1 Introduction

We'd like to approximate arbitrary finite metric spaces (X, ρ) with simpler ones. So far, the approximating (host) metric spaces we've considered have been $(Y \subset \mathbb{R}^d, \ell_p)$ for some small d and $p \geq 1$. For many applications (e.g. metric labeling, buy-at-bulk network design, minimum cost communication, group Steiner tree, metrical task system), tree metrics are also simple and convenient.

Definition 1. *A* tree metric *ρ_T is the shortest-path metric on a tree T.*

Problem 1. *Given a finite metric (X, ρ), find a tree $T \supseteq X$ such that for all $x, y \in X$:*

1. *(dominance) $\rho(x, y) \leq \rho_T(x, y)$, and*
2. *(bounded expansion) $\rho_T(x, y) \leq \alpha \cdot \rho(x, y)$.*

A solution to this problem for some $\alpha \geq 1$ reduces a number of optimization problems involving general finite metrics (X, ρ) to the same problems restricted to tree metrics, losing an α factor in solution quality.

Unfortunately, some spaces require $\alpha = \Omega(|X|)$. For example, consider C_n, the cycle on n vertices, with unit length edges. Say we delete a single edge $\{x, y\}$ so we have a tree T. Then, while $\rho(x, y) = 1$, we have $\rho_T(x, y) = n - 1$. Actually, $\alpha = \Omega(|X|)$ is required of any tree metric (not just spanning tree metrics) for approximating C_n (Rabinovich and Raz, 1998). Instead, we'll approximate (X, ρ) with a distribution over tree metrics. This is often sufficient to give *expected* approximation/competitive ratios for the above applications, provided that we can sample from the distribution.

Definition 2. *A distribution $\mathcal{D}$ over a family of tree metrics $\mathcal{S}$ α-*probabilistically approximates *a metric (X, ρ) if:*

1. *every metric in $\mathcal{S}$ dominates ρ (so $\rho(x, y) \leq \rho_T(x, y)$ for all $x, y \in X, \rho_T \in \mathcal{S}$), and*
2. *for all $x, y \in X$, $\mathbb{E}_{\rho_T \sim \mathcal{D}}[\rho_T(x, y)] \leq \alpha \cdot \rho(x, y)$.*

Problem 2. *Given a finite metric (X, ρ), find distribution $\mathcal{D}$ over a family of tree metrics $\mathcal{S}$ that α-probabilistically approximates (X, ρ).*

Exercise 1. *Show that C_n can be 2-probabilistically approximated by a distribution over (spanning) tree metrics (Karp, 1989).*

In this note, we'll show the construction of Fakcharoenphol, Rao, and Talwar (2003) that achieves $\alpha = \mathrm{O}(\log |X|)$ for any finite metric (X, ρ), which is optimal up to constant factors (Bartal, 1998). Thus, any finite metric space (X, ρ) embeds into a distribution over dominating tree metric spaces with $\mathrm{O}(\log |X|)$ distortion. The construction gives an efficient procedure for sampling from this distribution.

1.1 Outline

We'll show a randomized algorithm that constructs a sequence of decompositions of X, which are of interest in their own right. This sequence of decompositions in turn defines a tree $T \supset X$ for which ρ_T dominates ρ. Thus, we'll have implicitly (algorithmically) described a distribution over tree metrics. Letting ρ_T be a random variable following this distribution, we'll prove that $\mathbb{E}[\rho_T(x, y)] \leq \rho(x, y) \cdot \mathrm{O}(\log |X|)$.

2 The algorithm

Let (X, ρ) be the finite metric space with $|X| = n$ and diameter $\Delta = \max_{x,y \in X} \rho(x, y)$. Without loss of generality, we'll assume that $\rho(x, y) > 1$ for all $x, y \in X$, $x \neq y$, and that $\Delta = 2^\delta$ for some $\delta \in \mathbb{N}$. For some $x \in X$ and $r > 0$, the ball of radius r centered at x is $B(x, r) = \{y \in X : \rho(x, y) \leq r\}$.

2.1 From decompositions to trees

Definition 3. *An* r-cut decomposition *of* X *is a partitioning of* X *into clusters* $\{S_1, S_2, \ldots\}$ *such that each cluster* $S_i \subseteq B(x_i, r)$ *for some* $x_i \in X$.

