CMPUT 675: Approximation Algorithms	Winter 2005
Lecture 16: Mar 11	
Lecturer: Mohammad R. Salavatipour	*Scribe: Jianjun Zhou*

16.1 Steiner Forest (continue)

Recall the definition of Steiner forest problem:

Input: Given undirected graph $G(V, E)$, capacities $C : E \to Q^+$, and a collection of disjoint subsets $S_1, \cdots, S_k$, where each $S_i \subseteq V$.

Question: Find a min cost subgraph such that the vertices of each S_i are in one connected component.

We defined the connectivity requirement function r for any two vertices u and v in graph G as:

$$r(u,v) = \begin{cases} 1 & \text{if } u,\ v \text{ are in the same set } S_i; \\ 0 & \text{otherwise.} \end{cases}$$

Recall that for a set $S \subseteq V$, $\delta(S)$ is the set of edges with exactly one end-point in S. Then the minimum number of edges that must cross cut S is :

$$f(s) = \begin{cases} 1 & \exists u \in S, v \in \bar{S}, r(u,v) = 1; \\ 0 & \text{otherwise.} \end{cases}$$

From these definitions, we obtained the following primal and dual LP's for the (relaxation of) Steiner forest problem:

Primal:

$$\begin{array}{ll} \text{minimize} & \sum C_e x_e, \\ \text{such that} & \sum_{e \in \delta(S)} x_e \geq f(S) \text{ for all } S \subseteq V, \\ & x_e \geq 0. \end{array}$$

Dual:

$$\begin{array}{ll} \text{maximize} & y_s \sum f(S), \\ \text{such that} & \sum_{S: e \in S} y_s \leq C_e, \forall e \in E, \\ & y_s \geq 0. \end{array}$$

Definition 16.1 *In the LP formulation of the Steiner Forest problem, we say:*

- *Edge e feels dual y_s if $e \in \delta(S)$ and $y_s > 0$;*
- *Edge e is tight if the total amount of dual that it feels is C_e.*

The algorithm will have a primal-dual schema. We start with a primal and dual empty solution. Of course, the primal is infeasible and the dual is not optimal. At each iteration, we choose an unsatisfied constraint of the primal and raise the corresponding dual variables until some edge goes tight. Then we pick tight edges and continue. Here we state the primal and dual (relaxed) conditions:

Primal Condition: If $x_e > 0$, then $\sum_{S:e\in\delta(S)} y_s = C_e$. In other words, we only pick tight edges.

Relaxed Dual Condition: If we could prove that the following dual condition holds then it would yield a 2-approximation algorithm: for each set $S \subseteq V$, if $y_s > 0$, then $\sum_{e\in\delta(S)} x_e \leq 2f(S)$. In other words, for every cut S which has raised dual variable, the number of edges picked from that cut is at most 2. Unfortunately we cannot prove this condition. However, we can prove that *on average* the number of edges picked from every cut S with $y_S > 0$ is at most 2. This will be sufficient to yield a 2-approximation algorithm.

In the (informal) description of the algorithm above, we said that we find an unsatisfied constraint from the primal LP and raise the dual values. But there are exponentially many constraints. How do we choose the dual variables to be raised? We focus only on *minimal* unsatisfied cuts:

Definition 16.2 *Given an assignment to variables x_e's and y_S's:*

- *A set $S \subseteq V$ is unsatisfied if $f(S) = 1$ and no edge from $\delta(S)$ is picked.*
- *Active Set: a minimal unsatisfied set (with respect to inclusion).*

The following theorem characterizes active sets.

Lemma 16.3 *A set is active iff it is a connected component in the currently picked forest and $f(S) = 1$.*

Proof: For any active set S, by definition, $f(S) = 1$. Suppose S is not a connected component. There are two cases.

Case 1: S is within some connected component C_1 ($S \neq C_1$). Since S is connected with the other part of C_1, there is at least one edge in $\delta(S)$ is picked. So S is satisfied, which contradicts the definition of active sets.

Case 2: S contains more than one connected component. Since $f(S) = 1$, there is at least one vertex in S that needs to be connected to some vertex outside S. Denote this vertex by u. Let the connected component in S that contains u be C_1. By definition, $f(C) = 1$. Since C_1 is a connected component, there is no edge from $\delta(C_1)$ is picked. So C_1 is unsatisfied, which means that S is not minimal. This also contradicts the definition of active set.

In either case, we get a contradiction. So the lemma is true. ■

Steiner Forest Algorithm:

$F \leftarrow \emptyset$; for each $S \subseteq V$, $y_s \leftarrow 0$;

while there is an unsatisfied set do

  simultaneously raise y_s for each active set S until some edge e becomes tight;

  $F \leftarrow F \cup \{e\}$;

return $F' = \{e \in F | F - e \text{ is not feasible}\}$.

