CMPUT 675: Approximation Algorithms **Winter 2018**

Lecture 19 (March 20, 2018): Group Steiner Tree

Lecturer: Mohammad R. Salavatipour *Scribe: Jingjiao Ni*

19.1 Group Steiner Tree

Group Steiner tree is a problem that generalizes Steiner tree.

Group Steiner Tree:

- Input: weighted $G = (V, E)$, $c_e : E \to Q^+$, root r, k subsets (groups). $g_i \subseteq V$
- Goal: Find a tree T rooted at r that connects at least one node from each g_i to r, min-cost.

However, this problem is signinifacntly more difficult to approximate. In fact, even when the input graph is a tree, the problem is as hard as set cover (Exercise!).

Theorem 1 (Halperin Kraughtganer'03) *Group Steiner trees on trees is $\Omega(\log^{2-\epsilon} n)$-hard to approx for any $\epsilon > 0$.*

In this lecture we see a polylogarithmic approximation for this problem on general graphs. Given G the first step is to use the probabilistic embeddings of metrics into tree metrics that we saw last lecture and at a loss of $O(\log n)$-loss we reduce the problem into solving the problem on trees. Then we present an algorithm for it on trees. We prove the following Theorem.

Theorem 2 (Garg/Konjevod/Ravi'00) *Group Steiner problem on trees has an $O(\log\ n \cdot log\ k)$-approx.*

We will use the following LP relaxation of the problem.

$$min \sum_e c_e \cdot x_e$$
$$\sum_{e \in \delta(S)} x_e \quad \geq \quad 1 \quad \forall S \subseteq V,\ S \text{ separates a group } g_i \text{ from } r.$$

Our rounding algorithm outputs a set of edges E' such that:

1. $E[cost(E')] \leq OPT$.
2. For any g_i, $\Pr[E' \ covers\ g_i] \geq \frac{1}{\log |g_i|}$.

We repeat this algorithm $O(\log n \cdot \log k)$ times and take the union. It is easy to see that the expected cost of the solution is at most $O(\log n \cdot \log k)$ times the optimum. The probability that any group g_i is *not* connected to the root after all these iterations is bounded by $(1 - \frac{1}{n})^{O(\log n \log k)} \approx \frac{1}{k^2}$. Using union bound the probability that any group is not connected is very small.

For each edge e, let $p(e)$ be the parent of e. Suppose we first try a simple randomized rounding like set cover; i.e. pick each edge e with probability x_e. It will be easy to see that the expected cost of the solution is the same as optimum LP, however, it is far from satisfying the second condition. For example, consider the simple tree with only two leaves u and v and there are two paths of length $n/2$ from r to each of these. Let $k = 1$ and $g_1 = \{u, v\}$. Consider the LP solution in which all x_e's are 1/2. Then the probability that either of u or v are connected to r is at most $2 \times 2^{-n/2}$.

We modify the rounding in the following way:

Let $x'_e = 1$ with prob $\frac{x_e}{x_{p(e)}}$; if $p(e)$ doesn't exist $x'_e = 1$ with prob x_e.

We pick each edge e if $x'_e = 1$ and all its ancestors e' has $x'_{e'} = 1$.

Lemma 1 $\Pr[e \textit{ is picked}] = x_e$.

Proof. Consider any edge e. Say it has i ancestor edges. Then the probability that it gets picked is the probability that e and all its ancestors are marked. Therefore,

$$Pr[e \textit{ is picked}] = \frac{x_e}{x_{p(e)}} \cdot \frac{x_{p(e)}}{x_{p(p(e))}} \dots \; x_{e_{p^i(e)}} = x_e$$

■

Thus, the expected cost of the solution generated is the same as optimum LP. We have to prove that condition 2 holds. Given x, we build another LP solution $\widetilde{x}$ such that:

$$\forall g_i, \; \Pr[g_i \textit{ is not covered by } x] \leq \Pr[g_i \textit{ is not covered by } \widetilde{x}] \leq 1 - \frac{1}{\log |g_i|}.$$

Lemma 2 *If* $\widetilde{x}_e \leq x_e$ *for all* $e \in E$ *then* $\Pr[g_i \textit{ is not covered by } x] \leq \Pr[g_i \textit{ is not covered by } \widetilde{x}]$.

Proof. Prove lemma 2 by induction on the number of edges where x & $\widetilde{x}$ differ. Suppose that x & $\widetilde{x}$ differ on only one edge $e \; : \; x_e > \widetilde{x}_e$. We use T_e to denote the subtree below e. Note that for groups outside T_e, probabilities for x & $\widetilde{x}$ are the same. Let A be the event that a vertex from group g_i in T_e is covered.

$$\begin{aligned} \Pr[\overline{A}] &= \text{Pr[none of vertices of } g_i \text{ in } T_e \text{ are covered]} \\ &\leq (1 - x_e) + x_e \cdot \Pr[\overline{A} | e \text{ is picked}] \\ &= 1 - x_e(1 + \Pr[\overline{A} | e \text{ is picked}]). \end{aligned}$$

So if x_e increases then $\Pr[\overline{A}]$ decreases. Hence for any group, the probability that g_i is covered in T_e in x is larger than that of x'. ■

How to build $\widetilde{x}$ from x, fix group g_i

1. Delete all edges e incident to leaves from nodes not in g_i. Also remove unnecessary/irrelevant edges.
2. Reduce x values (if needed) so that it is minimally feasible. $\rightarrow$ 1 flows from g_i to r.

