CMPUT 675: Approximation Algorithms **Fall 2014**

Lecture 17 (Oct 15): Steiner Forest

Lecturer: Zachary Friggstad *Scribe: Yifeng Zhang*

17.1 Steiner Forest

Definition 1 *In the* Steiner Forest *problem, we are given an undirected graph $G = (V, E)$ with edge cost $c_e \geq 0$, $e \in E$. Additionaly, we are given of pairs of vertices $(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)$. The goal is to find the cheapest forest $F \subseteq E$ such that for each $1 \leq i \leq k$, s_i and t_i lie in the same component of (V, F)*

Consider the function $f : 2^v \to \{0, 1\}$ where for any $S \subseteq V$ we have $f(S) = 1$ if and only if $|S \cap \{s_i, t_i\}| = 1$ for some $1 \leq i \leq V$. This notation will be useful when we consider a generalization of the Steiner Forest problem next lecture.

Say that $F \subseteq E$ is feasible if and only if $\delta(S) \cap F \neq \emptyset$ for every $S \subseteq V$ with $f(S) = 1$. Note that F is feasible if and only if s_i and t_i lie in the same component of (V, F) for every $1 \leq i \leq k$.

17.2 An LP Relaxation

Recall that $\delta(S)$ denotes the set of all edges in E that have exactly one endpoint in S.

$$\begin{aligned} \text{minimize :} \quad & \sum_{e \in E} c_e \cdot x_e & & & \\ \text{subject to :} \quad & \sum_{e \in \delta(S)} x_e \geq 1 & \text{for any } S \subseteq V \text{ with } f(S) = 1 & & \textbf{(LP-Primal)} \\ & \mathbf{x} \geq 0 & & & \end{aligned}$$

Though this LP admits an efficient separation oracle (simply check that the minimum $\mathbf{x}$-capacity $s_i - t_i$ cut is at least 1 for each $1 \leq i \leq k$), we will not need to solve it directly in our approximation. Rather, we will use the *primal-dual* method: simultaneously build an integer primal solution and a feasible dual solution in a careful way to ensure that their costs are close to each other.

The dual of (**LP-Primal**) is the following LP that has a variable y_S for every $S \subseteq V$ with $f(S) = 1$.

$$\begin{aligned} \text{maximize :} \quad & \sum_{S \subseteq V : f(S) = 1} y_S & & & \\ \text{subject to :} \quad & \sum_{\substack{S \subseteq V : f(S) = 1 \\ e \in \delta(S)}} y_S \leq c_e & \text{for each } e \in E & & \textbf{(LP-Dual)} \\ & \mathbf{y} \geq 0 & & & \end{aligned}$$

Primal-dual algorithms are guided by complementary slackness conditions. Let's inspect these conditions for (**LP-Primal**) and (**LP-Dual**), where we use $x_e = 1$ instead of $x_e > 0$ because we are looking for an integer primal solution.

1. $x_e = 1 \Rightarrow \sum_{S:f(S)=1, e\in\delta(S)} y_S = c_e$ for each $e \in E$
2. $y_S > 0 \Rightarrow \sum_{e\in\delta(S)} x_e = 1$ for each $S \subseteq V$ with $f(S) = 1$

Of course, we cannot satisfy all of these conditions simultaneously; we want an *integer* primal solution. Something has to be relaxed.

The algorithm below ensures the first complementary slackness condition hold: an edge is not "bought" unless its dual constraint is tight. However, it only satisfies a relaxed version of the second constraints.

First note that if we could ensure the edges we use F satisfy $|\delta(S) \cap F| \leq 2$ for every S with $y_S > 0$, then this is a 2-approximation (c.f. previous lecture). However, even this is difficult to enforce. What the algorithm does is ensure that the constructed primal and dual solutions have these second conditions hold, in some appropriate sense, *on average*. This will be explained in more detail soon.

The following algorithm is our primal-dual approximation for Steiner Forest. For a given feasible dual solution, say that an edge e is *tight* if the dual constraint for that edge holds with equality. The Δ_i variables are only used to help the analysis.

Algorithm 1 Steiner Forest Approximation

$F_1 \leftarrow \emptyset$
$\mathbf{y} \leftarrow \mathbf{0}$
$i \leftarrow 1$ (iteration counter)
while F_i is not feasible **do**
  Let $\mathcal{C}_i$ be the set of components S of (V, F_i) with $f(S) = 1$
  Raise all y_S, $S \in \mathcal{C}_i$ uniformly until some edge e becomes tight
  $\Delta_i \leftarrow$ the amount we raised each y_S
  Let $F_{i+1} \leftarrow F_i \cup \{e\}$
  $i \leftarrow i + 1$
end while
$F \leftarrow F_i$
while there is some $e \in F$ s.t. $F - \{e\}$ is feasible **do**
  $F \leftarrow F - \{e\}$
end while
return F

Note that throughout the execution of the algorithm that $\mathbf{y}$ is feasible. This is because we stop raising y_S values when some edge goes tight to ensure that no dual constraint is violated. Also, if multiple edges go tight then pick any one of them to be e in the iteration.

