COM S 783 / OR&IE 634 Approximation Algorithms **Fall Semester, 2003**

Lecture 13: October 20, 2003

Lecturer: Yuval Rabani

In today's lecture, we will present a randomized algorithm that embeds any arbitrary metric into a dominating tree metric such that each edge is distorted, in expectation, in the worst case by a factor of $O(\log n)$ and such that each edge is not contracted.

1 Overview and Notation

Recall that a metric on graph $G = (V, E)$ is a function $d : E \to \mathbb{R}$. One can define a *metric* $\mathcal{M} = (X, d)$ on an arbitrary set of vertices X with a distance function d such that the following properties hold:

- $d : X \times X \to \mathbb{R}$
- $d(i, j) \geq 0$
- $d(i, j) = 0 \Leftrightarrow i = j$
- $d(i, j) = d(j, i)$
- $d(i, j) + d(j, k) \geq d(i, k)$

A *tree metric* is the shortest path metric of a weighted tree. In other words, $d(i, j)$ is the length of the unique shortest path between node i and node j.

A metric (V', d') is said to *dominate* metric (V, d) if $\forall u, v \in V, d'(u, v) \geq d(u, v)$.

WLOG, $\forall i, j, d(i, j) \geq 1$.

Let Δ denote the diameter of the metric (V, d). WLOG, $\Delta = 2^\delta$.

Let $\mathcal{S}$ be a family of metrics over V, and $\mathcal{D}$ be a distribution of $\mathcal{S}$. We say that $(\mathcal{S}, \mathcal{D})$ α-probabilistically approximates a metric (V, d) if every metric in $\mathcal{S}$ dominates d and for every pair of vertices $(u, v) \in V$, $E_{d' \in (\mathcal{S}, \mathcal{D})}[d'(u, v)] \leq \alpha \cdot d(u, v)$. Formally, we're interested in $O(\log n)$-probabilistically approximating an arbitrary metric (V, d) by a distribution over tree metrics.

For a parameter r, an *r-cut decomposition* of (V, d) is a partitioning of V into clusters, each centered around a node and having radius at most r. Thus each cluster will have diameter at most $2r$.

A *hierarchical cut decomposition* of (V, d) is a sequence of $\delta + 1$ nested cut decompositions $D_0, D_1, \ldots, D_\delta$ such that

- $D_\delta = \{V\}$, the trivial partition where all nodes are in a single cluster
- D_i is a 2^i-cut decomposition, and a refinement of D_{i+1}.

Note that each cluster in D_0 has radius at most 1. Hence, each cluster consists of just a unique node.

Recall that a *laminar family* $\mathcal{F} \subseteq 2^V$ is a family of subsets of V such that for any $A, B \in \mathcal{F}$, it si the case that $A \subseteq B$ or $B \subseteq A$ or $A \cap B = \emptyset$.

A hierarchical cut decomposition defines a laminar family. A root tree can be generated from the decomposition as follows: Each set in the laminar family is a node in the tree and the children of a node corresponding to a set S are the nodes corresponding to maximal subsets of S in the family.

Remark 1: The node corresponding to V is the root and the singletons are the leaves.

Remark 2: The children of a set in D_{i+1} are sets in D_i.

Define a distance function on this tree as follows: The links from a node S in D_i to each of its children in the tree have length equal to the 2^i (which is an upper bound on the radius of S). Hence, define the distance function $d^T(\cdot, \cdot)$ on the tree to be the length of the shortest (unique) path distance in T from node u to node v. It's obvious that $\forall u, v, d^T(u, v) \geq d(u, v)$.

An edge (u, v) is at *level* i if u and v are first separated in the decomposition D_i. Note that if (u, v) is at level i, then $d^T(u, v) = 2 \sum_{j=0}^{i} 2^j \leq 2^{i+2}$.

2 Algorithm

Below is the random process that defines a hierarchical cut decomposition of (V, d), such that the probability that an edge (u, v) is at level i decreases geometrically with i:

- Pick a random permutation π of $\{v_1, \ldots, v_n\}$,
- Select β uniformly at random in the interval $[1, 2]$
- For each i, we compute D_i from D_{i+1} as follows. First, set β_i to be $2^{i-1}\beta$. Let S be a cluster in D_{i+1}. Assign a node $u \in S$ to the first (as defined by π) node $v \in V$ closer than β_i to u. Each child cluster of S in D_i then consists of the set of vertices in S assigned to a single center v.

