CS294-1 On-line Computation & Network Algorithms	Spring 1997
Lecture 20: April 15	
Lecturer: Yair Bartal	*Scribe: Tzu-Yi Chen*

In this lecture we continue the discussion of probabilistic approximate metric spaces [Bartal96]. The main result is stated here as theorem 20.4, and we begin the proof.

20.1 Review

We begin by reviewing the definition of a probabilistic approximate metrical task system.

Definition 20.1 (α-probabilistically approximate) *A set of metric spaces S over V $\underline{\alpha\text{-probabilistically approximates}}$ a metric space M over V if:*

1. *$\forall N \in S$, N dominates M*
2. *$\forall u, v \in V, E(d_N(u,v)) \leq \alpha \cdot d_M(u,v)$*

Naturally, we need the set S. A natural choice is the set of weighted trees where the metric space is defined as the length of the shortest path between two nodes. This implies that any metric space can be approximated by a tree. The rest of this lecture states this as a theorem and begins the proof.

20.2 Tree Metric Spaces

Theorem 20.2 *If a tree metric space α-(deterministically) approximates the k-cycle, then $\alpha = \Omega(n)$.*

Last time we saw how a line with k points 2-approximates a k-cycle.

20.2.1 Defining Hierarchically Well-Separated Trees

Definition 20.3 (k-HST) *A k-hierarchically well-separated tree (k-HST) is defined as a weighted, rooted tree with the following properties:*

1. *For every node, the distance to each of its children is the same.*
2. *For every path from the root to a leaf, the edge weight drops by a factor $\geq k$ on every step.*

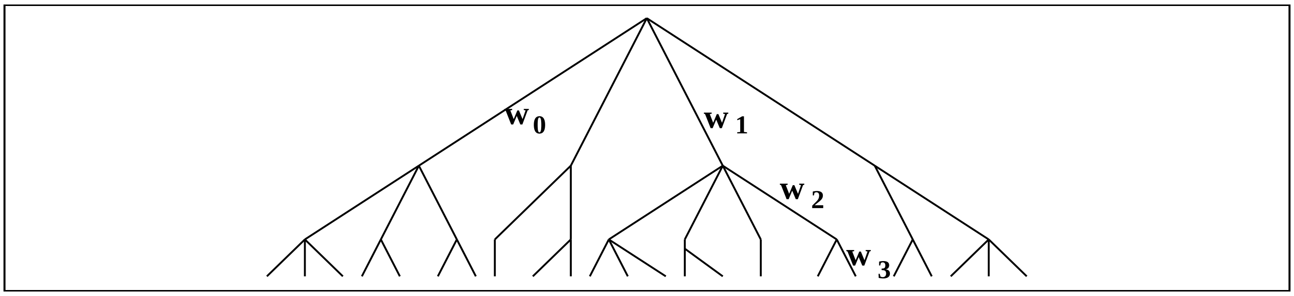

Figure 20.1: Example of a k-hst tree.

In the following example of a k-hst tree, property 1 means that $w_0 = w_1$, and property 2 means that $w_1 \geq kw_w$ and $w_2 \geq kw_3$.

The height of a tree T is the maximum number of edges on a path from the root to a leaf. We call this $h(T)$ and we know that $h(T) \leq \lceil \log_k(diam(T)) \rceil$.

Note that intuitively this definition for a k-hst tree makes sense given the approximations we want to make. Think of the children of a node as black boxes. Then, in some sense, the tree is *uniform* because the distances to the children of that node are all the same.

We now state the main theorem of this set of lectures on probabilistic approximate metric spaces:

Theorem 20.4 *Every metric space can be α-probabilistically approximated by k-HSTs with $\alpha = O(k \log^2 n) = O(k \log n \log_k(\min\{n, \Delta\}))$ and $diam(T) = O(diam(G))$.*

This result is complemented by the following:

Theorem 20.5 *There exists a metric space such that if it can be α-probabilistically approximated by any set of tree metric spaces, then $\alpha = \Omega(\log n)$.*

One result of this theorem is that it implies one should be able to get an α competitive ratio for the on-line Steiner tree problem.

20.2.2 Probabilistic Partitions and Hierarchically Well-Separated Trees

Intuitively, we don't want the clusters to be too close. Note that we can also think of building each successive level of a k-HST tree as finding a partition of the nodes in the root of the current subtree. In this subsection we formalize this intuition regarding the construction of k-HST trees.

