601.435/635 Approximation Algorithms **Lecturer:** Michael Dinitz
Topic: FRT Tree Embedding **Date:** 3/26/19
Scribe: Michael Dinitz

15.1 Introduction

Last class we defined probabilistic tree embeddings and showed how to use them to reduce Group Steiner Tree on general metrics to Group Steiner Tree on trees. Today we're going to actually show how to *do* probabilistic tree embeddings, through the FRT algorithm/embedding [FRT04]. In particular, we're going to prove the following theorem:

Theorem 15.1.1 *Let (V, d) be a metric. Then there is a randomized, polytime algorithm that produces a tree metric (V', T) for V such that*

1. *$d(u, v) \leq d_T(u, v)$ for all $u, v \in V$, and*
2. *$\mathbf{E}[d_T(u, v)] \leq O(\log n) \cdot d(u, v)$ for all $u, v \in V$.*

FRT is the best possible result of this form, but it built off of ideas pioneered by Bartal, who introduced the definition of probabilistic tree embeddings and gave one with $O(\log^2 n)$ distortion [Bar96], and then improved this to $O(\log n \log \log n)$ distortion [Bar98].

15.2 The FRT Algorithm

15.2.1 Review: Hierarchical Cut Decomposition

Recall from last class the following basic definitions. For any $u \in V$ and $r \in \mathbb{R}_{\geq 0}$, let $B(u, r) = \{v \in V : d(u, v) \leq r\}$ be the abll around u of radius r. Without loss of generality, we may assume (by scaling) that $\min_{u,v \in V : u \neq v} d(u, v) = 1$. For any set $S \subseteq V$, let $\text{diam}(S) = \max_{u,v \in S} d(u, v)$ denote its diameter. Let $\Delta = 2^{\lceil \log \text{diam}(V) \rceil}$ be the smallest power of 2 such that $\Delta \geq \text{diam}(V)$.

Hierarchical Cut Decomposition: A (rooted) tree metric (V', T) for (V, d) so that

1. Every vertex $\ell \in T$ is associated with some subset $S_\ell \subseteq V$ (note that by definition of tree metric the subset associated with a leaf vertex is a node in V).
2. The root r of T has $S_r = V$.
3. If u at level i of T, then $\text{diam}(S_u) < 2^i$ (leaves at level 0, root at level $\log \Delta$).
4. If u has children $w_1, \ldots w_k$, then $\{S_{w_i}\}_{i \in [k]}$ partition S_u (i.e., $S_u = \cup_{i=1}^k S_{w_i}$ and $S_{w_i} \cap S_{w_j} = \emptyset$ for all $i \neq j$).
5. Length of an edge between a level i node and a level $i + 1$ node is 2^{i+1}.

This can be summed up with the following picture, which is directly from the textbook [WS11]:

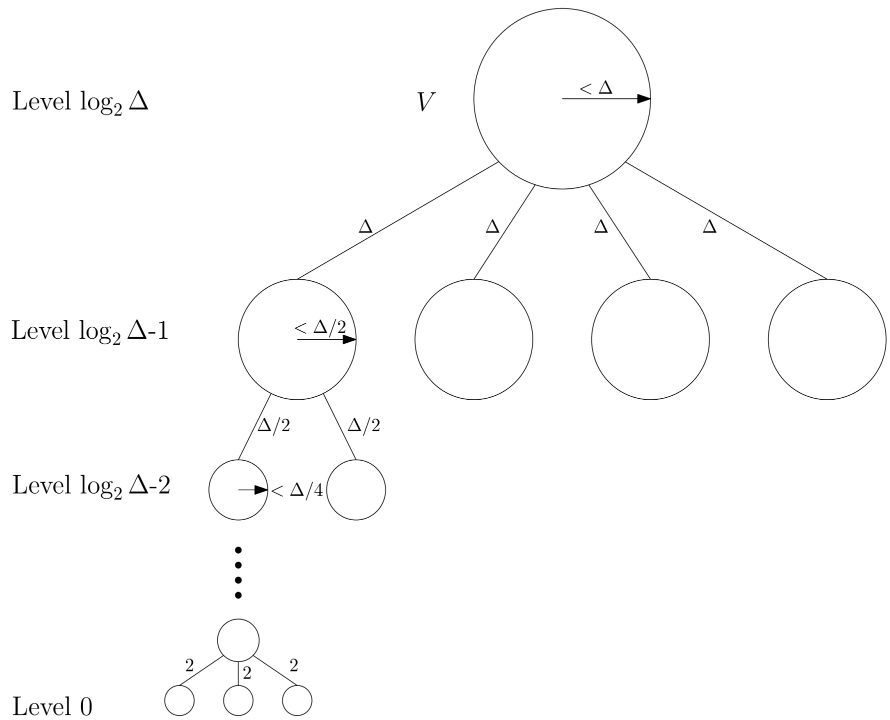

Figure 15.2.1: Hierarchical Decomposition

The FRT algorithm will construct a hierarchical cut decomposition, but before we give the algorithm, let's start by showing a simple lemma which holds for any hierarchical cut decomposition. Consider a hierarchical cut decomposition (V', T) of some metric (V, d).

