8 Embeddings into Random Trees

8.1 Single Trees

Can any metric be embedded into a single tree with low distortion? Unfortunately, the answer to this question is "no". Embedding C_n into a single subtree requires distortion at least $n-1$ (deleting a single edge from the cycle yields a tree metric with distortion $n-1$). Rabinovich and Raz (1995) proved that embedding the unit weight n-cycle C_n into a tree (which is not necessarily a subtree of C_n, and may have vertices and edges not in C_n) still requires a distortion of $\Omega(n)$. We will overcome this lower bound by embedding metrics into *distributions* of trees.

Definition 8.1 *Suppose $\mathcal{G}$ is a graph family. Then $(X,d) \overset{D}{\hookrightarrow} \mathcal{G}$ means that there exists a graph $H \in \mathcal{G}$ with edge lengths such that $(X,d) \overset{D}{\hookrightarrow} (v_H, d_H)$.*

Definition 8.2 $(X,d) \overset{D}{\hookrightarrow} \mathsf{distrib}(\mathcal{G})$ *means that there exists a distribution π on the graph family $\mathcal{G}$ and an $r > 0$ such that:*

$$r \le \frac{E_{H \leftarrow \pi}[d_H(x,y)]}{d(x,y)} \le Dr$$

It is easy to see that $d_\pi(x,y) = E_{H \leftarrow \pi}[d_H(x,y)]$ is a metric (because of linearity).

8.2 Line Metrics

Assume $\mathcal{L}$ is the set of all line metrics. ($\mathcal{L}$ is equal to the set of all metrics that isometrically embed into the real line ℓ_p^1.)

Theorem 8.3 *Let $\mathcal{L}$ be the set of all line metrics. For any metric (X,d),*

$$(X,d) \overset{\log n}{\hookrightarrow} \mathsf{distrib}(\mathcal{L})$$

This follows from the following (simple) result:

Lemma 8.4 *Let $\mathcal{L}$ be the set of all line metrics. Given a metric μ,*

$$\mu \in \ell_1 \iff \mu \in \mathsf{distrib}(\mathcal{L}).$$

Proof. For one direction, assume $\mu \in \ell_1$. Then $\mu = \sum_S y_S \delta_S$, where δ_S are all elementary cut metrics. But elementary cut metrics are line metrics with two points, so μ is equivalent to a distribution over line metrics, each one having probability mass $\frac{y_{S_i}}{\sum_S y_S}$. For the other direction, note that line metrics are in ℓ_1, and hence distributions over them, which are the same as convex combinations of them, are in ℓ_1 as well. ∎

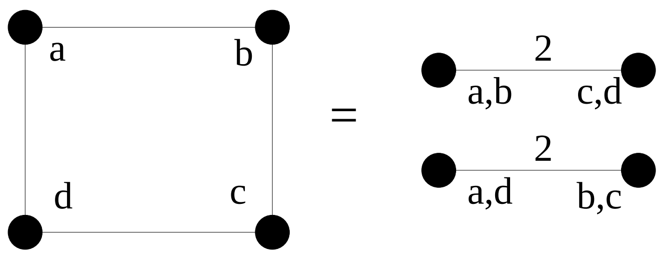

Figure 8.1: The metric d_{C_4} on the left can be thought of as a distribution over the two line metrics on the right, each having probability 1/2.

8.3 Distributions Over Trees and Dominating Trees

Through an argument similar to the one above we can see that distributions over tree metrics are equivalent to ℓ_1. Therefore, any metric embeds into a distribution over tree metrics with distortion $O(\log n)$. We will prove a stronger result due to Fakchapoenphol, Rao and Talwar (2003): any metric embeds into distributions of *dominating trees* with distortion $O(\log n)$.

A metric (X, d') *dominates* another metric (X, d) if $d'(x, y) \geq d(x, y)$ for all $x, y \in X$; hence, given a metric d, a *dominating tree* is merely a tree T such that d_T dominates d. Dominating trees are very useful in approximation and online algorithms.

