COMPSCI 632: Approximation Algorithms	October 18, 2017
Lecture 14	
Lecturer: Debmalya Panigrahi	*Scribe: Abe Frandsen*

1 Overview

We continue our discussion of degree-bounded spanning tree problem and given an algorithm which produces a solution that acheives optimal cost and violates the degree bounds by at most 1. We next address approximation algorithms for multiway cut problems.

2 Degree-bounded Spanning Tree

Recall the degree-bounded spanning tree relaxed linear program for a graph $G = (V,E)$, edge weights $w_e, \forall e \in E$ and degree bounds $B_v, \forall v \in W \subset V$ is given by

$$\begin{aligned} \min \quad & \sum_{e \in E} w_e x_e & \\ \text{subject to} \quad & x(S) \le |S| - 1, S \subset V & \text{(subtour constraints)} \\ & x(V) = |V| - 1 & \\ & x(\delta(v)) \le B_v, v \in W & \text{(degree bound constraints)} \\ & x_e \ge 0, e \in E & \end{aligned}$$

where $x(S) := \sum_{(u,v) \in E : u,v \in S} x_{(u,v)}$ and $x(\delta(v)) := \sum_{(u,v) \in E} x_{(u,v)}$. This problem generalizes the Hamiltonian path problem (set $B_v = 2$ for all $v \in V$), which is NP-hard, so there is little hope to find an efficient exact algorithm for our problem. However, last time, we saw an algorithm that produces a spanning tree of optimal cost that violates the degree bounds by at most 2 (Goemans [**?**]). We now improve this to an an algorithm that violates the degree bounds by at most 1:

While W is nonempty, do the following:

- solve the LP to obtain a (fractional) solution x.
- for each $e \in E$, if $x_e = 0$, remove e from E.
- if there is a vertex $v \in W$ with degree relative to E bounded by $B_v + 1$, remove v from W.

When the iteration terminates, return the minimum spanning tree of the final subgraph (V,E).

Note that as the iteration progresses, we are either removing edges that were assigned a weight of zero (which doesn't change the cost of the solution), or we are removing constraints (which can only improve the cost of the solution). At the final stage, there are no degree constraints, so the problem reduces to finding a minimum spanning tree, and this MST will have cost no more than the cost of the initial fractional optimal solution. Moreover, we only remove a degree constraint if the degree of that vertex in the current graph is no more than its degree bound plus 1. Hence, in the final solution, the degree of each vertex will exceed its degree bound by no more than 1. Therefore, this algorithm is correct up to $+1$ degree constraint violations as long as the iteration terminates. That is, we need to show that if all $x_e > 0$, then there exists some $v \in W$ such that $d(v) \leq B_v + 1$. If we can show this, then clearly the iteration terminates within $|E| + |W|$ steps.

To this end, let x be a fractional solution to the LP such that $x_e > 0$ for all $e \in E$, and suppose to the contrary that $d(v) > B_v + 1$ for all v. Let $\mathscr{L}$ be a maximal laminar family of tight subtour constraints, and let T be the set of tight degree bound constraints (we also use T to denote the set of vertices corresponding to these tight constraints, but this abuse of notation should be clear from the context). We have seen earlier that T and $\mathscr{L}$ are linearly independent and that $|\mathscr{L}| + |T| = |E|$. Let $\delta(v)$ denote the set of edges incident on v, and for $U \subset E$, let $\chi(U)$ denote the edge incidence vector for U (i.e. it is a binary vector that indicates which edges are contained in U). Note that

$$
\begin{aligned}
2\chi(E) &= \sum_{v \in V} \chi(\delta(v)) && (1)\\
&= \sum_{v \in T} \chi(\delta(v)) + \sum_{v \in V \setminus T} \chi(\delta(v)) && (2)\\
&= \sum_{v \in T} \chi(\delta(v)) + \sum_{v \in V \setminus T} \sum_{e \in \delta(v)} \chi(\{e\}). && (3)
\end{aligned}
$$