Lemma 15.2.1 *If the least common ancestor of two leaf nodes u and v in T is at level i, then $d_T(u, v) \leq 2^{i+2}$. Furthermore, $d_T(u, v) \geq d(u, v)$ for all $u, v \in V$*

Proof: Let u and v be leaf nodes in T, and let w be u and v's least common ancestor (so w is at level i). Then by construction we know that $d_T(u, w) = \sum_{j=1}^{i} 2^j$, so $2^i \leq d_T(u, w) < 2^{i+1}$. Similarly, $2^i \leq d_T(v, w) < 2^{i+1}$. Since $d_T(u, v) = d_T(u, w) + d_T(w, v)$, we get that $2^{i+1} \leq d_T(u, v) < 2^{i+2}$. That proves the first part of the lemma. And because u, v are both contained in S_w, we know that $d(u, v) \leq \text{diam}(S_w) \leq 2^i$, which implies the second part. ■

15.2.2 Constructing the FRT tree

We can now finally give the FRT algorithm for constructing a tree embedding. FRT constructs a hierarchical decomposition in a certain way, but since it does construct a hierarchical decomposition, we know that no pair is contracted, and the distance between two nodes in the tree depends only on the level of their LCA. This is going to make reasoning about distances in the tree much easier.

(As a side note, you might have noticed that these trees are not just trees, they're special trees where the distance between two nodes grows exponentially with the level of their LCA. So we're actually doing more than just giving a tree embedding: we're giving a tree embedding into a special class of trees known as Hierarchically Well-Separated Trees (HSTs). Occasionally it is useful to utilize this property algorithmically: for GST we didn't care whether we were in a HST or a general tree, but for other problems it is sometimes easier to handle HSTs than general trees, and thanks to FRT we only need to handle HSTs).

Algorithm 1 FRT embedding

```
Let π be a permutation of V, chosen uniformly at random
Let r_0 be a value in [1/2, 1), chosen uniformly at random
Let r_i = r_0 · 2^i for all i such that 1 ≤ i ≤ log Δ
Let T be a tree with only a root node (at level log Δ) which represents V
for i ← log Δ to 1 do
  Let 𝒞 be the set of nodes at level i
  for C ∈ 𝒞 do
    S ← C
    for j ← 1 to n do
      P ← B(π(j), r_{i−1}) ∩ S
      if P ≠ ∅ then
        S ← S \ P
        Add P to T as a child of C at level i − 1
      end if
    end for
  end for
end for

return T
```

Note that there are two sources of randomness in this algorithm: the choice of π, and the choice of r_0.

15.3 Analysis of FRT

Since FRT gives a hierarchical cut decomposition we know that no distance is smaller in T than it is in the original metric. So we just need to prove that the expected expansion is at most $O(\log n)$, i.e., we want to prove the following theorem.

Theorem 15.3.1 $\mathbf{E}[d_T(u,v)] \leq O(\log n) d(u,v)$ *for all* $u, v \in V$

For the rest of this section, let's fix u and v. Let's introduce a couple definitions. Recall that $B(w,r)$ denotes the ball with center w and radius r.

Definition 15.3.2 w settles u, v *at level* i *if* w *is the first vertex in* π *s.t.* $B(w, r_{i-1}) \cap \{u, v\} \neq \emptyset$.

Definition 15.3.3 w cuts u, v *at level* i *if* $|B(w, r_{i-1}) \cap \{u, v\}| = 1$.

From these definitions we can make the following obvious observation. Recall that $LCA(u, v)$ is the lease common ancestor of u and v, and we know from our previous analysis of hierarchical cut decompositions that the distance between u and v is essentially determined by their LCA.

