Lecture 29 (Nov 17 & 19): Bounded-Degree Spanning Trees

Lecturer: Zachary Friggstad *Scribe: Zachary Friggstad*

29.1 The Spanning Tree Polytope

Let $G = (V, E)$ be an undirected graph. Consider the following polytope over variables $x_e, e \in E$. For a set $S \subseteq V$ we let $E(S) = \{(u, v) \in E : u, v \in S\}$ and for a set $F \subseteq E$ we let $x(F) = \sum_{e \in F} x_e$.

$$\begin{array}{rcll} x(E(S)) & \leq & |S| - 1 & \text{for each } S \subseteq V, |S| \geq 2 \\ x(E) & = & |V| - 1 & \\ x & \geq & 0 & \end{array} \qquad \textbf{(LP-Span)}$$

We show how to separate over the constraints of this polytope and that the extreme points are precisely the $\{0, 1\}$-integer solutions corresponding to spanning trees.

Lemma 1 *There is a polynomial-time separation oracle for the constraints of (***LP-Span***).*

Proof. Let $\overline{x}$ be a proposed solution such that $\overline{x} \geq 0$ and $\overline{x}(E(V)) = |V| - 1$ (i.e. we checked them already).

Try all pairs of vertices $u, v \in V$. The idea is that we are guessing $u \in S, v \notin S$ for some set $S \subseteq V$ whose corresponding LP constraint is violated. Consider the directed graph $H(v) = (V, E')$ with edge capacities $z_e, e \in E'$ where E' consists of the following directed edges.

- For each $e = (a, b)$ edge in the original graph G, add both directed copies $(a, b), (b, a)$ to E' each and set $z_{(a,b)} = z_{(b,a)} = \overline{x}_e/2$.
- For each $a \in V - \{u, v\}$, add the arc (a, v) with capacity 1 and the arc (u, a) with capacity $\overline{x}(\delta(a))$.

Consider any $u - v$ cut S in H. The capacity of arcs exiting S is

$$z(\delta^{out}(S)) = |S| - 1 + \sum_{a \in V - S} \overline{x}(\delta(a))/2 + \sum_{e \in \delta(S)} \overline{x}_e/2.$$

The latter two sums count each $e \in E(V) - E(S)$ twice: if $e \in E(V - S)$ then it will be counted twice in the first sum and if $e \in \delta(S)$ it will be counted exactly once in the first sum and exactly once in the second. Therefore,

$$z(\delta^{out}(S)) = |S| - 1 + \overline{x}(E(V)) - \overline{x}(E(S)) = (|V| - 1) + (|S| - 1) - \overline{x}(E(S)).$$

Therefore, the minimum-capacity $u - v$ cut in H has capacity $< |V| - 1$ if and only if some violated constraint contains u and excludes v. Running this over all pairs $u, v \in V$ will find a violated constraint if there is any. ■

The proof used $n \cdot (n - 1)$ min-cut computations. It can be reduced to at most $2n - 2$ by only fixing one particular u, guessing the corresponding $v \neq u$, and trying to find the minimum $u - v$ and $v - u$ cuts in the corresponding graphs.

Lemma 2 *The feasible integer solutions are precisely the* $\{0,1\}$ *solutions corresponding to spanning trees of* G.

Proof. Let $\overline{x}$ be a feasible integer solution. Note that $\overline{x}_{(u,v)} \le 1$ for each $(u,v) \in E$ because $x(E(S)) \le 1$ is satisfied for $S = \{u,v\}$. Let $T = \{e : \overline{x}_e = 1\}$.

We have $|T| = \overline{x}(E(V)) = |V| - 1$. Furthermore, T cannot contain a cycle because if T contained a cycle with vertex set C, then we must have $\overline{x}(E(C)) \ge n$ which contradicts feasibility of $\overline{x}$. Any graph on n nodes that has $n-1$ edges and does not contain a cycle is a spanning tree, so T is a spanning tree.