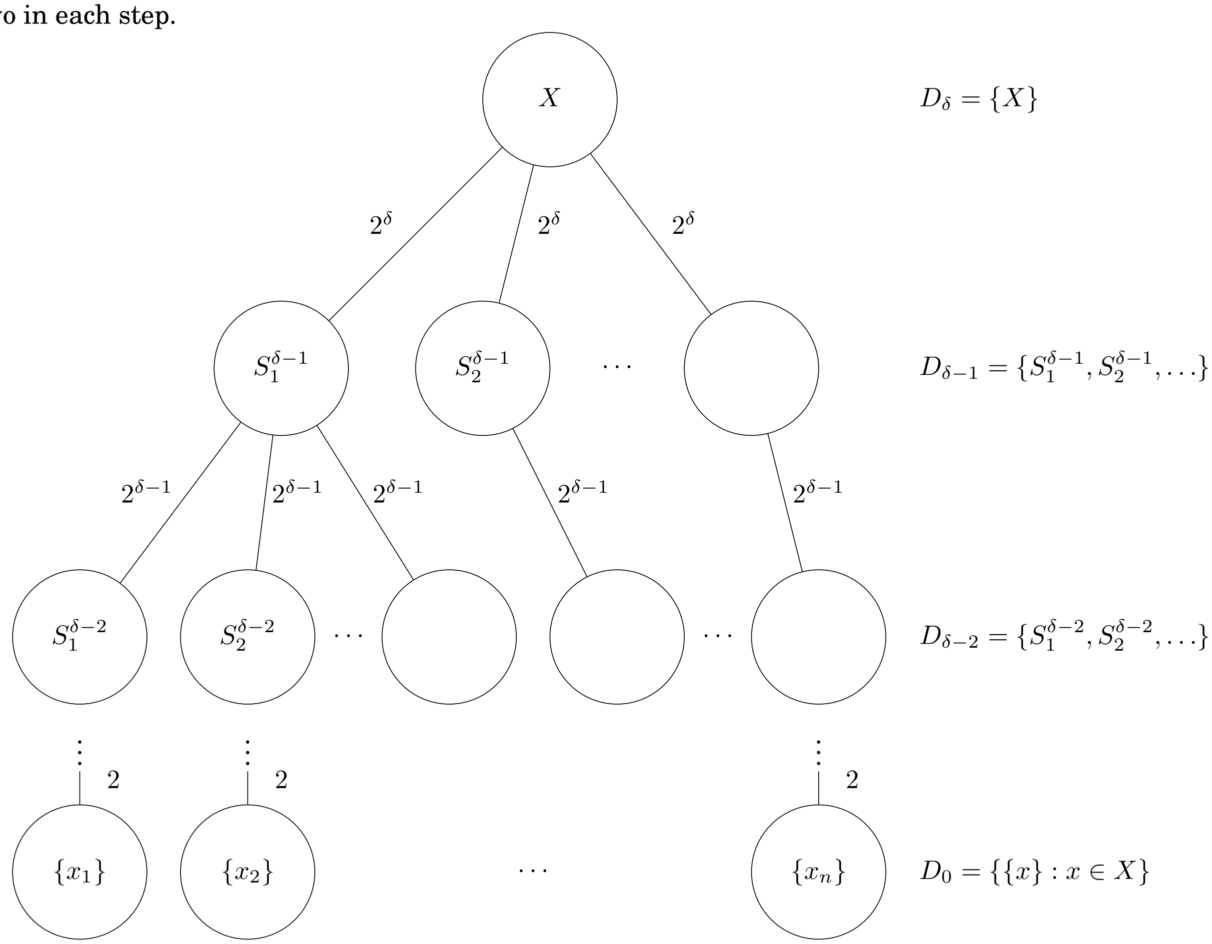
Definition 4. *A* hierarchical cut decomposition *of* X *is a sequence of* $\delta+1$ *partitionings* $D_0, D_1, \ldots, D_\delta$ *such that* $D_\delta = \{X\}$ *and each* D_i *refines* D_{i+1} *and is a* 2^i*-cut decomposition of* X. *Note that because* $\rho(x, y) > 1$ *for* $x \neq y$, *each cluster in* D_0 *is a singleton set, i.e.* $D_0 = \{\{x\} : x \in X\}$.

The following procedure constructs an r-cut decomposition of any $S \subseteq X$. It will be repeatedly called to construct a hierarchical cut decomposition of X. Let π be a permutation over the points in X.

Procedure Partition(S, π, r):

- For $j = 1, 2, \ldots$ until $S = \emptyset$:
  - $S_j \leftarrow B(\pi(j), r) \cap S$.
  - $S \leftarrow S \setminus S_j$.
- Return $\{S_1, S_2, \ldots\}$.

