15-859(J): Metric Embeddings and Methods **Lecturer:** Anupam Gupta
Topic: Tree Covers: Embeddings and Routing Algorithms **Date:** October 7, 2003
Scribe: Yiannis Koutis

11.1 Motivation and Definition

Suppose we are given C_n, the cycle with n-vertices. If we pick any spanning path P_n of C_n, there are two vertices which are at distance $n-1$ in P_n, an increase by a factor of $n-1$ compared to their original distance in C_n. However, if we were to now pick two paths P_1 and P_2 obtained by breaking C_n in two antidiametric edges, the distance of any two points in C_n is preserved either in P_1 or in P_2. Hence, this gives yet another way of representing the original distance metric by trees, and we say that the two paths P_1 and P_2 **cover** C_n. This motivates the following definition.

Definition 11.1.1 *A tree cover $\mathcal{F}$ of a graph G (and hence of the metric defined by it) is a family $\mathcal{F} = \{T_1, \ldots, T_k\}$ of trees such that*

1. *Each tree dilates distances, i.e., $d_{T_j}(x, g) \geq d_G(x, y)$ for all j.*
2. *For any two nodes (x, y) of G, some tree maintains the distance between x and y; i.e., there is an i such that $d_{T_i}(x, g) = d_G(x, y)$.*

There are certain situations where tree covers are useful.

Network routing. Given a complex network, and a routing protocol for trees, we can cover the network with a (hopefully small) family of trees; now given a packet, we can use the tree routing protocol to route it on the tree that acieves the shortest path distance between its source and destination.

Note that if the tree cover is of size k, then we use $\log k$ extra bits in the packet header to specify which tree to route on, and each node has to keep information for routing on k trees.

Distance Labeling. A distance labeling scheme is one which assigns labels to vertices such that one can infer the distance between any two vertices by just looking at their labels. It is trivial to obtain a labeling using n *numbers* for each vertex. (Why?)

If a graph has a k-cover, then it turns out that we can label each vertex by only $O(k \log n)$ numbers. This follows from the fact that trees have distance labels with $O(\log n)$ numbers, in turn an immediate consequence of the result of Linial, London and Rabinovich (1995) that for any tree T,

$$(V(T), d_T) \hookrightarrow \ell_\infty^{O(\log n)}$$

As an aside, there are many results on distance labelings: e.g., a result of Winkler (1983) shows that any unweighted graph admits a distance labeling with $O(n)$ bits, an improvement of $O(\log n)$ bits over the trivial bound. (A counting argument shows that this is the best we can do without restricting the graph family we are labeling.)

11.2 Tree Covers with stretch

Let us first observe that all graphs trivially have linear sized tree covers.

Lemma 11.2.1 *Every graph has a tree cover of n trees, consisting of the shortest path trees from each node.*

It is equally easy to show that this is nearly tight.

Lemma 11.2.2 *The unweighted complete graph K_n does not have a $n/2-1$-cover.*

Proof: Note that the union of the trees must cover all the edges, else the distance between the ends of the uncovered edge would not be maintained. However, this is not possible witn $n/2-1$ trees, since each tree has $n-1$ edges, and hence the total edge count is $(\frac{n}{2}-1)\times(n-1)<\binom{n}{2}$. ■

In view of the above lemma, we would like to introduce tree covers with **stretch**.

Definition 11.2.3 *A* stretch D *tree cover $\mathcal{F}$ of a graph G (and hence of the metric defined by it) is a family $\mathcal{F}=\{T_1,\ldots,T_k\}$ of trees such that*

1. *Each tree dilates distances, i.e., $d_{T_j}(x,g)\geq d_G(x,y)$ for all j.*
2. *For any two nodes (x,y) of G, some tree maintains the distance between x and y to with a factor of D; i.e., there is an i such that*

$$d_{T_i}(x,g)\leq D\times d_G(x,y).$$

11.3 Special Families of Graphs: Avoiding Stretch

Let us see how far we can go without having to introduce stretch; to this end, let us consider some special families of graphs, namely planar graphs and other graph families with small separators. To warm up, let us prove something for a basic planar graph, the square grid.

Theorem 11.3.1 *The unweighted n-vertex square grid G has a $O(\log n)$-sized tree cover.*

Proof: In fact, we prove this for rectangular grids as well: let us assume that the grid is wider than it is high (else we can rotate it). To build the first tree, let us take all the horizontal edges, and vertical edges in the median column of G. Observe that if a shortest path between any two vertices x and y hit the median column, then this tree T maintains the distance between x and y.