Conversely, any spanning tree T contains exactly $n-1$ edges and for each $S \subseteq V$, at most $|S|-1$ edges of T have both endpoints in S (otherwise there is a cycle contained in S) so the $\{0,1\}$ integer point corresponding to T is a point in (**LP-Span**). ■

29.1.1 Integrality of Extreme Points

Before proving that extreme points are integral, we introduce more important notation and concepts.

For a set of edges $F \subseteq E$, let $\chi(F) \in \mathbb{R}^E$ be the $\{0,1\}$ indicator vector for F. That is, $\chi(F)_e = 1$ for $e \in F$ and $\chi(F)_e = 0$ for $e \notin F$.

Say any two sets $A, B \subseteq V$ *cross* if $A \cap B \neq \emptyset$ but neither is a subset of the other. A family $\mathcal{L}$ of subsets of V is called *laminar* no two of its subsets cross, i.e. for any $A, B \in \mathcal{L}$ we have either $A \cap B = \emptyset, A \subseteq B$ or $B \subseteq A$.

Lemma 3 *Let* $\mathcal{L}$ *be a laminar family of subsets of* V *such that* $|A| \ge 2$ *for any* $A \in \mathcal{L}$. *Then* $|\mathcal{L}| \le |V| - 1$.

Proof. Assignment 5. ■

Theorem 1 *Any extreme point of (***LP-Span***) is integral.*

Proof. Let $\overline{x}$ be an extreme point. It is easy to see that $\overline{x}$ is an extreme point if and only if the corresponding solution we get after deleting $e \in E$ with $\overline{x}_e = 0$ is an extreme point, so we assume $\overline{x}_e > 0$ for each $e \in E$.

We show that in this case it must be that $|E| \le |V| - 1$. If so, then we are done because:

- $\overline{x}(E(V)) = |V| - 1$
- $\overline{x}_e \le 1$ for each $e \in E$

So if $|E| \le |V| - 1$ then we must have $\overline{x}_e = 1$ for each $e \in E$.

By the properties of extreme points, $|E|$ is equal to the rank of the collection of vectors $\mathcal{M} = \{\chi(E(S)) : \overline{x}(E(S)) = |S|-1\}$. We show that there is a laminar family $\mathcal{L}$ consisting only of S with $|S| \ge 2$ and $\overline{x}(S) = |S|-1$ such that the vectors $\chi(E(S)), S \in \mathcal{L}$ form a basis for the space spanned by all tight constraints (i.e. the space spanned by $\mathcal{M}$). If so, then by Lemma 3 we have

$$|E| = \text{rank}(M) = \text{rank}\left(\{\chi(E(S)) : S \in \mathcal{L}\}\right) = |\mathcal{L}| \le |V| - 1$$

which completes the proof.

Let $\mathcal{L}$ be the largest laminar collection of subsets of V such that $\chi(E(S)), S \in \mathcal{L}$ are linearly independent. If $|\mathcal{L}| < |E|$ then there is some $R \subseteq V, |R| \ge 2$ such that $\overline{x}(R) = |R| - 1$ but $\chi(R) \notin \text{span}\{\chi(E(S)) : S \in \mathcal{L}\}$.

Because R cannot be added to $\mathcal{L}$, we know that R crosses $S \in \mathcal{L}$. Choose such an R that crosses the fewest sets in $\mathcal{L}$ and let S be any set in $\mathcal{L}$ such that R and S cross.

Let F' denote the edges with one endpoint in $S - R$ and the other in $R - S$. Then we have

$$\begin{array}{rcll} |R| - 1 + |S| - 1 & = & \overline{x}(E(R)) + \overline{x}(E(S)) & \text{(the corresponding constraints are tight)} \\ & = & \overline{x}(E(R \cap S)) + \overline{x}(E(R \cup S)) - \overline{x}(F') & \text{(count how many times each edge contributes to each side)} \\ & \leq & |R \cap S| - 1 + |R \cup S| - 1 & (\overline{x} \text{ is feasible}) \\ & = & |R| - 1 + |S| - 1 & \end{array}$$

Therefore all inequalities hold with equality. In particular:

- $\overline{x}(E(S \cup R)) = |S \cup R| - 1$
- $\overline{x}(E(S \cap R)) = |S \cap R| - 1$
- $\overline{x}(F') = 0$

(if $|S \cap R| = 1$ then just ignore any term involving it and the proof works fine)

Because $\overline{x}_e > 0$ for each $e \in E$, then $F' = \emptyset$ which means $\chi(E(R)) + \chi(E(S)) = \chi(E(R \cup S)) + \chi(E(R \cap S))$. Both $R \cup S$ and $R \cap S$ can only cross sets in $\mathcal{L}$ that R crossed. Since both do not cross S, then both cross fewer sets in $\mathcal{L}$ than R.

Finally, it cannot be that both $\chi(E(R \cup S)), \chi(E(R \cap S)) \in \text{span}\{\chi(E(S')) : S' \in \mathcal{L}\}$, otherwise $\chi(E(R)) = \chi(E(R \cup S)) + \chi(E(R \cap S)) - \chi(E(S)) \in \text{span}\{\chi(E(S')) : S' \in \mathcal{L}\}$. Therefore, at least one of $R' \in \{R \cap S, R \cup S\}$ is such that $\chi(E(R')) \notin \text{span}\{\chi(E(S')) : S' \in \mathcal{L}\}$, R' crosses fewer sets of $\mathcal{L}$ than R, and the constraint for R' is tight. This contradicts our choice of R. ■

29.2 The Minimum Bounded-Degree Spanning Tree Problem

Now we tackle the main problem. Given a graph $G = (V, E)$ with edge costs $c_e \geq 0, e \in E$ and integer vertex bound $B_v \geq 1, v \in V$, the goal is to find the cheapest spanning tree T of G such that $|\delta(v) \cap T| \leq B_v$ for each $v \in V$. It is NP-hard to determine if there is a feasible solution even when $B_v = 2$ for all $v \in V$ because this is precisely the problem of determining if G has a Hamiltonian path.

We will see the next best thing: a polynomial-time algorithm that either (correctly) states there is no such tree or it returns a spanning tree T with $|\delta(v) \cap T| \leq B_v + 1$. Furthermore, if there is in fact a spanning tree satisfying the original degree bounds then the cost of the returned tree T is at most OPT. We are not losing anything in the objective function value here, just the degree bounds!

We consider the following linear programming relaxation. Here, $x(\delta(v))$ denotes $\sum_{e \in \delta(v)} x_e$. The relaxation is slightly more general in that we only have variables for a subset of edges $F \subseteq E$ and degree constraints for a subset of vertices $W \subseteq V$.

$$\begin{array}{rrcll} \text{minimize :} & \displaystyle\sum_{e \in F} c_e \cdot x_e & & & \\ \text{subject to :} & x(F(S)) & \leq & |S| - 1 & \text{for each } S \subseteq V, |S| \geq 2 \\ & x(F(V)) & = & |V| - 1 & \\ & x(\delta(v)) & \leq & B_v & \text{for each } v \in W \\ & x & \geq & 0 & \end{array} \qquad (\textbf{LP-BDST}(W, F))$$

The algorithm we consider is an *iterative relaxation* algorithm. It iterates the process of solving the LP, deleting edges with x-value 0, and dropping some constraints until the set of feasible solutions is given by the normal spanning tree LP (**LP-Span**).

Algorithm 1 Minimum Bounded-Degree Spanning Tree Approximation

if (**LP-BDST**(V, E)) is infeasible **then**
  return *no solution*
end if
$F \leftarrow E$
$W \leftarrow V$
while $W \neq \emptyset$ **do**
  Solve (**LP-BDST**(W, F)) to get an optimum extreme point $\overline{x}$
  $F \leftarrow \{e \in F : \overline{x}_e > 0\}$
  $W \leftarrow \{v \in V : |\delta(v) \cap F| \geq B_v + 2\}$
end while
return An optimum extreme point solution to (**LP-BDST**$(\emptyset, F)$)

If Algorithm 1 returns *no solution* then clearly there is none as the $\{0, 1\}$ solution corresponding to the optimal degree-bounded spanning tree would be feasible. Next, since the main loop only drops constraints and edges with $\overline{x}$-value 0 then the cost $\sum_{e \in F} c_e \cdot \overline{x}_e$ does not increase over the iterations. Since we only drop degree constraints for vertices v with $|\delta(v) \cap F| \leq B_v + 1$, then any resulting integer solution must satisfy this slightly relaxed degree bound. Finally, the feasible solutions of (**LP-BDST**$(\emptyset, F)$) are precisely the feasible solutions of (**LP-Span**) for the graph $G = (V, F)$.