Definition 20.6 (l-partition) *An l-partition of a weighted, connected graph G is a collection of subsets of vertices $P = \{V_1, V_2, \ldots, V_S\}$ such that:*

1. $\forall \; 1 \leq i \leq s, \;\; V_i \leq V$ *and* $\cup_i V_i = V$

2. $\forall\ 1 \leq i, j \leq s, i \neq j,\ \ V_i \cap V_j = \emptyset$
3. *Let* $C(V_i)$ *(a cluster) denote the subgraph of* G *induced by* V_i. $\forall\ 1 \leq i \leq s,\ \ diam(C(V_i)) \leq l$

The following is an example of a graph partition into four clusters.

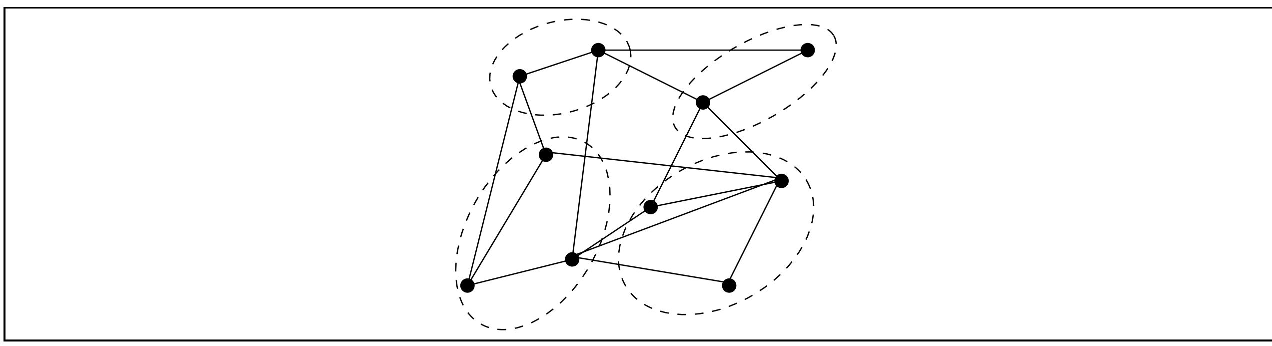

Figure 20.2: Example of a graph partition into four clusters.

Now let D be a probability distribution over l-partitions of G. Consider the edge $e = (u, v) \in E$ and the corresponding $x(e)$, which we define to be the probability that u and v belong to different clusters. In other words, $x(e)$ is the probability that edge e is in a cut. This gives us the terminology necessary for the following definition.

Definition 20.7 ((r, ρ, λ)-approximate partition) *An* (r, ρ, λ)*-approximate partition of* G *is a probability distribution* D *over* $(r\rho)$*-partitions of* G *s.t.* $\max_{e \in E} \left\{ x(e) \frac{r}{w(e)} \right\} \leq \lambda$.

The insight into this definition, as partially illustrated in the following figure, is that the probability that an edge e goes between clusters should be proportional to its length. Hence, $x(e) \leq \lambda \frac{w(e)}{r}$.

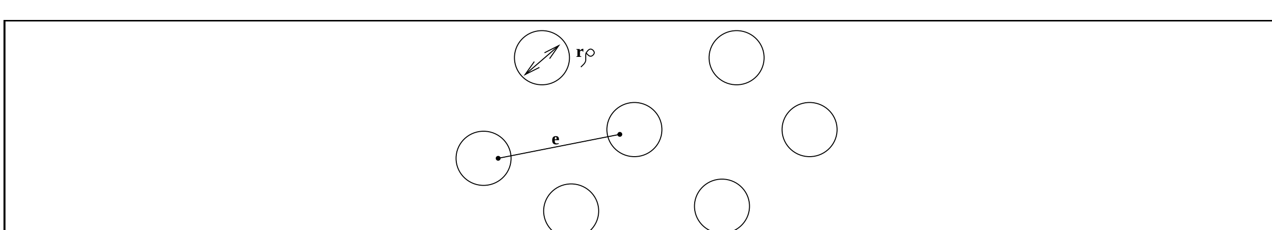

Figure 20.3: Insight into definition 20.7.

Definition 20.8 (ϵ-forcing) *A probabilistic partition ie* ϵ*-forcing if the following holds:*

$$\forall e\ \ s.t.\ \ \frac{w(e)}{r} \leq \epsilon \Rightarrow x(e) = 0$$

Now let $x(u, v)$ be the probability that the shortest path between u and v goes between clusters.