Observation 15.3.4 *$LCA(u, v)$ is at level i if i is the largest value such that the vertex w which settles u, v at level i also cuts u, v at level i.*

To analyze the expected distortion we'll need to analyze a few random variables:

$$S_{iw} = \begin{cases} 1, & \text{if } w \text{ settles } u, v \text{ at level } i, \\ 0, & \text{otherwise.} \end{cases}$$

$$X_{iw} = \begin{cases} 1, & \text{if } w \text{ cuts } u, v \text{ at level } i, \\ 0, & \text{otherwise.} \end{cases}$$

We can now start analyzing the expected distortion, although we'll have to stop a few places along the way to prove useful lemmas. Using our random variables, there is a vertex which both settles and cuts u, v at level i if $\sum_{w \in V} S_{iw} X_{iw} = 1$. Let i^* be the level of $LCA(u, v)$. Then using our observation, i^* is the largest i such that $\sum_{w \in V} S_{iw} X_{iw} = 1$. Moreover, we know from Lemma 15.2.1 that $d_T(u, v) \leq 2^{i^*+2}$. Putting this together and changing the order of summation, we get that

$$d_T(u, v) \leq 2^{i^*+2} = \max_{i : \sum_{w \in V} S_{iw} X_{iw} = 1} 2^{i+2} \leq \sum_{i=1}^{\log \Delta} 2^{i+2} \sum_{w \in V} S_{iw} X_{iw} = \sum_{w \in V} \sum_{i=1}^{\log \Delta} 2^{i+2} S_{iw} X_{iw}.$$

Now if we take the expectation, by using linearity of expectations and the definition of conditional probabilities, we get that

$$\begin{aligned} \mathbf{E}[d_T(u, v)] &\leq \sum_{w \in V} \sum_{i=1}^{\log \Delta} 2^{i+2} \mathbf{E}[S_{iw} X_{iw}] \\ &= \sum_{w \in V} \sum_{i=1}^{\log \Delta} 2^{i+2} \mathbf{Pr}[S_{iw} = 1 \wedge X_{iw} = 1] \\ &= \sum_{w \in V} \sum_{i=1}^{\log \Delta} 2^{i+2} \mathbf{Pr}[S_{iw} = 1 | X_{iw} = 1] \mathbf{Pr}[X_{iw} = 1] \end{aligned}$$

So we've (very formally) broken this up into analyzing two events: that w cuts u, v at level i (which has nothing to do with π), and that w settles u, v at level i conditioned on it cutting u, v at level i. We're going to prove a few lemmas which let us analyze these events, but consider the following intuition. w cutting u, v is independent of π: it only has to do with r_0. On the other hand, if we assume that w does cut u, v, then whether it also settles depends on π (and on r_0). So the hope is that this will be easier to analyze since we've removed the dependence on π from one of them.

The first lemma gives us a bound on the conditional event.

Lemma 15.3.5 *For every vertex w there is some $b_w \in \mathbb{R}_{\geq 0}$ such that:*

1. $\mathbf{Pr}[S_{iw} = 1 | X_{iw} = 1] \leq b_w$, *and*
2. $\sum_{w \in V} b_w \leq O(\log n)$.

The second lemma gives us a bound on the cutting probability.

Lemma 15.3.6 $\sum_{i=1}^{\log \Delta} 2^{i+2} \mathbf{Pr}[X_{iw} = 1] \leq 16 d(u,v)$ *for all $w \in V$.*

Let's now finish the proof of the main theorem, assuming these two lemmas. Continuing from our previous inequalities:

$$\begin{aligned}
\mathbf{E}[d_T(u,v)] &\leq \sum_{w \in V} \sum_{i=1}^{\log \Delta} 2^{i+2} \mathbf{Pr}[S_{iw} = 1 | X_{iw} = 1] \mathbf{Pr}[X_{iw} = 1] \\
&\leq \sum_{w \in V} b_w \sum_{i=1}^{\log \Delta} 2^{i+2} \mathbf{Pr}[X_{iw} = 1] \\
&\leq \sum_{w \in V} b_w 16 d(u,v) = 16 d(u,v) \sum_{w \in V} b_w \\
&\leq O(\log n) d(u,v)
\end{aligned}$$

So now we just need to prove these two lemmas!

Proof of Lemma 15.3.5: We're trying to analyze $\mathbf{Pr}[S_{iw} = 1 | X_{iw} = 1]$ for every $w \in V$. To do this, let's order V by distance to $\{u, v\}$, so

$$d(w_i, \{u,v\}) \leq d(w_{i+1}, \{u,v\})$$

for all i.

Now let's fix some w_j, and suppose that w_j cuts $\{u,v\}$ at level i, i.e., $|B(w_j, r_{i-1}) \cap \{u,v\}| = 1$. Then by the definition of our ordering, every w_k with $k < j$ must have $|B(w_j, r_{i-1}) \cap \{u,v\}| > 0$. Thus if *any* of these nodes come before w_j in π, we know that w_j will not settle u, v at level i, since at least one of u, v will have already been clustered by the time w_j gets to form clusters. Since π is a random permutation, the probability that w_j comes before the x_k for all $k < j$ is exactly $1/j$. Thus $\mathbf{Pr}\left[S_{iw_j} = 1 | X_{iw_j} = 1\right] \leq 1/j$. So by setting $b_{w_j} = 1/j$, we have proved the first part of the lemma.

