsest-cut problem.*

The key ingredient of the algorithm of [7] is a non-expansive embedding from ℓ_2^2 (negative-type metrics) into ℓ_1, which has $\tilde{O}\left(\sqrt{\log r}\right)$ contraction for all demand pairs. We will use the strong terminal embedding for negative type metrics given in item (7) of Corollary 4 to improve the distortion to $\tilde{O}\left(\sqrt{\log k}\right)$.

We now elaborate on how to use the embedding of ℓ_2^2 into ℓ_1 to obtain an approximation algorithm for the sparsest-cut, all the details can be found in [35, 9, 7], and we provide them just for completeness. First, write down the following SDP relaxation with triangle inequalities:

Algorithm 1 Sparsest Cut SDP Relaxation

min $\sum_{\{u,v\}\in E} c(u,v) \cdot \|\bar{u} - \bar{v}\|_2^2$
s.t. $\sum_{i=1}^{r} D_i \cdot \|\bar{s}_i - \bar{t}_i\|_2^2 = 1$
For all $u, v, w \in V$, $\|\bar{u} - \bar{v}\|_2^2 + \|\bar{v} - \bar{w}\|_2^2 \geq \|\bar{u} - \bar{w}\|_2^2$
For all $u \in V$, $\bar{u} \in \mathbb{R}^n$

Note that this is indeed a relaxation: if S is the optimal cut, set $\rho = \sum_{i=1}^{r} D_i \cdot |\mathbf{1}_S(s_i) - \mathbf{1}_S(t_i)|$; for $u \in S$ set $\bar{u} = (\frac{1}{\sqrt{\rho}}, 0, ..., 0)$, and for $u \notin S$, set $\bar{u} = (0, ..., 0)$. It can be checked to be a feasible solution of value equal to that of the cut S.

Let $K \subseteq V$ be a vertex cover of the demand graph $(V, \{\{s_i, t_i\}_{i=1}^r\})$ of size at most $2k$ (recall that we can find such a cover in polynomial time). Let $X = \{\bar{v} \in \mathbb{R}^n \mid v \in V\}$ be an optimal solution to the SDP (it can be computed in polynomial time), which is in particular an ℓ_2^2 (pseudo) metric. By Corollary 4 there exists a non-expansive embedding $f : X \to \ell_1$ with terminal distortion $\tilde{O}\left(\sqrt{\log k}\right)$ (where K is the terminal set).[11] This implies that for any $u, v \in V$ and any $1 \leq i \leq r$,

$$\begin{aligned} &\|\bar{u} - \bar{v}\|_2^2 \geq \|f(\bar{v}) - f(\bar{u})\|_1 \\ &\|\bar{s}_i - \bar{t}_i\|_2^2 \leq \tilde{O}(\sqrt{\log k}) \cdot \|f(\bar{s}_i) - f(\bar{t}_i)\|_1 . \end{aligned} \tag{6}$$

Let $\|f(\bar{v}) - f(\bar{u})\|_1 = \sum_{S\subseteq V} \alpha_S |\mathbf{1}_S(v) - \mathbf{1}_S(u)|$ be a representation of the ℓ_1 metric as a nonnegative linear combination of cut metrics (it is well known that there is such a

[11] The embedding of Corollary 4 is in fact into ℓ_2, but there is an efficient randomized algorithm to embed ℓ_2 into ℓ_1 with constant distortion [24].

representation with polynomially many cuts S having $\alpha_S > 0$). We conclude

$$\begin{aligned}
\mathrm{opt(SDP)} &= \sum_{\{u,v\}\in E} c(u,v)\cdot\|\bar{u}-\bar{v}\|_2^2 \\
&= \frac{\sum_{\{u,v\}\in E} c(u,v)\cdot\|\bar{u}-\bar{v}\|_2^2}{\sum_{i=1}^r D_i\cdot\|\bar{s}_i-\bar{t}_i\|_2^2} \\
&\overset{(6)}{\geq} \frac{\sum_{\{u,v\}\in E} c(u,v)\cdot\|f(\bar{v})-f(\bar{u})\|_1}{\sum_{i=1}^r D_i\cdot\tilde{O}\left(\sqrt{\log k}\right)\cdot\|f(\bar{s}_i)-f\left(\bar{t}_i\right)\|_1} \\
&= \frac{1}{\tilde{O}\left(\sqrt{\log k}\right)}\cdot\frac{\sum_{\{u,v\}\in E} c(u,v)\cdot\sum_{S\subsetneq V}\alpha_S\,|\mathbf{1}_S(v)-\mathbf{1}_S(u)|}{\sum_{i=1}^r D_i\cdot\sum_{S\subsetneq V}\alpha_S\,|\mathbf{1}_S(s_i)-\mathbf{1}_S(t_i)|} \\
&\geq \frac{1}{\tilde{O}\left(\sqrt{\log k}\right)}\min_{S:\alpha_S>0}\frac{\sum_{\{u,v\}\in E} c(u,v)\cdot|\mathbf{1}_S(v)-\mathbf{1}_S(u)|}{\sum_{i=1}^r D_i\cdot|\mathbf{1}_S(s_i)-\mathbf{1}_S(t_i)|} \\
&= \min_{S:\alpha_S>0}\frac{\phi(S)}{\tilde{O}\left(\sqrt{\log k}\right)}\ .
\end{aligned}$$