We will present a series of results. The first one, due to Karp (1989), states that the n-cycle C_n embeds into a distribution of dominating trees with distortion $2(1 - \frac{1}{n})$. The next one, due to Bartal, shows that any metric embeds into distributions of dominating trees with distortion $O(\log^2 n)$. We will finish with the recent result of Fakcharoenphol et al. (2003), which shows that any metric embeds into a distribution of dominating trees with distortion $O(\log n)$.

We start with Karp's theorem:

Theorem 8.5 *Let C_n be the n-cycle with unit edge lengths. Then d_{C_n} can be embedded into a distribution of dominating trees with distortion $2(1 - \frac{1}{n})$.*

The embedding is simple: starting from C_n, delete a single edge at random to produce a tree. It is easy to see that $d_\pi(x, y) \geq d_{C_n}(x, y)$ for any x, y. If x, y are adjacent in C_n, $d_\pi(x, y) = 2(1 - \frac{1}{n})d_{C_n}(x, y)$. For non-adjacent vertices, look at the shortest path between x and y, say $x = x_0, x_1, ..., x_t = y$. Then $d_\pi(x_i, x_{i+1}) \geq 2(1 - \frac{1}{n})d_{C_n}(x_i, x_{i+1})$. The theorem follows by linearity of expectation.

Remark 8.6 *The above observation holds in general, and implies that the expansion is worst for adjacent vertices; hence we will only worry about adjacent vertices from now on.*

8.3.1 Applications

Embeddings of metrics into a distribution over dominating trees have many important application due to the fact that many problems that appear difficult on general graphs often have nice and simple solutions on tree networks. The general framework for applying the embedding-technique is, as follows. Suppose we are given a graph G and a problem for which the cost of an optimum solution can be expresses as a conic combination of pairwise distances in the graph. Examples for such problems are the Traveling Salesman Problem, Metric Labeling, Steiner Tree, Buy-at-Bulk Network Design, etc.

Further, suppose that we have an embedding with distortion D into a distribution over dominating trees, and that we can solve the problem on a tree with some approximation guarantee c, i.e., for a tree T we can output a solution S_T such that

$$\mathrm{cost}(S_T, T) \leq \mathrm{cost}(S_T^*, T) \ ,$$

where S_T^* denotes the optimum solution in T and for a solution S, and a graph H with $V(G) \subset V(H)$, $\mathrm{cost}(S, H)$ denotes the cost of the solution when distances are measured in H.

The algorithm for the graph G is as follows. We sample a random tre T according to the distribution defined by the embedding. Then we compute the approximate optimum solution on T and interpret this solution as a solution in G. Since distances in T are larger than distances in G the solution will have smaller cost in G. This means we have

$$\mathbf{E}[\mathrm{cost}(S_T, G)] \leq \mathbf{E}[\mathrm{cost}(S_T, T)] \leq c \cdot \mathbf{E}[\mathrm{cost}(S_T^*, T)] \ .$$

Let S_G^* denote the optimum solution for G. Since the distances in the tree T are (in expectation) only a factor D larger we get

$$\mathbf{E}[\mathrm{cost}(S_T^*, T)] \leq \mathbf{E}[\mathrm{cost}(S_G^*, T)] \leq D \cdot \mathrm{cost}(S_G^*, G) \ .$$

Altogether this means that the solution S_T has an expected cost that is at most a $c \cdot D$ factor larger than the cost of the optimum solution S_G^*. Note that the above technique does not only work for approximation algorithms but can also be applied in the same way for the analysis of online algorithms when the adversary does not know the random choices (i.e., the random tree T) of the algorithm (i.e., the adversary is oblivious).