By Theorem 1, the fact that the optimum solution LP solution does not increase over the iterations, and the fact that $|\delta(v) \cap F| \leq B_v + 1$, the last step returns an integer solution corresponding to a spanning tree with cost at most the optimum of (**LP-BDST**(V, E)) that violates each degree bound by at most +1.

All that is left to prove is that each iteration of the algorithm makes progress.

Theorem 2 *Consider an extreme point $\overline{x}$ for (**LP-BDST**(W, F)) such that $\overline{x}_e > 0$ for each $e \in F$. If $W \neq \emptyset$, then there is some $v \in W$ such that $|\delta(v) \cap F| \leq B_v + 1$.*

Proof. By way of contradiction, suppose $|\delta(v) \cap F| \geq B_v + 2$ for every $v \in W$. Using essentially the same arguments as in the proof of Theorem 1, we find a laminar collection $\mathcal{L}$ of subsets of V such $|S| \geq 2$ for each $S \in \mathcal{L}$ and such that the corresponding vectors form a basis for $\{\chi(E(S)) : \overline{x}(E(S)) = |S| - 1\}$. Then we find $U \subseteq W$ whose corresponding degree constraints are tight such that the vectors

$$\{\chi(F(S)) : S \in \mathcal{L}\} \cup \{\chi(\delta(v)) : v \in U\} \tag{29.1}$$

form a basis for the space spanned by tight constraints. This can be done by greedily adding vertices $u \in W$ such that $\overline{x}(\delta(u)) = B_v$ to U while ensuring the vectors (29.1) remain linearly independent.

Note that have $|\mathcal{L}| + |U| = |F|$ by the characterization of extreme points. Now, if $U = \emptyset$ then $\overline{x}$ is an extreme point of (**LP-Span**) for the graph $G = (V, F)$, so it is integral already by Theorem 1 and it is clear that integer solutions to (**LP-BDST**(W, F)) must satisfy the degree bounds for nodes in W without any violation. So, we now assume $U \neq \emptyset$.

We will assign a charge of 1 to each $e \in F$ and distribute some of this charge to sets in $\mathcal{L}$ and vertices in W. We will count the amount of charge that is redistributed in two ways. On one hand, we see that strictly less than $|F|$ units of charge is sent to these sets. On the other hand, we will see that at least $|F|$ units of charge were collected by $\mathcal{L}$ and W. This is a contradiction, so it must be that some $v \in W$ satisfies $|\delta(v) \cap F| \leq B_v + 1$.

For each $e = (u, v) \in F$, send $\overline{x}_e$ units of charge to the smallest $S \in \mathcal{L}$ with $u, v \in S$ (if there is none, then do not distribute this charge). Also, send $(1 - \overline{x}_e)/2$ units of charge to each of u and v that lies in U. Note that e sends out at most 1 unit of charge, so the total charge sent out by all edges is at most $|F|$ (we will soon see it is, in fact, strictly less than $|F|$).

Next we show that each $v \in U$ and each $S \in \mathcal{L}$ collect at least one unit of charge. To start, consider some $v \in U$. Then the charge that v collects is precisely

$$\sum_{e \in \delta(v) \cap F} \frac{1 - \overline{x}_e}{2} = \frac{|\delta(v) \cap F| - B_v}{2} \geq 1$$

where the equality is because the degree constraint for $u \in W$ is tight and the inequality is because we are assuming $|\delta(v) \cap F| \geq B_v + 2$.