Theorem 20.9 *There exists a probabilistic partition with* $\rho = 2\ln n + 1, \lambda = 2, \epsilon = \frac{1}{n}$ *for all weighted graphs.*

Theorem 20.10 *If every subgraph of* G *has* ϵ*-forcing* (r, ρ, λ)*-probabilistic partitions, then* G *can be alpha-probabilistic approimxated by* k*-HSTs with* $\alpha = O(k\rho\lambda \log_k(\min\{\Delta, n\log n\}))$.

20.2.3 Constructing Probabilistic Partitions

In this subsection we present an algorithm for constructing probabilistic partitions and then an analysis of the algorithm.

For a graph H, let $B_z^H(v) = \{u \in V(H) | d(v,u) \leq z\}$ denote the ball of radius z around v. An example follows:

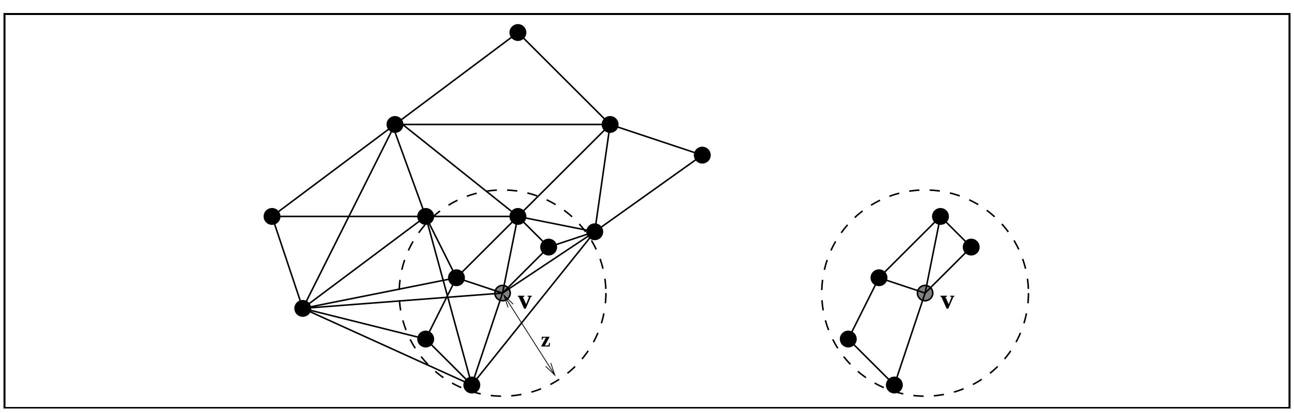

Figure 20.4: Example of a subgraph induced by a ball. The left hand side of the figure shows a graph, a vertex v, and a ball of radius z about v. The right hand side of the figure shows the subgraph induced by this ball.

20.2.3.1 Algorithm

We follow the following algorithm to recursively construct the probabilistic partition. At each stage we construct a cluster. At the t^{th} stage, H_t is the subgraph of G induced by $U_t \subseteq V$.

$H_0 = G, V_0 = V$
while $U_t \neq \emptyset$
    Pick an arbitrary $v_t \in U_t$.
    if the connected component of H_t including v_t has diameter at most $2r\ln n$
        then set the next cluster C_t of the partition equal to that connected component.
        else Choose radius z using the probability distribution.
            G_t is the subgraph of H_t induced by the ball $B_t^{H_t}(v)$.
    $U_{t+1} = U_t \backslash v(C_t)$ and H_{t+1} as the subgraph of G induced by U_{t+1}.
end while

20.2.3.2 Analysis

Clearly we need to know the appropriate probability distribution. It turns out that the correct distribution is

$$p(z) = \left(\frac{n}{n-1}\right)\frac{1}{r}e^{-\frac{z}{r}}$$

on the interval $[0, r\ln n)$.

Now we want a bound on the probability that an edge is in a cut.

Consider $e = (u, w)$. Fix t Set $V = V_t$. Now assume wlog that $d(v,u) \leq d(v,w)$. Let $\tilde{d}(u,v) = \min\{d(u,v), r\ln n\}$. Notice that the triangle inequality is preserved. The following definitions are also useful:

$$\begin{array}{rcl} A_t & : & u, w \in U_t \\ X_t & : & (v,w) \text{ is in none of the clusters } c_j \text{ for } j \geq t, \text{ conditional on } A_t \\ M_t^I & : & \tilde{d}(v,w) \leq z, \text{ conditional on } A_t \\ M_t^X & : & \tilde{d}(v,u) \leq z \leq \tilde{d}(v,w), \text{ conditional on } A_t \\ M_t^N & : & z \leq \tilde{d}(v,u), \text{ conditional on } A_t \end{array}$$

Note that if u, w is not already in a cluster, X_t is the probability that they will be cut eventually. We're interested in X_0.