A hierarchical cut decomposition naturally arrange in a tree T: let the children of $S \in D_{i+1}$ be the $S' \in D_i$ that partition S (i.e. the $S' \in D_{i+1}$ with $S' \subseteq S$). So, the $\delta + 1$ levels of the tree are exactly $D_0, D_1, \ldots, D_\delta$ with the singleton sets of D_0 as the leaves. Let all edges between nodes in D_i and D_{i+1} have length 2^{i+1}, so the edge lengths on the path from the root X to any leaf $\{x\}$ decreases by a factor of two in each step.



2.2 Distances in T

Before describing the randomized algorithm for decomposing X (it will be repeated invocations of Partition), we first consider the metric ρ_T induced by the tree T corresponding to the hierarchical cut decomposition

$D_0, D_1, \ldots, D_\delta$ (equate the singleton sets $\{x\} \in D_0$ with the respective $x \in X$, so ρ_T is indeed a metric over X). For any $x, y \in X$, $x \neq y$, we say the pair (x, y) is *at level* i if x and y last appear in different clusters in D_i, in order of increasing i. Equivalently, (x, y) is at level i if x and y first appear in the same cluster S in D_{i+1}. The path from x to y in T includes the edges leading up from x to this joining cluster S, and those leading back down to the leaf y. Thus, the distance $\rho_T(x, y)$ is

$$\rho_T(x,y) = \underbrace{2^1 + 2^2 + \ldots 2^{i+1}}_{\text{path up the tree}} + \underbrace{2^{i+1} + 2^i + \ldots + 2^1}_{\text{path down the tree}} = 2\sum_{j=1}^{i+1} 2^j. \tag{1}$$

Thus $2^{i+2} \leq \rho_T(x, y) < 2^{i+3}$. The metric ρ_T dominates ρ because each $v \in S \in D_{i+1}$ is at most distance 2^{i+1} from some $u \in X$, so $\rho(x, y) \leq \rho(x, u) + \rho(u, y) \leq 2 \times 2^{i+1} = 2^{i+2} \leq \rho_T(x, y)$. The goal is now clear. We want to (randomly) construct the decompositions $D_0, D_1, \ldots, D_\delta$ so that any pair (x, y) with $\rho(x, y) \approx 2^i$ is (in expectation) at level $\approx i$.

2.3 The main algorithm

We now state the algorithm for decomposing X into the hierarchical cut decomposition $D_0, D_1, \ldots, D_\delta$. First, choose a permutation π of the points in X uniformly at random. Then, independently choose β uniformly at random from the interval $[1, 2]$. Set $D_\delta = \{X\}$. Then, for $i = \delta - 1, \delta - 2, \ldots, 0$, let

$$D_i = \bigcup_{S \in D_{i+1}} \mathsf{Partition}(S, \pi, 2^{i-1}\beta). \tag{2}$$

It's clear from this description that for $0 \leq i < \delta$, we have that D_i is a 2^i-cut decomposition and that D_i refines D_{i+1}. The resulting hierarchical cut decomposition $D_0, D_1, \ldots, D_\delta$ is arranged into a tree T in the fashion described above.

3 Analysis

Fix a pair of points (x, y). These points x and y are in the same cluster for some initial phase of the algorithm. The fate of $\rho_T(x, y)$ is determined when x and y are first put into separate clusters by some partitioning D_i. This is the event we must analyze.

Suppose x and y are in the same cluster in D_{i+1}. In the formation of D_i, let $c \in X$ be the first (in order of the permutation π) cluster center to which one of x and y is assigned. We say that c *settles* the pair (x, y) in D_i. Without loss of generality, assume x is assigned the the cluster centered at c. Now, either y is assigned to the same cluster, or it is not. In the latter case, we say that c *cuts* the pair (x, y) in D_i; it is this event that determines the fate of $\rho_T(x, y)$, because if a vertex cuts (x, y) in D_i, then (x, y) is at level i and so $\rho_T(x, y) \approx 2^{i+3}$. Since only one cluster center can cut a pair (x, y) in the entire run of the algorithm, we have

$$\rho_T(x,y) < \sum_{i=0}^{\delta-1} \sum_{c \in X} \mathbb{1}[c \text{ cuts } (x,y) \text{ in } D_i] \times 2^{i+3}. \tag{3}$$

3.1 Cutting (x, y)

Order all the points (centers) in X by increasing distance from $\{x, y\}$, so $\rho(c_1, \{x, y\}) \leq \rho(c_2, \{x, y\}) \leq \ldots \leq \rho(c_n, \{x, y\})$. For c_k to cut (x, y), we must have:

- A: $\rho(c_k, \{x, y\}) \leq 2^{i-1}\beta < \max\{\rho(c_k, x), \rho(c_k, y)\}$, and
- B: c_k settles (x, y).

