I give an account of the paper [HJ06] on the prize-collecting generalized Steiner tree problem.

1 A Useful Lemma

Lemma 1 *Let F be a tree with at least two nodes whose nodes are labelled red or black. All leaves are assumed to be red. Then*

$$\sum_{v \text{ is red}} \deg_F(v) \leq 2 \cdot \textit{number of red nodes}.$$

Proof: Let r and b be the number of red and black nodes, respectively. The sum of the degrees of all nodes is $2(b+r)-2$. The sum of the degrees of the black nodes is at least $2b$. Therefore the sum of the degrees of the red nodes is at most $2r-2$. ∎

2 Generalized Steiner Tree Problem with Penalties

We are given an undirected graph $G=(V,E)$ with positive edge weights $c : E \to \mathbb{R}_{>0}$. We are also given subsets T_1, T_2, to T_k of V, called demands. Each subset T has an associated penalty $p_T \in \mathbb{R}_{>0}$.

Let P be a partition of V and let T be a subset of V. We say that P splits T if T is not contained in any block of P. We write $P|T$ and use $w(P)$ to denote the sum of the penalties of the demands that are split by P, i.e,

$$w(P) = \sum_{T;P|T} p_T.$$

A forest F induces a partition; the blocks of the partition are the connected components of F. We say that F splits T ($F|T$) if the induced partition splits T.

A partition P refines a set S ($P \sqsubseteq S$) if S is a union of blocks of P.

A set S is unsaturated if it splits some T_i, i.e., $\emptyset \neq S \cap T_i \neq T_i$ and saturated otherwise. Define $f(S)=1$ for unsaturated sets and $f(S)=0$ for saturated sets.

The cost of a solution F is the cost of the edges in F (denoted $c(F)$) plus the sum of the penalties of the demands T_i that are split by F (denoted $w(F)$). In the ILP formulation we have a decision variable x_e for each edge e and a decision variable z_P for each partition of V. The goal is to minimize

$$\sum_e c_e x_e + \sum_P w(P) z_P$$

subject to

$$x(\delta(S)) + \sum_{P;P\sqsubseteq S} z_P \geq f(S) \quad \text{for all } S.$$

Lemma 2 *The ILP captures our problem.*

Proof: Let F be an optimal solution. Let $x_e = 1$ iff $e \in F$ and let P_F be the partition induced by F. Set $z_P = 1$ iff $P = P_F$. This is a feasible solution of the ILP with objective value equal to the cost of F. Thus the optimal objective value of the ILP is at most the cost of an optimal solution to our problem.

Consider any optimal solution to the ILP and let F be the set of edges picked. We may assume

- there is at most one P with $z_P = 1$. Assume $z_P = z_Q = 1$ and let $R = P \sqcap Q$ (the blocks of R are the nonempty intersections of blocks of P and blocks of Q. Change z_P and z_Q to 0 and set $z_R = 1$. This does not increase the cost of the solution.
- All edges of F are essential, i.e., removal of any edge will split some T which is realized by F. Otherwise, we could reduce the cost of the solution.
- If $z_P = 0$ for all P, F realizes all T's and hence the objective value of the LP is equal to the cost of F.
- Assume $z_P = 1$ for some P and let T be any demand that is not realized by F. Let S be a connected component of F that contains some but not all terminals in T. Then P must refine S and hence P splits T. Thus $w(P) \geq w(F) = w(P_F)$. Thus we may assume that $P = P_F$.

We conclude that an optimal solution to the ILP is of the form (F, P_F). ∎

The dual of the linear programming relaxation has one nonnegative variable y_S for each set. The goal is to maximize

$$\sum_S y_S$$

subject to the edge constraints

$$\sum_{S; e \in \delta(S)} y_S \leq c_e \quad \text{for all edges } e$$

and the partition constraints

$$\sum_{S; P \sqsubseteq S} y_S \leq w(P) \quad \text{for all partitions } P.$$

Let $y(P) = \sum_{S; P \sqsubseteq S} y_S$.

The primal-dual algorithm operates in two phases. In the first phase, it constructs a dual solution y and a forest F. This forest is then pruned to the solution F' by deleting edges. We say more about the pruning below.

In the first phase, we have a current set F of edges. Initially, $F = \emptyset$ and $y_S = 0$ for all S. A *current set* is a component with respect to the current edge set F. A current set S is active if there is no tight partition P that refines S and inactive otherwise. A saturated set S is inactive as the partition $\{S, V \setminus S\}$ is tight.