8.4 Bartal's Theorem

In this section, we will prove a result due to Bartal (1996):

Theorem 8.7 *Given a metric (X, d) with diameter Δ, let $\mathcal{DT}$ be the set of all tree metrics that dominate d. Then*

$$(X, d) \xhookrightarrow{O(\log n \log \Delta)} \mathsf{distrib}(\mathcal{DT}).$$

Before we start with proving Bartal's theorem, let us recall the definition of diameter, and define the "weak" diameter:

Definition 8.8 (Diameter) *The* diameter *of a graph G (denoted by $\mathsf{diam}(G)$) is the least δ such that for all pairs of vertices $x, y \in V(G)$, we have $d_G(x, y) \leq \delta$.*

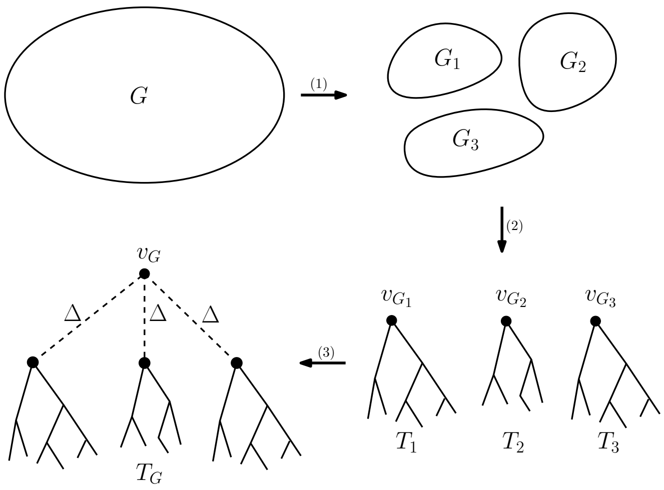

Figure 8.2: Bartal's construction: (1) decompose the graph into pieces G_i with weak diameter $\mathsf{weak}_G\mathsf{diam}(G_i) \leq \Delta/2$, (2) recusively obtain trees T_i for each of the G_i, and (3) attach them to a new vertex v_G by edges of length Δ (dashed lines).

Definition 8.9 (Weak diameter) *Given a graph G and a* subgraph $G' \subseteq G$, *the* weak diameter of G' with respect to G *(denoted by $\mathsf{weak}_G\mathsf{diam}(G')$) is the least δ such that $d_G(x,y) \leq \delta$ for all $x, y \in V(G')$.*

Given a graph H and a subgraph H', it follows that $\mathsf{weak}_H\mathsf{diam}(H') \leq \mathsf{diam}(H')$; however, these two parameters could be very far apart. (Can you give an example?)

8.5 The Embedding

Assume $G = (X, E)$ generates the metric (X, d). In Bartal's construction, we will partition G (the *level* 0 graph) into "smaller" subgraphs $G_1, ..., G_n$ (called the *level* 1 *graphs*), recursively build a tree for each of the G_i's, and then join these trees to get a tree for G.

Formally, let Δ be the diameter of G. Our construction will guarantee that the weak diameter of each of the G_i's (i.e., of the level 1 components) is smaller than $\Delta/2$. We will recursively build rooted trees (called the *level* 1 *trees*) T_i for each of the G_i's, where the root of T_i will be a vertex v_{G_i}. We will then take a fresh vertex v_G, and attach each of the v_{G_i}'s to v_G with edges of length Δ. (See Figure 8.2 for an illustration.) In the base case for the recursion, the graph has one vertex and there is nothing to be done.

Let us follow the recursion one more level: the tree for the level 1 component G_i is built in exactly the same way: G_i will be divided into *level* 2 components $G_{i_1}, G_{i_2}, ..., G_{i_k}$, and the corresponding rooted *level* 2 trees $T_{i_1}, T_{i_2}, ..., T_{i_k}$ will be built recursively, with T_{i_j} having a root vertex $v_{G_{i_j}}$. We will then create the new vertex v_{G_i} and connect each of the rooted trees to v_{G_i} with edges of length $\Delta/2$ to yield the tree T_i with root v_{G_i}.

8.6 The Analysis

Let us prove that the above embedding has the properties claimed in Theorem 8.7. To begin, let us note certain properties that can be proved by simple inductions: consider a level i component H, and a level $i+1$ component H' formed by partitioning H. Then the above algorithm ensures that the weak diameter of the level $i+1$ component H' is $\mathsf{weak}_H\mathsf{diam}(H') \leq \Delta/2^{i+1}$.