The figure below shows an example run of the algorithm. Suppose we have two disjoint subset $S_1 = \{u, v\}$ and $S_2 = \{s, t\}$. At the beginning of the algorithm, u, v, s, t are four active sets, each of which contains one vertex only. The algorithm raises their y_s values simultaneously, and stops at the value of 6 when edge ua and bv are tight. ua and bv are added to F. Then the algorithm find the next layer of active set, raise their y_s, and so on. The bold edges are added to F in the while loop. At the end, all edges in F except the redundant edge ua are added to F' and returned.

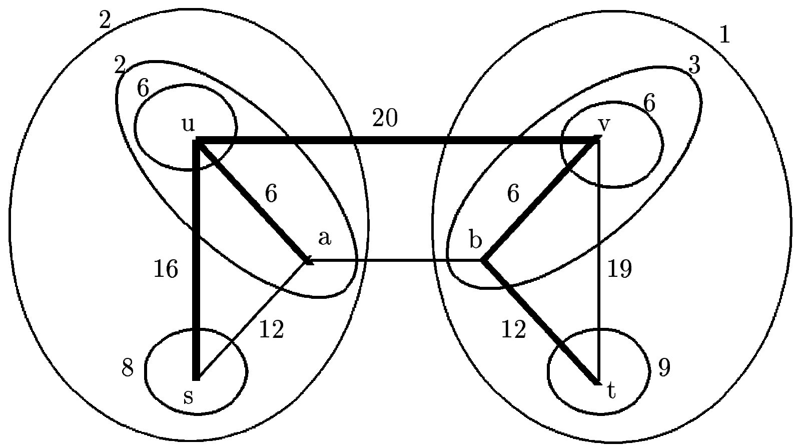

Figure 16.1: An example run of the algorithm

First we show that F' and y are feasible primal and dual solutions. It is easy to see that since we never over-pack an edge, no edge becomes over-tight. Therefore, y is a feasible solution. Before the last step, F satisfies all the connectivity requirements. Furthermore, in each iteration, only the dual variables of the cuts that are running between different connect components are raised. Therefore, F will be a forest. Also, at the last step, only redundant edges are removed. So F' is a feasible solution.

Definition 16.4 *$deg_{F'}(C)$ is the number of edges that leave C in F'.*

Lemma 16.5 *if C is a connected component in any iteration and $f(C) = 0$, then $deg_{F'}(C) \neq 1$, i.e. $deg_{F'}(C) = 0$ or $deg_{F'}(C) \geq 2$.*

Proof: Suppose $deg_{F'}(C) = 1$ and let e be the unique edge of F' coming out of C, since e is not redundant, then there exists a pair u, v with $r(u, v) = 1$ and $u \in C$ and $v \in \bar{C}$. Therefore, $f(C) = 1 \neq 0$, which contradicts the condition of the lemma. ■

Lemma 16.6 $\sum_{e \in F'} C_e \leq 2 \sum_{S \subseteq V} y_s$.

Proof:

$$\begin{aligned}
\sum_{e \in F'} C_e &= \sum_{e \in F'} \left(\sum_{S: e \in \delta(S)} y_s \right) && \text{(Edges picked are tight)} \\
&= \sum_{S \subseteq V} \left(\sum_{e \in F' \cap \delta(S)} y_s \right)
\end{aligned}$$

$$= \sum_{S \subseteq V} deg_{F'}(S) \cdot y_s$$

We need to show: $\sum_{S \subseteq V} deg_{F'}(S) y_s \leq 2 \sum_{S \subseteq V} y_s$.

Let Δ be the amount to which active sets were raised in one iteration. We show $\Delta \times \sum_{S \text{ is active}} deg_{F'}(S) \leq 2 \times \Delta \times$ (number of active sets).

Consider the graph, called H, on the same vertex set and edge set as F'. For every set of vertices of a connected component with respect to F, contract all those vertices in H into one single (big) vertex. Call this new graph H'. Note that H' is a forest and that the degree of each (big) vertex in H' is the same as the degree of corresponding set of vertices that are contracted. Each vertex of H' that corresponds to an active set has non-zero degree. By Lemma 16.5, the degree of every vertex in H' that corresponds to a non-active set is at least 2. Also, because H' is a forest, its average degree is at most 2. Therefore, the average degree of nodes of H' that correspond to active sets is at most 2. Therefore, $\Delta \times \sum_{S \text{ is active}} deg_{F'}(S) \leq 2 \times \Delta \times$ (number of active sets), and

$$\sum_{e \in F'} C_e = \sum_{S \subseteq V} deg_{F'}(S) y_s \leq 2 \sum_{S \subseteq V} y_s.$$

■

Theorem 16.7 *This is a 2-approximation algorithm.*

Proof: As we mentioned earlier, F' and y are primal and dual feasible solutions. By Lemma 16.6 their value is within a factor of 2 from each other. This implies that the value of F' within a factor at most 2 of the optimal (fractional) solution. ■