Event A occurs when

$$\beta \in \left[\frac{\min\{\rho(c_k,x), \rho(c_k,y)\}}{2^{i-1}}, \frac{\max\{\rho(c_k,x), \rho(c_k,y)\}}{2^{i-1}}\right). \tag{4}$$

This interval has length $|\rho(c_k, x) - \rho(c_k, y)|/2^{i-1}$; if it intersects $[1, 2]$ (an interval of length 1), then

$$\Pr(A) \leq \frac{|\rho(c_k, x) - \rho(c_k, y)|}{2^{i-1}} \leq \frac{\rho(x, y)}{2^{i-1}}, \tag{5}$$

where second the inequality follows from the triangle inequality. Given that A occurs, event B occurs when c_k occurs first in π among $\{c \in X : \rho(c, \{x, y\}) \leq 2^{i-1}\beta\}$. That is, it must be c_k that settles (x, y), rather than some other center similarly close to (x, y). The aforementioned set contains at least $\{c_1, c_2, \ldots, c_k\}$; since π is chosen uniformly at random, we have $\Pr(B|A) \leq 1/k$. Thus, the chance that c_k cuts their pair (x, y) in D_i is at most $\rho(x, y)/2^{i-1} \times 1/k$. Plugging this probability into the expectation of (3), we have

$$\mathbb{E}[\rho_T(x, y)] < \sum_{i=0}^{\delta-1} \sum_{k=1}^{n} \frac{\rho(x, y)}{2^{i-1}} \times \frac{1}{k} \times 2^{i+3} = 16\rho(x, y) \times H_n \times \delta, \tag{6}$$

where $H_n = \sum_{j=1}^{n} 1/j = \mathrm{O}(\log n)$ is the nth harmonic sum. However, this bound is $\rho(x, y) \cdot \mathrm{O}(\log n \log \Delta)$; we were hoping for $\rho(x, y) \cdot \mathrm{O}(\log n)$.

3.2 A tight bound

The key to tightening the bound is to notice that a given c_k cannot cut (x, y) at all scales; rather it only has a chance of cutting (x, y) at a few scales (at most two, in fact). This amounts to only restricting the sum in (6) to the c_k for which c_k has a non-zero probability of cutting (x, y).

Specifically, we don't need to consider c_k for which the interval in (4) doesn't intersects $[1, 2]$:

- If $\rho(c_k, \{x, y\}) > 2^i$, then c_k can never settle (x, y) because it is too far away.
- If $\max\{\rho(c_k, x), \rho(c_k, y)\} \leq 2^{i-1}$, then even if c_k settles (x, y), it will not cut (x, y).

So, the second summation in (6) can be restricted to k for which $2^{i-1} \leq \rho(c_k, \{x, y\}) \leq 2^i$:

$$\mathbb{E}[\rho_T(x, y)] < \sum_{i=0}^{\delta-1} \sum_{\substack{k: \\ \rho(c_k, \{x,y\}) \in [2^{i-1}, 2^i]}} \frac{\rho(x, y)}{2^{i-1}} \times \frac{1}{k} \times 2^{i+3} \tag{7}$$

$$= \sum_{i=0}^{\delta-1} \sum_{\substack{k: \\ \rho(c_k, \{x,y\}) \in [2^{i-1}, 2^i]}} \frac{2^4 \rho(x, y)}{k} \tag{8}$$

Notice that each c_k can contribute to at most two inner summations; this simplifies the double-summation, so we get

$$\mathbb{E}[\rho_T(x, y)] < 2 \sum_{k=1}^{n} \frac{16\rho(x, y)}{k} = 32\rho(x, y) \times H_n = \rho(x, y) \times \mathrm{O}(\log n), \tag{9}$$

which is the desired bound.

References

Y. Bartal (1998). On approximating arbitrary metrics by tree metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*.

J. Fakcharoenphol, S. Rao, and K. Talwar (2003). A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*.

R. Karp (1989). A $2k$-competitive algorithm for the circle. Manuscript.

Y. Rabinovich and R. Raz (1998). Lower bounds on the distortion of embedding finite metric spaces in graphs. *GEOMETRY: Discrete & Computational Geometry* 19.