Remark 1: The center v itself need not be in S. Thus, one center v may correspond to more than one cluster, each inside a different level $(i+1)$ cluster.

Remark 2: Since $\beta_i \leq 2^i$, the radius of each cluster is at most 2^i. Thus, we get a 2^i-cut decomposition.

3 Analysis

For a fixed edge (u, v), we'll show $E[d^T(u, v)] \leq O(\log n) \cdot d(u, v)$.

From previous remarks after the algorithm presented above, it follows that

$$E[d^T(u,v)] \leq \sum_{i=0}^{\delta} Pr[(u,v) \text{ is at level } i] \cdot 2^{i+2}$$

If nodes u and v are in separate clusters in D_i, we say that D_i *separates* (u, v).

Based on that definition, (u, v) is at level i if

(a) D_i separates (u, v).

(b) D_j does not separate (u, v) for any $j > i$.

Clearly if $d(u, v) > 2^{i+2}$, then u and v cannot be in the same cluster in D_{i+1}, i.e. D_{i+1} separates (u, v). From (b) above, (u, v) cannot be at level i. Let j^* be the smallest i such that $d(u, v) \leq 2^{i+2}$. Thus, $\Pr[(u, v)$ is at level $i] = 0$ for any $i < j^*$. For $i \geq j^*$, we'll find an upper bound of the probability that (u, v) is at level i.

From (a) and (b) above, for any $i \geq j^*$,

$$\begin{aligned} &Pr[(u,v) \text{ is at level } i] \\ &= Pr[D_i \text{ separates } (u,v)] \cdot Pr[\nexists j > i : D_j \text{ separates } (u,v) | D_i \text{ separates } (u,v)] \\ &\leq Pr[D_i \text{ separates } (u,v)] \end{aligned}$$

For any $j^* \leq j \leq \delta$, let K_j^u be the set of vertices in V closer than 2^j to node u, and let $k_j^u := |K_j^u|$. Define K_j^v and k_j^v similarly. For $j < j^*$, let $k_j^u = 0$.

Now consider the clustering step at level $i \geq j^*$. In each iteration, all unassigned nodes v such that $d(v, \pi(l)) \leq \beta_i$ assign themselves to $\pi(l)$. For some initial iterations of this procedure, both u and v remain unassigned. Then at some step l, at least one of u and v gets assigned to the center $\pi(l)$.

Center $\pi(l)$ *settles* the edge (u, v) at level i if its is the first center to which at least one of u and v gets assigned. Note that exactly one center settles any edge (u, v) at any particular level.

Center $\pi(l)$ *cuts* the edge $e = (u, v)$ at level i if it settles e at this level, but exactly one of u and v is assigned to $\pi(l)$ at level i. Clearly, D_i separates (u, v) iff some center w cuts it at this level. Hence $Pr[D_i \text{ separates } (u, v)] = \sum_w Pr[w \text{ cuts } (u, v) \text{ at level } i]$.

Center w *cuts u out of* (u, v) *at level* i if w cuts (u, v) at this level and u is assigned to w (and v is not assigned to w) at this level.

For each center w, we'll find an upper bound for the probability that w cuts u out of (u, v) at level i. Arrange the centers in K_i^u in increasing order of distance from u, say $w_1, w_2, \ldots, w_{k_i^u}$. For a center w_s to cut (u, v) such that only u is assigned to w_s, it must be the case that

(a) $d(u, w_s) \leq \beta_i$.

(b) $d(v, w_s) > \beta_i$.

(c) w_s settles e.

Thus β_i must lie in $[d(u, w_s), d(v, w_s)]$. By the triangle inequality, $d(v, w_s) \leq d(v, u) + d(u, w_s)$, and hence the interval $[d(u, w_s), d(v, w_s)]$ is of length at most $d(u, v)$. Since β_i is distributed uniformly in $[2^{i-1}, 2^i]$, the probability that β_i falls in the bad interval is at most $d(u, v)/2^{i-1}$. Moreover, for such a value of β_i, any of $w_1, w_2, \ldots, w_s$ can settle (u, v) at level i and hence the first amongst these in the permutation π will. Since π is a random permutation, the probability that w_s is the one to settle (u, v) at level i is at most $1/s$.