Recall the token distribution process, in which we have $|E|$ tokens overall, and each edge $e \in E$ distribution a total of one token as follows: x_e tokens to the smallest set in $\mathscr{L}$ containing the endpoints of e, and $(1 - x_e)/2$ tokens to each endpoint of e. We saw previously that the number of tokens given to any set corresponding to a constraint in $\mathscr{L}$ is at least 1, and the number of tokens given to any vertex in T is also at least 1. Since $|E| = |T| + |\mathscr{L}|$, we see that all tokens are distributed to $\mathscr{L}$ and T.

Now let $v \in V \setminus T$ and take $e \in \delta(v)$. If $x_e < 1$ then $(1 - x_e)/2$ tokens are given to v. But then this token amount is *not* distributed to T or $\mathscr{L}$, which contradicts the fact that we established above. Hence, we conclude that $x_e = 1$, and so the endpoints of e correspond to a tight subtour constraint, i.e. $\chi(\{e\})$ is in the span of $\mathscr{L}$. Therefore, every term on the righthand side of (3) is in the span of $T \cup \mathscr{L}$, and $\chi(E)$ is in $\mathscr{L}$. This contradicts the linear independence of T and $\mathscr{L}$.

3 Multiway Cuts

3.1 Minimum $s - t$ Cut

Recall the minimum $s - t$ cut LP with edge weights c_e is given as

$$
\begin{aligned}
\min \quad & \sum_{e \in E} c_e l_e \\
\text{subject to} \quad & \sum_{e \in p} l_e \geq 1,\ p \in \mathscr{P}_{s,t} \\
& l_e \geq 0,\ e \in E,
\end{aligned}
$$

where $\mathscr{P}_{s,t}$ is the set of paths from s to t. Observe that this LP has possibly an exponential number of constraints, but it is separable (e.g. the separation oracle is to check that the shortest $s - t$ path according to lengths l_e is at least 1).

We can think of this problem as a system of pipes. Each edge e represents a pipe segment of cross-sectional area c_e and length l_e. In this analogy, the objection function of the LP is just the total volume of the pipe system, and the constraints mean that each $s - t$ path has length at least 1 (as measured by the pipe lengths l_e).

Now choose r uniformly at random from $(0,1)$, and define $B(s,r) = \{v \in V : d(s,v) \leq r\}$ (we take distances with respect to the lengths l_e). Further, let $\delta(B(s,r))$ be the edges in the cut formed by $B(s,r)$. We have

Lemma 1. $\mathbb{E}\left[\sum_{e\in\delta(B(s,r))} c_e\right] \leq \sum_{e\in E} c_e l_e$

Proof. Observe that

$$\mathbb{E}\left[\sum_{e\in\delta(B(s,r))} c_e\right] = \sum_{e\in E} c_e \cdot Pr(e \in \delta(B(s,r))).$$

Now consider an edge $e = (u,v)$. This edge crosses the $B(s,r)$ cut if $d(s,u) \leq r < d(s,v)$. Then

$$\begin{aligned} Pr(e \in \delta(B(s,r))) &= Pr(d(s,u) \leq r < d(s,v)) \\ &= d(s,v) - d(s,u) \\ &\leq l_e, \end{aligned}$$

where the second line follows from the fact that r is drawn uniformly at random from $(0,1)$, and the thrid line follows from the fact that the length l_e of the edge e is an upper bound for the difference in lengths between the shortest path from s to u and the shortest path from s to v. ☐

Corollary 2. *The $s-t$ mincut LP is integral.*

Proof. By the previous lemma, we know there exists some $r \in (0,1)$ such that

$$\sum_{e\in\delta(B(s,r))} c_e \leq \sum_{e\in E} c_e l_e.$$

But the edges of $\delta(B(s,r))$ form a feasible integral solution to the LP, and the objective function value of this feasible solution is given by the left-hand side of the above inequality. On the other hand, the right-hand side of the inequality is optimal LP objective function value. Hence, the integral solution is just as good as the optimal fractional solution, and hence the integrality gap is 1. ☐