In particular, among the polynomially many sets $S\subseteq V$ with $\alpha_S>0$, there exists one which has sparsity at most $\tilde{O}(\sqrt{\log k})$ times larger than the optimal one.

7.2 Min Bisection

In the min-bisection problem, we are given a graph $G=(V,E)$ on an even number n of vertices, with capacities $c:E\to\mathbb{R}_+$. The purpose is to find a partition of V into two equal parts $S\subseteq V$ and $\bar{S}=V\setminus S$, that minimizes $\sum_{e\in E(S,\bar{S})} c(e)$. This problem is NP-hard, and the best known approximation is $O(\log n)$ by [40]. We obtain the following generalization.

▶ **Theorem 20.** *There exists a $O(\log k)$ approximation algorithm for min-bisection, where k is the size of a minimal vertex cover of the input graph.*

Proof. Our algorithm follows closely the algorithm of [40], the major difference is that we use our embedding into trees with terminal congestion. Let $K\subseteq V$ be the set of terminals, which is a vertex cover of size at most $2k$, and $\mathcal{D}$ a distribution over trees with strong terminal congestion $(O(\log k), O(\log n))$ given by Corollary 17. The algorithm will sample a tree $T=(V,E_T)$ from $\mathcal{D}$, find an optimal bisection in T and return it. We refer the reader to Section 6 for details on notation and on the definition of capacities $C_T:E_T\to\mathbb{R}_+$ for T. We note that there is polynomial time algorithm (by dynamic programming) to find a min-bisection in trees.

It remains to analyze the algorithm. Let $S\subseteq V$ be the optimal solution in G, and S_T be the optimal bisection for the tree T. The expected cost of using S_T in G can be bounded using Lemma 18 as follows

$$\begin{aligned}\sum_{T\in\mathrm{supp}(\mathcal{D})}\Pr[T]\sum_{e\in E\left(S_T,\bar{S}_T\right)}c(e) &\overset{(4)}{\leq} \sum_T\Pr[T]\sum_{e'\in E_T\left(S_T,\bar{S}_T\right)}C_T(e') \\ &\leq \sum_T\Pr[T]\sum_{e'\in E_T\left(S,\bar{S}\right)}\mathcal{C}_T(e') \\ &\overset{(4)}{\leq} \sum_T\Pr[T]\sum_{e\in E\left(S,\bar{S}\right)}\mathrm{load}_T(e) \\ &= \sum_{e\in E\left(S,\bar{S}\right)}\mathbb{E}_{T\sim\mathcal{D}}[\mathrm{load}_T(e)] \\ &\leq \sum_{e\in E\left(S,\bar{S}\right)}O(\log k)\cdot c(e) \\ &= O(\log k)\cdot\mathrm{opt}(G) \quad ,\end{aligned}$$

where the last inequality uses that every edge touches a terminal, so its expected congestion is $O(\log k)$. The algorithm can be derandomized using standard methods, see e.g. [6]. ◀

7.3 Online Algorithms: Constrained File Migration

We illustrate the usefulness of our probabilistic terminal embedding into ultrametric via the constrained file migration problem. This is an online problem, in which we are given a graph $G = (V, E)$ representing a network, each node $v \in V$ has a memory capacity m_v, and a parameter $D \geq 1$. There is some set of files that reside at the nodes, at most m_v files may be stored at node v in any given time. The cost of accessing a file that currently lies at v from node u is $d_G(u, v)$ (no copies of files are allowed). Files can also be migrated from one node to another, this costs D times the distance. When a sequence of file requests arrives online, the goal is to minimize the cost of serving all requests. The *competitive ratio* of an online algorithm is the maximal ratio between its cost to the cost of an optimal (offline) solution. For randomized algorithms the expected cost is used.