At this point, it's obvious that

$$
\begin{aligned}
& Pr[D_i \text{ separates } (u, v)] \\
& \leq \sum_{s=1}^{k_i^u} (d(u, v)/2^{i-1}) \cdot \frac{1}{s} + \sum_{s=1}^{k_i^v} (d(u, v)/2^{i-1}) \cdot \frac{1}{s} \\
& \leq (d(u, v)/2^{i-1})(\ln k_i^u + \ln k_i^v)
\end{aligned}
$$

Thus, each i contributes at most $O(\log n)$ to the expected value of $d^T(u, v)$. Hence, the expected length is bounded by $O(\log n \log \Delta)$.

However, we want an upper bound of $O(\log n)$. Observe that the total number of centers over all δ levels is n. A more careful analysis of the above procedure will give the desire result.

First consider some $i \geq j^* + 4$. Since the radius of the cluster at level i is at least 2^{i-1}, centers very close to both u and v can never cut the edge (u, v). More precisely, for any w in K^u_{i-2}, if u is assigned to w, it must be the case that v gets assigned to w also, because $d(v, w) \leq d(v, u) + d(u, w) \leq 2^{i-2} + 2^{i-2} \leq 2^{i-1} \leq \beta_i$ (since $i \geq j^* + 4$). Thus, no center in $w_1, w_2, \ldots, w_{k^u_{i-2}}$ can ever cut u out of (u, v). This implies that the probability that u gets cut out of edge e is in fact bounded by

$$\sum_{s=k^u_{(i-2)}+1}^{k^u_i} \frac{1}{s}(d(u,v)/2^{i-1})$$

$$= (d(u,v)/2^{i-1}) \cdot (H_{k^u_i} - H_{k^u_{(i-2)}})$$

Since (u, v) can be cut when either u or v is cut out by some node, the overall probability that D_i separates (u, v) is then at most $(d(u,v)/2^{i-1}) \cdot [H_{k^u_i} + H_{k^v_i} - H_{k^u_{(i-2)}} - H_{k^v_{i-2}}]$.

For $i = j^*+1, j^*+2, j^*+3$, this probability is bounded by $(d(u,v)/2^{i-1}) \cdot (H_{k^u_i} + H_{k^v_i}) \leq (d(u,v)/2^{i-1} \cdot 2H_n)$.

Hence,

$$E[d^T(u,v)]$$

$$\leq \sum_{i=0}^{\delta} Pr[(u,v) \text{ is at level } i] \cdot 2^{i+2}$$

$$\leq \sum_{i=j^*}^{\delta} Pr[D_i \text{ separates } (u,v)] \cdot 2^{i+2}$$

$$\leq \sum_{i=j^*}^{j^*+3} 2H_n \cdot \frac{d(u,v)}{2^{i-1}} \cdot 2^{i+2} + \sum_{i=j^*+4}^{\delta} (H_{k^u_i} + H_{k^v_i} - H_{k^u_{(i-2)}} - H_{k^v_{i-2}}) \cdot \frac{d(u,v)}{2^{i-1}} \cdot 2^{i+2}$$

$$\leq 8d(u,v)(4 \cdot 2H_n + H_{k^u_\delta} + H_{k^v_\delta} + H_{k^u_{\delta-1}} + H_{k^v_{\delta-1}})$$

$$\leq 8d(u,v)(12H_n)$$

$$= 96 \ln n \cdot d(u,v)$$

The third to last inequality follows because alternate terms of the summation $\sum_i (H_{k^u_i} - H_{k^u_{(i-2)}})$ telescope.

References

[1] J. Fakcharoenphol, S. Rao, K. Talwar. A Tight Bound on Approximating Arbitrary Metrics by Tree Metrics. *Proceedings of the Thirty-fifth ACM Symposium on Theory of Computing*, 448-455, 2003.