IOE 691: Approximation Algorithms	Date: 03/06/2017

Lecture Notes: Group Steiner Tree in General Graphs

Instructor: Viswanath Nagarajan — Scribe: Xiangkun Shen

1 Group Steiner Tree

The input to the group Steiner tree (GST) problem is a graph $G = (V, E)$ with costs d_e on edges $e \in E$, groups $S_i \subseteq V$ (for $i = 1, \ldots, k$) and a root vertex r. The goal is to find minimal cost set of edges F that connects at least one $v \in S_i$ to r for each $i = 1, \ldots, k$. Recall from the last lecture:

Theorem 1.1 *There is an $O(H \log k)$-approximation algorithm for GST when the graph G is a tree of depth H.*

When the given $G = (V, E)$ is a general graph, we will show:

Theorem 1.2 *There is an $O(\log^2 n \log k)$-approximation algorithm for GST on general graphs.*

2 Tree Embedding

The idea of tree embedding is to associate (a group of) tree T with edge-costs c to the given general graph G with edge-costs d such that the shortest path distances between any pair of vertices in G is bounded by the shortest path distances in T. We use $\bar{d}(u, v)$ to denote the shortest path distance of u, v in G. Similarly, $\bar{c}(u, v)$ is defined to be the shortest path distance in T.

The intuitive idea may be to find a deterministic T and edge cost c such that $\forall u, v$, $\bar{d}(u, v) \le \bar{c}(u, v) \le \alpha \cdot \bar{d}(u, v)$. But such a deterministic tree embedding is not very useful because α has to be very large even on simple graphs. (The example is a cycle with n vertices. If remove one edge, one of the shortest path distances increases $n - 1$ times. It can be proved that the *any* deterministic tree embedding incurs loss $\alpha = \Omega(n)$.)

We can approximate the given graph much better by using a randomized collection of trees.

Definition 2.1 *Consider any graph G on vertices V and edge-costs d. Let $\mathcal{D}$ be a distribution of trees (with edge-costs) on the same vertices V. Then $\mathcal{D}$ is said to α-probabilistically approximate distances in (G, d) if:*

1. *for all $u, v \in V$ and $(T, c) \in \mathcal{D}$, $\bar{c}(u, v) \ge \bar{d}(u, v)$, and*
2. *for all $u, v \in V$, $\mathbb{E}_{(T,c)\leftarrow\mathcal{D}}[\bar{c}(u, v)] \le \alpha\bar{d}(u, v)$.*

Theorem 2.1 *[1, 2] Given an graph (G, d), there is a randomized, polynomial-time algorithm that produces a distribution $\mathcal{D}$ of trees that $\alpha = O(\log n)$ probabilistically approximate distances in (G, d). Moreover, the depth of each tree is $O(\log n)$.*

3 GST algorithm in general graphs

Using the tree embedding theorem, we have the following algorithm for GST in general graphs.

1. Randomly sample tree (T, c) from the distribution $\mathcal{D}$ of tree embedding of (G, d).
2. Run [GKR] algorithm (Theorem 1.1) on (T, c) to get F, which is a subset of edges on T.
3. Solution R is given by the union of shortest paths in G corresponding to each edge $e \in F$.

Recall Theorem 1.2, now we can prove it.

Proof: First note that R is always feasible. We want to show $\mathbb{E}_{\mathcal{D}}[d(R)] \leq \alpha\rho OPT$ where α is the factor in tree embedding and ρ is the approximation ratio on tree. Let R^* be the optimal solution in G. Let S^* be the union of all shortest paths in T corresponding to edges of R^*; note that S^* is random as it depends on T. Then we have

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}[c(S^*)] \leq \mathbb{E}_{\mathcal{D}}[\sum_{(u,v)\in R^*} \bar{c}(u,v)] &= \sum_{(u,v)\in R^*} \mathbb{E}_{\mathcal{D}}[\bar{c}(u,v)] \\ &\leq \alpha \sum_{(u,v)\in R^*} \bar{d}(u,v) \leq \alpha OPT,\end{aligned}$$

which comes from the second property of tree embedding. Since ρ is the approximation ratio on the tree instance, we have

$$\mathbb{E}_{\mathcal{D}}[c(F)] \leq \rho\mathbb{E}_{\mathcal{D}}[c(S^*)] \leq \rho\alpha OPT.$$

And by the first property of the tree embedding,

$$d(R) \leq \sum_{(u,v)\in F} \bar{d}(u,v) \leq \sum_{(u,v)\in F} \bar{c}(u,v) = c(F)$$

Take expectation of $d(R)$ over $\mathcal{D}$, we have $\mathbb{E}_{\mathcal{D}}[d(R)] \leq \mathbb{E}_{\mathcal{D}}[c(F)] \leq \alpha\rho OPT$. ∎

References

[1] Fakcharoenphol, J., Rao, S., and Talwar, K. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing* (2003), ACM, pp. 448–455.

[2] Gupta, A. Steiner points in tree metrics don't (really) help. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms* (2001), Society for Industrial and Applied Mathematics, pp. 220–227.