The event M_t^I corresponds to the case where an edge (u,w) is included in C_t. The event M_t^X corresonds to the case where u is included in C_t, but w is not — therefore the edge (u,w) will be exluded from any cluster in the partition. The event M_t^N corresponds to the case that both u and w are not included in C_t, and are therefore passed onto the next stage.

We now show how to get an upper bound on $X(e) = Pr(X_0)$. Note that $Pr(A_0) = 1$.

We have the following recursive relation:

$$Pr(X_t) = Pr(M_t^X) + Pr(M_t^N) \cdot Pr(X_{t+1})$$

We now derive formulas for $Pr(M_t^X)$ and $Pr(M_t^N)$.

Since we know the following standard integral:

$$\int_a^b \frac{1}{r}e^{-\frac{z}{r}}\,dz = (1 - e^{-\frac{b-a}{r}})e^{-\frac{a}{r}}$$

we know the following:

$$\begin{aligned} Pr(M_t^X) &= \int_{\tilde{d}(v,r)}^{\tilde{d}(v,w)} P(z)\,dz \\ &= \left(\frac{n}{n-1}\right)\left(1 - e^{-\frac{\tilde{d}(v,w)-\tilde{d}(v,u)}{r}}\right)e^{-\frac{\tilde{d}(v,u)}{r}} \end{aligned}$$

The triangle inequality tells us that $\tilde{d}(v,w) - \tilde{d}(v,u) \leq \tilde{d}(u,w)$, which gives us:

$$Pr(M_t^X) \leq \left(\frac{n}{n-1}\right)\left(\frac{\tilde{d}(u,w)}{r}\right) e^{-\frac{\tilde{d}(v,u)}{r}}$$

We can also derive the following for $Pr(M_t^N)$:

$$\begin{aligned} Pr(M_t^N) &= \int_0^{\tilde{d}(v,w)} P(z)\,dz \\ &= \left(\frac{n}{n-1}\right)\left(1 - e^{-\frac{\tilde{d}(v,u)}{r}}\right) \end{aligned}$$

We are now ready to derive a formula for $Pr(X_t)$. We do so by induction. Assume:

$$Pr(X_t) \leq \left(2 - \frac{t}{n-1}\right)\frac{\tilde{d}(u,w)}{r}$$

The base case is trivial: if some $t < n$ is the last step, then $Pr(X_t) = 0$. Now assume true for $t+1$ and we show it to be true for t.

$$\begin{aligned} Pr(X_t) &= Pr(M_t^X) + Pr(M_t^N)\cdot Pr(X_{t+1}) \\ &\leq \left(\frac{n}{n-1}\right)\left(\frac{\tilde{d}(u,w)}{r}\right) e^{-\frac{\tilde{d}(v,u)}{r}} + \left(\frac{n}{n-1}\right)\left(1 - e^{-\frac{\tilde{d}(v,u)}{r}}\right)\left(2 + \frac{t+1}{n}\right)\frac{\tilde{d}(u,w)}{r} \\ &= \left(\frac{n}{n-1}\right)\left[1 + \frac{n-t-2}{n-1}\left(1 - e^{-\frac{\tilde{d}(v,u)}{r}}\right)\right]\frac{\tilde{d}(u,w)}{r} \end{aligned}$$

We know $e^{-\frac{\tilde{d}(v,u)}{r}} \geq \frac{1}{n}$, so:

$$Pr(X_t) \leq \left(2 - \frac{t}{n-1}\right)\frac{\tilde{d}(u,w)}{r}$$

Therefore $Pr(X_0) \leq 2\frac{\tilde{d}(u,w)}{r}$, which gives us the upper bound we wanted.

Referring to Theorem 20.9 we see a +1 term which needs explaining. This term comes from a $\frac{1}{n}$-forcing term. The intuition is that for every edge that has the property $\frac{w(e)}{r} \leq \frac{1}{n}$, we contract it so that there are no edges left with the property. At the end we replace the edges contracted. However, we know that the replaced edges cannot increase the diameter by more than r. This gives us the +1 term in Theorem 20.9.

References

[Bartal96] Y. Bartal, *Probabilistic Approximation of Metric Spaces and its Algorithmic Applications*, in Proc. of the 37rd Ann. IEEE Symp. on Foundations of Computer Science, 1996.