Now consider some $S \in \mathcal{L}$. Let $R_1, R_2, \ldots, R_k$ denote the maximal subsets of S in $\mathcal{L}$. That is, each $R_i \in \mathcal{L}$ is a proper subset of S and no other $R' \in \mathcal{L}$ satisfies $R_i \subsetneq R' \subsetneq S$. Then the total charge collected by $\mathcal{L}$ is precisely

$$\overline{x}(F(S)) - \sum_{i=1}^{k} \overline{x}(F(R_i)) = (|S| - 1) - \sum_{i=1}^{k} (|R_i| - 1).$$

We have $|R_1| + \ldots + |R_k| \leq |S|$ so the last expression is a nonnegative integer. Furthermore, we have $\chi(F(S)) \neq \sum_i \chi(F(R_i))$ (by linear independence) so there is some edge $e \in F$ in $F(S)$ but not in any $F(R_i)$. Thus, S collects a positive integer amount of charge, meaning it collects at least 1 charge.

So far, we have shown that the edges distribute at most $|F|$ units of charge and that $\mathcal{L}$ and U collectively receive at least $|\mathcal{L}| + |U| = |F|$ units of charge. We will show that some edge did not distribute exactly 1 unit of charge, so in fact the total charge that was distributed is strictly less than $|F|$, a contradiction.

First, two simple cases:

- If $V \notin \mathcal{L}$ then there is some $e \in F$ that is not contained in any $S \in \mathcal{L}$ so the charge $\overline{x}_e > 0$ is not distributed.
- If there is some vertex $v \in V - U$ such that $\overline{x}_e < 1$ for some $e \in \delta(v) \cap F$, then the charge $(1 - \overline{x}_e)/2 > 0$ is not distributed.

Now assume that none of these happen. We also note that if $\overline{x}_e = 1$ for some $e = (u, v) \in F$ then $\chi(E(\{u, v\})) \in \text{span}\{\chi(E(S)) : S \in \mathcal{L}\}$ because the constraint $\overline{x}(E(\{u, v\})) \leq 1$ is tight and we chose $\mathcal{L}$ so that the associated vectors span $\{\chi(E(S)) : \overline{x}(E(S)) = |E| - 1\}$.

Putting all of this together, we have

$$2 \cdot \chi(E(V)) = \sum_{v \in U} \chi(\delta(v)) + \sum_{v \in V - U} \chi(\delta(v)) = \sum_{v \in U} \chi(\delta(v)) + \sum_{v \in V - U} \sum_{e \in \delta(v)} \chi(\{e\}).$$

We just argued that each vector $\chi(\{e\})$ in the last sum is spanned by $\{\chi(E(S)) : S \in \mathcal{L}\}$. Furthermore, the first sum in the last expression is nonzero because $U \neq \emptyset$. Therefore, we have expressed a non-zero linear combination of the vectors $\{\chi(\delta(v)) : v \in U\}$ by a linear combination of the vectors in $\{\chi(E(S)) : S \in \mathcal{L}\}$, which contradicts the fact that the vectors in (29.1) are linearly independent. ■

The spanning tree polytope (**LP-Span**) was presented and proved to be integral by Edmonds [E71]. Singh and Lau described the +1 approximation for the Minimum Degree Bounded Spanning Tree problem [SL11], which improved over the +2 approximation by Goemans [G06]. Earlier work had considered the unweighted problem: given degree bounds B_v determine if there is any spanning tree with these bounds. An algorithm by Fürer and Raghavachari [FR94] will either find some spanning tree where the degree of each node v is at most $B_v + 1$ or else determine there is no such tree.

References

E71 J. Edmonds, Matroids and the greedy algorithm, Mathematical Programming, 1:125–136, 1971.

FR94 M. Fürer and B. Raghavachari, Approximating the minimum-degree steiner tree to within one of optimal, Journal of Algorithms, 17(3):409–423, 1994.

SL11 M. Singh and L. C. Lau, Approximating minimum bounded degree spanning trees to within one of optimal, In Proceedings of STOC, 2007.