3.2 General Multiway Cut

We now move on to the general multiway cut problem. Let $s_1, \ldots, s_k$ be vertices (called *terminals*). We seek to find a partition of the vertex set that separates the terminals and that minimizes the cost of cut edges. This problem can be expressed as follows:

$$\begin{aligned} \min \quad & \sum_{e\in E} c_e l_e \\ \text{subject to} \quad & \sum_{e\in p} l_e \geq 1,\, p \in \mathscr{P}_{s_i,s_j}, \quad \forall i \neq j \\ & l_e \geq 0,\, e \in E. \end{aligned}$$

This LP is a straightforward generalization of the minimum $s-t$ cut LP, but the multiway cut LP is not integral. To see why, consider a star graph, where the terminals are the leaves of the star, and each edge has cost 1. The optimal integral solution has cost $k-1$ (one terminal is grouped with the center of the star, and all other terminals are in isolated groups), but we have a fractional solution of cost $k/2$ (just set each edge variable $l_e = 1/2$). Hence, there is an integrality gap of $2(1-1/k)$.

Can we devise an algorithm that acheives this approximation factor? Let's slightly alter the random ball approach we took when analyzing the $s-t$ mincut LP. In particular, draw radius r_i uniformly at random from $(0, \frac{1}{2})$ for $i = 1, \ldots, k$, and form the balls $B(s_i, r_i)$ (where the distances are taken with respect to the fraction solutions l_e, as before). Note that these balls are all disjoint, since each pair of terminals is distance at least 1 apart from each other (by the LP constraints). Let's analyze the cost of the cut of these balls. Note that an edge $e = (u, v)$ can possibly cross two balls, if for example there exist s_i and s_j such that $d(s_i, u) \leq r_i$ and $d(s_j, v) \leq r_j$. In such a case, we charge half of the cost of e to each of $B(s_i, r_i)$ and $B(s_j, r_j)$. Now the probability of edge e crossing some ball $B(s_i, r_i)$ can be bounded as follows:

$$\begin{aligned} Pr(e \in \delta(B(s_i, r_i))) &= Pr(d(s_i, u) \leq r_i < d(s_i, v)) \\ &\leq 2(d(s_i, v) - d(s_i, u)) \\ &\leq 2l_e. \end{aligned}$$

Hence, in our charging scheme, the expected cost of the integral solution formed by the balls $B(s_i, r_i)$ is bounded above by twice the fractional cost, which yields a 2-approximation factor.

To improve this factor, consider the $1/2$-radius ball around the terminal that has the largest "volume" (according to the pipe system metaphor introduced earlier). Since all of the $1/2$-radius balls centered at each terminal are disjoint and their volumes sum up to (at most) the fractional LP cost, we have that this largest volume is greater than or equal to $\frac{1}{k} \sum_{e \in E} c_e l_e$. Therefore, by removing this ball, we can still form a partition that separates the terminals using random-radius balls around the remaining $k-1$ terminals, and this now gives a 2-approximation with respect to the smaller "pipe system", which has cost (volume) at most $(1-1/k) \sum_{e \in E} c_e l_e$. Overall, this yields a $2(1-1/k)$ approximation factor, as desired.

We finish by noting that better approximations can be acheived for this problem, although they will necessarily require analysis that goes beyond the LP we considered here.

4 Summary

In these notes, we discussed an algorithm for solving the degree-bounded spanning tree problem that produced a solution of optimal cost that violates the degree bound constraints by at most 1. We next considered the minimum $s-t$ cut problem, and showed an LP for this problem that has integrality gap of 1. We then considered the more general multiway cut problem, and saw that its LP has an integrality gap of $2(1-1/k)$. We finished by giving a rounding algorithm that acheived this approximation factor in expectation.