We now delete the median column, and consider the left and right subgrids produced. Each can be covered by $O(\log(n/2))$ trees, and we can pair these trees up to get $O(\log(n/2))$ trees that maintain all distances except between the left and right parts. However, the tree T maintains these very distances, and adding this to the cover gives us $O(\log(n/2))+1\leq O(\log n)$ trees, proving the theorem. ■

Problem: Can you prove that we need at least $\log n$ trees?

To continue to more complicated graphs, we will need one additional definition

Definition 11.3.2 *A graph G has a* $1/3$-balanced $r(n)$-sized vertex separator *if there is a $r(n)$-sized set of vertices, which if removed, breaks the graph in two components of size at least $n/3$.*

The graph G has a recursive 1/3-balanced $r(n)$-sized vertex separator *if the same holds recursively for the components formed after deleting the separator (all the way down to singletons).*

Outerplanar graphs are graphs that do not contain K_4 and $K_{2,3}$ as minors, and can be drawn in the plane such that all nodes lie on a cycle. It can be shown that outerplanar graphs have 2-sized recursive separators. Similarly, *Series-parallel* graphs have 2-sized recursive separators, and *treewidth-k* graphs have k-sized recursive separators. One of the most basic theorems for *planar graphs* is the following:

Theorem 11.3.3 (Lipton and Tarjan (1979)) *Every planar graph has recursive* $1/3$*-balanced* $O(\sqrt{n})$*-sized vertex separators.*

We can prove the following simple result:

Theorem 11.3.4 *Every graph with recursive* $1/3$*-balanced* $r(n)$*-sized vertex separators has an* $r(n)\log_{3/2} n$*-sized tree cover.*

Proof: The proof proceeds similarly to the proof for the square grid. We first find the $r(n)$-sized separator S. For each of the points in the separator S, we grow shortest paths trees in the entire graph G. Note that if a shortest path between u and v went through a vertex $s \in S$, then the distance between them in the shortest-path tree rooted at s would be exactly $d_G(u, v)$. Hence, for all the vertex-pairs whose distances in G and $G - S$ are not the same, there is a tree amongst the $r(n)$ trees that maintains their distance. Now we just recurse on each of the components and obtain tree covers for them, pair them up to get at most $r(n)\log_{3/2}(2n/3)$ trees in the cover. Adding the $r(n)$ trees at the top level to this proves the theorem. ■

If $r(n) = \text{poly}(n)$, then one can show that the analysis of the above theorem can be tightened to give $O(r(n))$ trees; thus we get $O(\sqrt{n})$-sized tree-covers for planar graphs. This result is tight up to constant factors, as the following result shows.

Theorem 11.3.5 (Gupta, Kumar and Rastogi (2001)) *There exists a planar graph G such that any stretch-1 tree cover has size* $\Omega(\sqrt{n})$*.*

Let us just point out the graphs for which the lower bound holds: we start with the square grid. The fact that allowed us to obtain the $O(\log n)$-sized cover is the symmetry in the distances. The lower bounds graphs are obtained by breaking the symmetry of each cell of the grid by placing edge weights such that the shortest path between the left-upper corner and the bottom-right corner is obtained by going "first south, then east", and the shortest-path between bottom-left corner and the top-right corner is obtained by going "first north, then east". This simple change is enough to prove the theorem; details are in the cited paper.

11.4 Tree Covers for General Graphs

Theorem 11.4.1 *For general graphs, every tree cover with stretch* $2k$ *must have size* $\Omega(n^{\frac{2}{3k-6}})$*. It is conjectured that the tightest bound is* $\Omega(n^{\frac{1}{k}})$*.*

Proof: The proof is very similar to Lemma 11.2.2. Suppose we are given a graph G with $f(n, k)$ edges and girth $2k + 1$. Note that every edge in G must be contained in some tree in any $2k - 1$-stretch tree cover; i.e., $\bigcup_i E(T_i) \supseteq E(G)$. Indeed, observe that if we miss an edge (x, y) we must a travel a distance of at least $(2k+1)-1$ (the girth minus one) to go from x to y, which would violate the stretch guarantee.