To implement this efficiently, we do not have to explicitly initialize each entry of $\mathbf{y}$ to 0. Rather, only keep track of the subsets S with $y_S > 0$. Each iteration of the first loop has some edge go tight, so there are at most $m = |E|$ of these iterations. Each iteration raises at most $n = |V|$ variables y_S, so the total number of nonzero variables is at most $n \cdot m$. Clearly every iteration can be performed in polynomial time.

The second "pruning" loop is necessary because the cost of F can be huge otherwise. Consider the following simple example with $V = \{s, t, v_1, \ldots, v_{n-2}\}$ and edges $E = \{(s,t), (s,v_1), (s,v_2), \ldots, (s,v_{n-2})\}$ where $c_{(s,t)} = 3$ and $c_{(s,v_i)} = 1$ for every $1 \leq i \leq n-2$. The only pair is the (s,t) pair. All edges (s,v_i) will go tight before the

(s,t) edge, so the cost of the tight edges before pruning is $n-1$ whereas the optimum solution has cost 3. The pruning phase will discard all edges except the (s,t) edge.

Our main result is the following.

Theorem 1 *Algorithm 1 returns a solution F with cost at most $2 \cdot OPT_{LP}$.*

To show this, we first note the following claim which will be proven in the next lecture. This is the averaging argument alluded to above that will be used in place of the relaxed complementary slackness condition $y_S > 0 \Rightarrow |\delta(S) \cap F| \leq 2$.

Claim 1 *Consider any iteration i and let F be the final set of returned edges. We have $\sum_{S \in \mathcal{C}_i} |F \cap \delta(S)| \leq 2|\mathcal{C}_i|$, i.e. the average degree of the "active" sets in iteration i is at most 2.*

Proof of Theorem 1. We emulate the proof of why relaxed complementary slackness conditions ensure approximately optimal solutions while making appropriate adjustments for the averaging argument.

Let $\overline{\mathbf{x}}$ be the integer solution where $\overline{x}_e = 1$ for $e \in F$ and $\overline{x}_e = 0$ for $e \notin F$.

$$\begin{aligned} \text{cost}(F) &= \sum_{e \in F} c_e \\ &= \sum_{e \in E} c_e \cdot \overline{x_e} \\ &= \sum_{e \in \delta(S)} \left(\sum_{\substack{S: f(S)=1 \\ e \in \delta(S)}} y_S \right) \overline{x_e} \end{aligned} \tag{17.1}$$

The last equality is by the fact that the dual constraint for each $e \in F$ is tight under $\mathbf{y}$. Continuing, we rearrange (17.1) to get

$$\sum_{S: f(S)=1} y_S \left(\sum_{e \in \delta(S)} \overline{x}_e \right) = \sum_{S: f(S)=1} y_S \cdot |\delta(S) \cap F|. \tag{17.2}$$

We break this last sum into the iterations. That is, in iteration i each $y_S, S \in \mathcal{C}_i$ is raised by Δ_i so the total amount this iteration contributes to expression (17.2) is exactly $\sum_{S \in \mathcal{C}_i} \Delta_i \cdot |\delta(S) \cap F|$. Summing over all iterations and using Claim (1), we see

$$\begin{aligned} \sum_{S: f(S)=1} y_S \cdot |\delta(S) \cap F| &= \sum_i \Delta_i \sum_{S \in \mathcal{C}_i} |\delta(S) \cap F| \\ &\leq 2 \cdot \sum_i \Delta_i |\mathcal{C}_i| \\ &= 2 \cdot \sum_{S: f(S)=1} y_S. \end{aligned}$$

The final equality holds because each y_S value is equal to the sum of the Δ_i values over all iterations where $S \in \mathcal{C}_i$.

Finally, because $\mathbf{y}$ is a feasible dual solution then by weak duality we have $2 \cdot \sum_{S: f(S)=1} y_S \leq OPT_{\text{LP}}$. That is, the cost of F is at most $2 \cdot OPT_{\text{LP}}$. ■

This algorithm, along with generalizations that are discussed next lecture and in Assignment 3, are due to Goemans and Williamson (see [GW95] for the Steiner Forest approximation). This is the best-known approximation for Steiner Forest and it is a major open problem to find a better than 2-approximation. This may be possible; in the special case of Steiner Tree that we have a $\ln(4) + \epsilon$ approximation for any constant $\epsilon > 0$.

References

- GW95 M. X. Goemans and D. P. Williamson, A general approximation technique for constrained forest problems, SIAM Journal on Computing, 24, 296–317, 1995.