Furthermore, let T' be the level $(i+1)$ tree corresponding to H' (with root r'), and let T be the level i tree corresponding to H (with root r). Then the length of the edge (r, r') in T is $\Delta/2^i$. Furthermore, the distance of any leaf in T from the root r is at most $2\Delta/2^i$; hence, the diameter of T is at most $4\Delta/2^i$.

Lemma 8.10 *Any tree T generated as above dominates d, i.e., $d_T(x, y) \geq d(x, y)$.*

Proof. Let x, y be such that $\Delta/2^j < d_G(x, y) \leq \Delta/2^{j-1}$. Since x and y are at distance greater than $\Delta/2^j$, they cannot lie in the same level j component. Hence they were separated at some level $j' < j$, and the edges of length $\Delta/2^{j'}$ added at this level to connect their subtrees ensure that x and y are at distance at least $2\Delta/2^{j'} > d(x, y)$ in the tree. ■

We will use the following graph decomposition theorem, which will be proved in the next lecture:

Theorem 8.11 *Given a graph $G = (V, E)$ with edge lengths, and a parameter δ, there exists a procedure that deletes edges E' such that:*

1. *Each connected component C in $(V, E - E')$ has (weak) diameter smaller than δ.*
2. *$Pr[$edge e is cut$] \leq 4 \log n \times (d_e/\delta)$.*

We can now use the properties promised in Theorem 8.11 to prove Theorem 8.7.

Proof of Theorem 8.7. Fix an edge $e = (x, y)$ and consider the expected distance between x and y in the random tree created by our procedure. It is merely

$$\begin{aligned}
&E_R[d_T(x, y)] \\
&\qquad = 4\Delta \cdot \Pr[x \text{ and } y \text{ are cut in level } 0] \\
&\qquad + 2\Delta \cdot \Pr[x \text{ and } y \text{ are cut in level } 1 \mid \text{they are in the same level 1 graph}] \\
&\qquad + \ldots \cdot \ldots \\
&\qquad + \tfrac{4\Delta}{2^j} \cdot \Pr[x \text{ and } y \text{ are cut in level } j \mid \text{they are in same level } j \text{ graph }] \\
&\qquad + \ldots \cdot \ldots \ ,
\end{aligned}$$

where the sum has $\log \Delta$ terms. However, since we invoke the procedure in Theorem 8.11 with the parameter $\delta = \Delta/2^{j+1}$ in level j, we get that

$$\Pr[x \text{ and } y \text{ are cut in level } j \mid \text{ they are in same level } j \text{ graph }] \leq 4 \log n \cdot \frac{d(e)}{\Delta/2^{j+1}} \ ,$$

and hence each of the terms in the summation above can be upper bounded by

$$\frac{4\Delta}{2^j} \cdot 4 \log n \cdot \frac{d(e)}{\Delta/2^{j+1}} = 32 \log n \cdot d(x,y) \ .$$

The fact that there are $\log \Delta$ terms now implies that

$$E_R[d_T(x,y)] \leq (32 \log n \log \Delta) d(x,y) \ ,$$

which proves the theorem. ■

References

[AKPW95] Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM Journal on Computing*, 24(1):78–100, 1995.

[Bar96] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 184–193, 1996.

[Bar98] Yair Bartal. On approximating arbitrary metrics by tree metrics. In *Proceedings of the 30th ACM Symposium on Theory of Computing (STOC)*, pages 161–168, 1998.

[FRT03] Jittat Fakcharoenphol, Satish B. Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th ACM Symposium on Theory of Computing (STOC)*, pages 448–455, 2003.

[Kar98] Richard M. Karp. A $2k$-competitive algorithm for the circle. Manuscript, 1998.

[RR98] Yuri Rabinovich and Ran Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete & Computational Geometry*, 19(1):79–94, 1998.