We raise the duals of all active sets until either an edge constraint or a partition constraint becomes tight.

Assume the edge constraint for edge $e = (u,v)$ becomes tight, where u and v lie in distinct current sets say C_u and C_v, at least one of them active. We add e to F. C_u and C_v cease to be current sets and $C_u \cup C_v$ becomes current. It is active if unsaturated. The y-value of the new current set is zero at this point.

Assume a partition constraint for P becomes tight. All active sets that are refined by P are declared inactive and any demand T that is split by P is marked as "penalty accounted for".

Lemma 3 *If partitions P and Q are tight then $P \sqcup Q$ and $P \sqcap Q$ are also tight. The blocks of $P \sqcap Q$ are the non-empty intersections of blocks of P and Q and the blocks of $P \sqcup Q$ are the smallest sets (with respect to set inclusion) that are unions of blocks of P and blocks of Q.*

Proof: We show

$$y(P \sqcup Q) + y(P \sqcap Q) \geq y(P) + y(Q) = w(P) + w(Q) \geq w(P \sqcap Q) + w(P \sqcup Q).$$

For the first inequality we observe that any set that is refined by P or Q is refined by $P \sqcap Q$ and that any set that is refined by P and Q is refined by $P \sqcup Q$. For the second inequality, we observe that any T that is split by $P \sqcap Q$ is split by either P or Q and that any T that is split by $P \sqcup Q$ is split by P and Q. ∎

It is time that we detail the pruning step.

Lemma 4 *At the end of phase I: If T is not marked as "penalty accounted for", T is realized by F.*

Proof: A demand T is marked if T is split by a tight partition. So if T is not marked it is not split by any tight partition. When phase I stops, all current sets are inactive and hence refined by a tight partition. Thus no current set splits T and hence T is realized by F. ∎

We set

$$F' = \{\, e \in F;\ e \text{ is necessary to realize an unmarked demand} \,\}.$$

We next bound the cost of F'. We show that the cost of the edges in F' is bounded by twice the sum of the dual variables and that the penalties incurred by F' is bounded by once the sum of the dual variables.

Lemma 5 $w(F') \leq \sum_S y_S$.

Proof: By the preceding Lemma all unmarked T are realized by F'. Hence $w(F')$ is bounded by the penalties of the marked demands.

If a demand is marked it is split by a tight partition. Let P_1, $\ldots$, P_k be the tight partitions. Then $P = P_1 \sqcap P_2 \ldots \sqcap P_k$ is tight. P splits every demand that is split by some P_i and hence splits all marked demands. Thus $w(F') \leq w(P)$.

Finally, observe that $w(P) = y(P) \leq \sum_S y_S$. ∎

Lemma 6 $c(F') \leq 2\sum_S y_S$.

Proof: Since all edges in F' are tight, we have

$$\sum_{e \in F'} c_e = \sum_{e \in F'} \sum_{S; e \in \delta(S)} y_S = \sum_S y_S |F' \cap \delta(S)|.$$

We show $\sum_S y_S |F' \cap \delta(S)| \leq 2\sum_S y_S$. Consider any iteration of the algorithm with $\varepsilon > 0$. In this iteration the left-hand side grows by $\varepsilon \cdot \sum_{S \text{ is active}} |F' \cap \delta(S)|$ and the right-hand side grows by $\varepsilon \cdot$ number of active sets. So it suffices to show

$$\sum_{S \text{ is active}} |F' \cap \delta(S)| \leq \text{number of active sets.}$$

Consider the following auxiliary forest. The vertices are the current sets at the beginning of the iteration and the edges are the edges of F' with endpoints in different current sets.

Lemma 7 *In this graph, the degree of any inactive set is different from one.*

Proof: An inactive set S is refined by a tight partition and hence any demand T that is split by S is marked. If the degree of S would be one, removal of the single edge incident to S would not split any unmarked demand and hence the edge would not be in F'. Thus the degree of any inactive set is either zero or at least two. ∎

We can now apply the red-black lemma to each tree (in the auxiliary forest) with at least two vertices. Active sets are red and inactive sets are black. ∎

References

[HJ06] Mohammad Taghi Hajiaghayi and Kamal Jain. The prize-collecting generalized steiner tree problem via a new approach of primal-dual schema. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 631–640, New York, NY, USA, 2006. ACM.