We consider the case where there exists a small set of vertices which are allowed to store files (i.e. $m_v > 0$). One may think about these vertices as servers who store files, while allowing file requests from all end users. Let $K \subseteq V$ be the set of terminal vertices that are allowed to store files, with $|K| = k$. Our result is captured by the following theorem.

▶ **Theorem 21.** *There is a randomized algorithm for the constrained file migration problem with competitive ratio $O(\log m \cdot \log k)$, where k vertices can store files and m is the total memory available.*

This theorem generalizes a result of [12], who showed an algorithm with competitive ratio $O(\log m \cdot \log n)$ for arbitrary graphs on n nodes. Both results are based on the following theorem. (Recall that a 2-HST is an ultrametric (see Definition 13) such that the ratio between the label of a node to any of its children's label is at least 2.)

▶ **Theorem 22** ([12])**.** *For any 2-HST, there is a randomized algorithm with competitive ratio $O(\log m)$ for constrained file migration with total memory m.*

By Theorem 14 there is a distribution $\mathcal{D}$ over embeddings of G into ultrametrics with expected terminal distortion $O(\log k)$, but in fact every tree in that distribution is a 2-HST.

Assume that in the optimal (offline) solution there are s_{uv} times a file residing on v was accessed by u, and t_{uv} files were migrated from v to u. Let $c_{uv} = s_{uv} + D \cdot t_{uv}$ be the total cost of file traffic from v to u. Note that as $m_v = 0$ for any $v \notin K$, then for any $u \in V$ we have $c_{uv} = 0$. Using the fact that the terminal distortion guarantee of $\mathcal{D}$ applies to all of the relevant distances, we obtain that

$$\begin{aligned} \mathrm{opt}_G &= \sum_{u\in V, v\in K} c_{uv} \cdot d_G(u,v) \\ &\geq \frac{1}{O(\log k)} \cdot \sum_{u\in V, v\in K} c_{uv} \cdot \mathbb{E}_{T\sim\mathcal{D}}[d_T(u,v)] \\ &= \frac{1}{O(\log k)} \cdot \mathbb{E}_{T\sim\mathcal{D}}\Big[\sum_{u\in V, v\in K} c_{uv} \cdot d_T(u,v)\Big] \ . \end{aligned} \tag{7}$$

Observe that for any tree $T \in \mathrm{supp}\,(\mathcal{D})$ we could have served the request sequence in the same manner as the optimal algorithm, which would have the cost $\sum_{u\in V, v\in K} c_{uv} \cdot d_T(u,v)$. In particular, the optimal solution opt_T for the same requests with the input graph T cannot be larger than that, i.e.

$$\sum_{u\in V, v\in K} c_{uv} \cdot d_T(u,v) \geq \mathrm{opt}_T \ . \tag{8}$$

Our algorithm will operate as follows: Pick a random tree according to the distribution $\mathcal{D}$, pick a random strategy S for transmitting files in T according to the distribution $\mathcal{S}(T)$ guaranteed to exists by Theorem 22, and serve the requests according to S. Denote by $\mathrm{cost}_H(S)$ the cost of applying strategy S with distances taken in the graph H. For any possible $T \in \mathrm{supp}\,(\mathcal{D})$ it holds that

$$\mathrm{opt}_T \geq \frac{\mathbb{E}_{S\sim\mathcal{S}(T)}[\mathrm{cost}_T\,(S)]}{O(\log m)} \geq \frac{\mathbb{E}_{S\sim\mathcal{S}(T)}[\mathrm{cost}_G\,(S)]}{O(\log m)}, \tag{9}$$

where the last inequality holds since T dominates G (i.e. $d_T(u,v) \geq d_G(u,v)$ for all $u,v \in V$). Combining equations (7), (8) and (9) we get that

$$\mathrm{opt}_G \geq \frac{\mathbb{E}_{T\sim\mathcal{D}}\mathbb{E}_{S\sim\mathcal{S}(T)}[\mathrm{cost}_G\,(S)]}{O(\log k \log m)} \ .$$

Hence our randomized algorithm has $O\left(\log m \log k\right)$ competitive ratio, as promised.

Acknowledgements. We would like to thank Robert Krauthgamer, Yair Bartal and Manor Mendel for fruitful discussions, and to an anonymous referee for an idea leading to Theorem 3.