The proof of the second part of the lemma is now straightforward:

$$\sum_{w \in V} b_w = \sum_{j=1}^{n} b_{w_j} = \sum_{j=1}^{n} \frac{1}{j} = H_n = O(\log n),$$

as claimed. ∎

Proof of Lemma 15.3.6: Now we're trying to prove that $\sum_{i=1}^{\log\Delta} 2^{i+2}\mathbf{Pr}[X_{iw} = 1] \leq 16d(u, v)$ for all $w \in V$. Without loss of generality, let's assume that $d(w, u) \leq d(w, v)$. In order for w to cut u, v at level i (i.e., for $X_{iw} = 1$), it needs to be the case that $r_{i-1} \in [d(w, u), d(w, v))$. Moreover, r_{i-1} is distributed uniformly in $[2^{i-2}, 2^{i-1})$. Thus

$$\mathbf{Pr}[X_{iw} = 1] = \frac{|[2^{i-2}, 2^{i-1}) \cap [d(w, u), d(w, v))|}{|[2^{i-2}, 2^{i-1})|} = \frac{|[2^{i-2}, 2^{i-1}) \cap [d(w, u), d(w, v))|}{2^{i-2}}.$$

So we have that

$$\begin{aligned} 2^{i+2}\mathbf{Pr}[X_{iw} = 1] &= \frac{2^{i+2}}{2^{i-2}}|[2^{i-2}, 2^{i-1}) \cap [d(w, u), d(w, v))| \\ &= 16|[2^{i-2}, 2^{i-1}) \cap [d(w, u), d(w, v))|. \end{aligned}$$

Thus

$$\begin{aligned} \sum_{i=1}^{\log\Delta} 2^{i+2}\mathbf{Pr}[X_{iw} = 1] &\leq \sum_{i=1}^{\log\Delta} 16|[2^{i-2}, 2^{i-1}) \cap [d(w, u), d(w, v)]| \\ &= 16|[d(w, u), d(w, v)]| = 16(d(w, v) - d(w, u)) \leq 16d(u, v), \end{aligned}$$

where the final inequality is from the triangle inequality. ■

15.4 Steiner Point Removal

Recall that our definition of a tree embedding for (V, d) involved us creating a tree where V was the leaves. A natural question is whether this is actually necessary: can we probabilistically embed into trees on V itself (so without any "extra" nodes)? Or even more basically, forgetting the probabilistic embedding:

Question 15.4.1 *If (V', T') is a tree metric for V, is there a (weighted) tree $T = (V, E)$ such that $d_{T'}(u, v) \leq d_T(u, v) \leq \alpha d_{T'}(u, v)$ for all $u, v \in V$, where $\alpha = O(1)$?*

This question asks whether we can turn any tree metric which uses steiner nodes ("extra" nodes) into a tree without any steiner nodes. This questions was resolved in a seminal paper by Anupam Gupta [Gup01], who showed that this was possible with $\alpha = 8$. Today we're going to prove an easier result which only holds for the kinds of tree embeddings that we construct, i.e., for hierarchical cut decompositions.

Theorem 15.4.2 *If (V', T') is a tree embedding for T which is a hierarchical cut decomposition, then can find some other T s.t. $d_{T'}(u, v) \leq d_T(u, v) \leq 4d_{T'}(u, v)$ for all $u, v \in V$.*

Proof: Use the following algorithm to construct T.

1. While there exists a node $x \in V$, s.t. $p(x) \notin V$, contract $(x, p(x))$. This gives a tree with vertex set V.

2. Multiply all edge weights by 4.

Here contracting edge $(x, p(x))$ means we just merge the subtree at x into $p(x)$ and identify the newly merged node as x. Contracting makes distance go down, and hence $d_T(u,v) \leq 4d_{T'}(u,v)$.

Suppose the least common ancestor of u, v in T' is w at level i. Then $d_{T'}(u,v) \leq 2^{i+2}$. After contractions, their distance in T is at least 2^i (consider w and its child). So $d_T(u,v) \geq 2^{i+2}$ as we multiply each edge weights by 4. So $d_{T'}(u,v) \leq d_T(u,v)$. ■

References

[Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, FOCS '96, pages 184–193, 1996.

[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, 1998.

[FRT04] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3):485 – 497, 2004. Special Issue on STOC 2003.

[Gup01] 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, 2001.

[WS11] David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms.* Cambridge University Press, New York, NY, USA, 1st edition, 2011.