3. Round down x values to the nearest power of $\frac{1}{2}$. $\rightarrow \geq \frac{1}{2}$ flow from g_i to r.
4. Delete all edges with x values $< \frac{1}{4|g_i|}$; Therefore the remaining flow from g_i to r is at least $\frac{1}{2} - |g_i| \cdot \frac{1}{4|g_i|} \geqslant \frac{1}{4}$
5. $\forall e$ if $x_e = x_p(e)$ and $p(e)$ has only one child then merge the edges. If $p(e)$ has at least two children, then contract e such that children of e become children of $p(e)$. (Note we can do this because when we pick $p(e)$ then x_e is picked with prob 1).

Lemma 3 *The height of the new tree is* $O(\log |g_i|)$.

Proof. At each node, x values decrease by factor ≥ 2 going down and they are all $1 \ldots \frac{1}{4 \log |g_i|} \rightarrow \leq \log(4|g_i|)$ steps from root to any leaf. ∎

Suppose U is a set and $S_1, S_2, \ldots$ are some subsets of U. Also, assume we generate S' by randomly and independently adding each $e \in u$ to S'. Let event $\mathcal{E}_i$ be the event that $S_i \subseteq S'$, $\mu = \sum_i \Pr[\mathcal{E}_i]$ and $\Delta = \sum_{i \sim j} \Pr[\mathcal{E}_i \cap \mathcal{E}_j]$ where $i \sim j$ means $\mathcal{E}_i$ and $\mathcal{E}_j$ are dependant.

Lemma 4 (Janson's Inequality) *If* $\mu \leq \Delta$ *then* $\Pr[\cap_i \overline{\mathcal{E}}_i] \leq e^{\frac{-\mu^2}{2\Delta}}$.

In our setting, let $U = E$, $\forall v \in g_i$, S_v is the set of edges on the path from v to r and good event $\mathcal{E}_v$ is when all edges of S_v are picked. S' is the set of edges we pick in our randomized rounding.

Note that $\Pr[\mathcal{E}_v] = \tilde{x}_{p(v)}$ and therefore $\frac{1}{4} \leq \sum_{v \in g_i} \Pr[\mathcal{E}_v] \leq 1$. Also $\Pr[g_i \textit{ not covered}] = \Pr[\cap_{v \in g_i} \overline{\mathcal{E}}_v]$

Lemma 5 $\Delta = \sum_{u \sim v} \Pr[\mathcal{E}_u \cap \mathcal{E}_v] \in O(\log n)$.

First, assume this lemma. Then:

$$\Pr[\cap_v \overline{\mathcal{E}}_v] = e^{\frac{-O(1)}{O(\log |g_i|)}} \approx 1 - \frac{1}{\log |g_i|}$$

and this completes the proof. So it remains to prove the lemma.

Proof. Suppose that the height of the tree is H. Let $u \in g_i$. We prove that

$$\Delta_u = \sum_{v \sim u} \Pr[\mathcal{E}_u \cap \mathcal{E}_v] \leq O(H) \cdot \widetilde{x}_u.$$

Noting that $\sum_{u \in g_i} \widetilde{x}_u \leq 1$ and $H \in O(\log n)$ implies the bound claimed. Let e be the edge going up the least common ancestor (LCA) of two vertices u, v and let $c(e)$ be the child of LCA towards v. Recall that $\Pr[\mathcal{E}_u] = \widetilde{x}_u$. Also:

$$\Pr[\mathcal{E}_v | \mathcal{E}_u] = \frac{\widetilde{x}_v}{\widetilde{x}_{p(v)}} \cdot \frac{\widetilde{x}_{p(v)}}{\widetilde{x}_{p(p(v))}} \cdots \frac{\widetilde{x}_{c_e}}{\widetilde{x}_e} = \frac{\widetilde{x}_v}{\widetilde{x}_e} \Pr[\mathcal{E}_u \cap \mathcal{E}_v] = \widetilde{x}_u \cdot \frac{\widetilde{x}_v}{\widetilde{x}_e}$$

Sum over all v's that have e as L.C.A. with u and noting that $\sum_{such\ v's} \widetilde{x}_v \leq O(\widetilde{x}_e)$, since the flow of all of those vertices goes up from e towards r, therefore:

$$\sum_{such\ v's} \Pr[\mathcal{E}_u \cap \mathcal{E}_v] \leq O(\widetilde{x}_u).$$

∎