This implies that the tree cover must have at least $f(n,k)/(n-1)$ trees. It is known that $f(n,k) \geq n^{1+\frac{2}{3k-6}}$. A famous conjecture of Erdős states that $f(n,k) \geq n^{1+\frac{1}{k}}$, which implies the conjectured bound on tree covers. ■

It is not known whether there are constructions of tree covers that match the above lower bound; however, a slightly weaker guarantee is offered by the following theorem:

Theorem 11.4.2 (Thorup and Zwick (2001)) *For all graphs G, there is a stretch-$2k-1$ tree cover $\mathcal{F} = T_1, \ldots, T_l$ such that every vertex lies in $O(n^{1/k} \log^{1-1/k} n)$ trees.*

Note that these trees are not necessarily spanning trees of the graph; hence there is no explicit bound on the number of trees in the cover, except that each tree lies in only a small number of these trees. It turns out the this weaker guarantee suffices for most applications of tree covers, including those given at the beginning of this lecture.

Let us finish the lecture with the following theorem for planar graphs:

Theorem 11.4.3 (Gupta, Kumar and Rastogi (2001)) *Every planar graph has a stretch-3 $O(\log n)$-sized tree cover.*

Similar results were independently obtained for planar digraphs by Thorup (2001); while he does not obtain better tree covers, his constructions can be used to obtain better (i.e., stretch-$(1+\epsilon)$ for any ϵ) routing schemes and distance labelings.

Before we prove the result, let us make one final definition:

Definition 11.4.4 *The graph G has a t-*path separator *if there are t subpaths $P_1, P_2, \ldots, P_t$ in G such that*

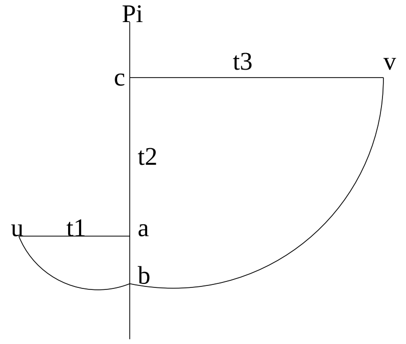
1. *p_i are* geodesics *(i.e., shortest paths in G). Formally, $P_i \subseteq E(G)$ such that if $x, y \in P_i$, then $d_{P_i}(x,y) = d_G(x,y)$*
2. *The set $S = \bigcup_i P_i$ is a 1/3-balanced separator.*

We will need the following claim:

Claim 11.4.5 (Gupta, Kumar and Rastogi (2001), Thorup (2001)) *Each planar graph has a 2-path separator.*

We only sketch the proof here, which follows the lines of the proof of Theorem 11.3.3 by Lipton and Tarjan. We can take a shortest-path tree in the planar graph, triangulate the graph, and show that there is some edge $e = (u,v)$ such that the fundamental cycle corresponding to the edge is a separator. However, this fundamental cycle consists of two shortest paths (those from u and v to $lca(u,v)$, and these are the claimed paths.

Proof of Theorem 11.4.3: We can prove in general that if G has a t-path recursive separator, then we there is a $O(t \log n)$-sized tree cover. For each path P_i in the separator, we contract the vertices of P_i, and build a shortest path tree in the resulting graph. Then, we expand again path P_i, getting a tree T_i. This T_i, contains P_i, and all the shortest paths from each node not in P_i to nodes in P_i. Now we need to prove that any other path split by P_i is stretched by a factor of at most 3. Consider the following figure.



The shortest path between vertices u and v intersects P_i at some point b. The path P' between u and v in T_i can be divided in three parts, the shortest path t_1 from u to a node a in P_i, the shortest path t_3 from v to a node c in P_i, and the shortest (and unique) path t_2 between a and c in P_i. We know that

$$d_G(u,a) \leq d_G(u,b) \quad , \quad d_G(v,c) \leq d_G(v,b)$$

Also since the segment between a and c is the shortest path we have

$$d_G(a,c) \leq d_G(a,u) + d_G(u,v) + d_G(v,c)$$

Then, for the length of the path between u and v in T_i, we have

$$\begin{aligned} d_{T_i}(u,v) &= d_G(u,a) + d_G(a,c) + d_G(v,c) \\ &\leq d_G(u,a) + (d_G(a,u) + d_G(u,v) + d_G(v,c)) + d_G(v,c) \\ &\leq 2(d_G(u,b) + d_G(b,v)) + d_G(u,v) = 3d_G(u,v) \end{aligned}$$

The proof concludes by applying recursively the same procedure in the two graphs in $G - S$. ■