References

1 Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *SODA*, pages 502–511, 2007.
2 Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026–3126, 2011.
3 Ittai Abraham and Ofer Neiman. Using petal-decompositions to build a low stretch spanning tree. In *STOC*, pages 395–406, 2012.
4 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 I. Althöfer, G. Das, D. P. Dobkin, D. Joseph, and J. Soares. On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 9:81–100, 1993.

6 Reid Andersen and Uriel Feige. Interchanging distance and capacity in probabilistic mappings, 2009.

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

8 Sanjeev Arora, James R. Lee, and Assaf Naor. Fréchet embeddings of negative type metrics. *Discrete & Computational Geometry*, 38(4):726–739, 2007.

9 Sanjeev Arora, Satish Rao, and Umesh V. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2), 2009.

10 Yonatan Aumann and Yuval Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing*, 27(1):291–301, 1998.

11 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.

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

13 Yair Bartal, Amos Fiat, and Yuval Rabani. Competitive algorithms for distributed data management. *J. Comput. Syst. Sci.*, 51(3):341–358, 1995.

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

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

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

17 Moses Charikar, Tom Leighton, Shi Li, and Ankur Moitra. Vertex sparsifiers and abstract rounding algorithms. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 265–274, 2010.

18 Don Coppersmith and Michael Elkin. Sparse source-wise and pair-wise distance preservers. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'05, pages 660–669, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

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

20 Michael Elkin, Arnold Filtser, and Ofer Neiman. Prioritized metric structures and embedding. In *Proceedings of the 47th ACM Symposium on Theory of Computing*, STOC'15, 2015.

21 Michael Elkin and Shay Solomon. Steiner shallow-light trees are exponentially lighter than spanning ones. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 373–382, 2011.

22 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.

23 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.

24 T. Figiel, J. Lindenstrauss, and V.D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139(1):53–94, 1977.

25 E. N. Gilbert and H. O. Pollak. Steiner minimal trees. *SIAM Journal on Applied Mathematics*, 16(1):1–29, Jan 1968.
26 Anupam Gupta. Steiner points in tree metrics don't (really) help. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'01, pages 220–227, Philadelphia, PA, USA, 2001. Society for Industrial and Applied Mathematics.
27 Anupam Gupta, Viswanath Nagarajan, and R. Ravi. An improved approximation algorithm for requirement cut. *Oper. Res. Lett.*, 38(4):322–325, 2010.
28 Sariel Har-Peled, Piotr Indyk, and Anastasios Sidiropoulos. Euclidean spanners in high dimensions. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'13, pages 804–809. SIAM, 2013.
29 William Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
30 Lior Kamma, Robert Krauthgamer, and Huy L. Nguyen. Cutting corners cheaply, or how to remove steiner points. In *SODA*, pages 1029–1040, 2014.
31 Telikepalli Kavitha and Nithin M. Varma. Small stretch pairwise spanners. In *Proceedings of the 40th International Conference on Automata, Languages, and Programming – Volume Part I*, ICALP'13, pages 601–612, Berlin, Heidelberg, 2013. Springer-Verlag.
32 Samir Khuller, Balaji Raghavachari, and Neal E. Young. Balancing minimum spanning trees and shortest-path trees. *Algorithmica*, 14(4):305–321, 1995.
33 Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. Measured descent: a new embedding method for finite metrics. *Geometric and Functional Analysis*, 15(4):839–858, 2005.
34 Frank Thomson Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.
35 N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
36 Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. In *FOCS*, pages 109–118, 2006.
37 Ankur Moitra. Approximation algorithms for multicommodity-type problems with guarantees independent of the graph size. In *FOCS*, pages 3–12, 2009.
38 Yuri Rabinovich and Ran Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete & Computational Geometry*, 19(1):79–94, 1998.
39 Harald Räcke. Minimizing congestion in general networks. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS'02, pages 43–52, Washington, DC, USA, 2002. IEEE Computer Society.
40 Harald Räcke. Optimal hierarchical decompositions for congestion minimization in networks. In *STOC*, pages 255–264, 2008.
41 Liam Roditty, Mikkel Thorup, and Uri Zwick. Deterministic constructions of approximate distance oracles and spanners. In *Proceedings of the 32Nd International Conference on Automata, Languages and Programming*, ICALP'05, pages 261–272, Berlin, Heidelberg, 2005. Springer-Verlag.