15-854: Approximations Algorithms **Lecturer:** Anupam Gupta
Topic: Metric Rounding: Cut Problems **Date:** November 14, 2005
Scribe: Michael Dinitz

19.1 $s-t$ mincut

In the $s-t$ mincut problem we are given a directed graph $G=(V,A)$, capacities c_e for arcs e, and two designated nodes, a source s and a sink t. The goal is to find a minimum capacity set of arcs A' such that $G-A'$ has no $s-t$ path. Equivalently, we want to find a cut $(S,\overline{S})$ with $s \in S$ and $t \in \overline{S}$ such that $c(\delta^+(S))$ is minimized, where $\delta^+(S)$ denotes the set of outgoing edges from S.

A formulation of this problem as an integer linear program is

$$\begin{array}{llll} \text{min} & \sum_{(u,v)\in A} c_{uv} l_{uv} & & \\ \text{subject to} & & & \\ & \sum_{(u,v)\in P} l_{uv} & \geq 1 & \forall s-t \text{ paths P} \\ & l_{uv} & \in \{0,1\} & \forall (u,v) \in A \end{array}$$

We can relax this formulation to get a linear program by changing the last constraint to $l_{uv} \geq 0$. Note that this is not a compact formulation, since there could be an exponential number of constraints, but we can still solve it efficiently by using the ellipsoid method with a shortest path computation under the distance function l_{uv} as the separation oracle. This works as a separation oracle since if any constraint is violated there will be a $s-t$ path of length less than 1 corresponding to the violated constraint. As we saw in the homework there is a compact formulation of the problem, but we do not use it here.

Now suppose that l^* is an optimal feasible solution to the LP relaxation, and let $Z^*_{LP} = \sum_{(u,v)\in A} c_{uv} l^*_{uv}$. First we will show how to round a feasible LP solution to an integer solution of no greater cost, which implies that we can round an optimal solution to an optimal integer solution, and then we show that any basic feasible solution to the LP is actually integral anyway.

Theorem 19.1.1 *If l^* is a feasible to the LP relaxation, then we can find a cut of cost no more than Z^*_{LP}.*

Proof: Let α_v be the distance from s to v according to the length function l^*. Also, let $B_s(r) = \{x : \alpha_x \leq r\}$. We create an integral cut by choosing a random radius $R \in_R [0,1]$ uniformly at random and setting $S = B_s(R)$.

Claim 19.1.2 *If we pick $R \in_R [0,1]$ uniformly at random, then $\mathbf{Pr}[\text{arc } e \text{ is cut}] \leq \frac{l^*_e}{1}$*

Proof of Claim: $\mathbf{Pr}[\text{arc } e \text{ is cut}] = \frac{\alpha_v - \alpha_u}{1-0} \leq \frac{l^*_e}{1}$ ■

Claim 19.1.3 $\mathbf{E}[c(\delta^+(B_s(R)))] \leq Z^*_{LP}$

Proof of Claim: We use linearity of expectations.

$$\mathbf{E}[\text{capacity of the cut}] = \mathbf{E}\big[c(\delta^+(B_s(R)))\big] = \sum_{e \in A} c_e \mathbf{Pr}[e \text{ is cut}] \le \sum_{e \in A} c_e l_e^* = Z_{LP}^*$$

■

Since the expected cost of the cut is at most Z_{LP}^*, there must be at least one cut S in the support that costs at most Z_{LP}^*. Thus we can simply try all the cuts in the support (which there are polynomially many of since there are only a polynomial number of cuts that form distinct balls around s) to find the cut with smallest cost, which will be at most Z_{LP}^* ■

This implies that

$$Z_{IP}^* \le c(\delta^+(S)) = Z_{LP}^* \le Z_{IP}^*$$

Note that since the cut we eventually return is a feasible solution to the LP relaxation we actually know that its cost is precisely Z_{LP}^*. Also, it is obvious that $\mathbf{E}[\text{capacity of the cut}]$ also equals $\sum_{\text{cuts } (S,\overline{S})} c(\delta^+(S)) \times \mathbf{Pr}\big[\text{the algorithm outputs } (S, \overline{S})\big]$. Since all of these probabilities are nonnegative, it is a convex combination of feasible cuts. However, each valid cut $(S, \overline{S})$ has $c(\delta^+(S)) \ge Z_{LP}^*$, since l^* is the optimal solution to the LP relaxation. Since each valid cut costs at least Z_{LP}^* and a convex combination of them costs at most Z_{LP}^*, every cut $(S, \overline{S})$ that is output with nonzero probability has $c(\delta^+(S)) = Z_{LP}^*$.

Now we prove that in fact any basic feasible solution to the LP is integral.

Theorem 19.1.4 *If l^* is a basic feasible solution of the LP relaxation, then l^* is integral*

Proof: First we show that $l_e^* = \alpha_v - \alpha_u$ for edges $e = (u, v)$ on a $s - t$ path. We only have to worry about these edges since any edge not on an $s - t$ path will have $l_e^* = 0$ since l^* is a basic feasible solution. Let P be an $s - t$ path containing e. Without loss of generality we assume that u is closer to s than v is. By the triangle inequality $\alpha_v \le \alpha_u + l_e^*$, so $l_e^* \ge \alpha_v - \alpha_u$. Assume that $l_e^* > \alpha_v - \alpha_u$. Then there is some $s - v$ path that does not use e and is shorter than $\alpha_u + l_e^*$. Let p be this path, and let a be the difference between $\alpha_u + l_e^*$ and the length of p. Then since there is an $s - t$ path using e, there is an $s - t$ path P' consisting of p plus the $v - t$ portion of P. Since l^* is feasible the total length of P' is at least 1. But clearly $l(P') = l(P) - a$. Then by setting $l'_e = l_e^* - a$ and $l''_e = l_e^* + a$ for the arc e and $l'_a = l''_a = l_a^*$ for all other arcs, we get two feasible solutions such that $l^* = (l' + l'')/2$, which implies that l^* is not a basic feasible solution. Since this is a contradiction, we can assume that $l_e^* = \alpha_v - \alpha_u$.

Now Claim 19.1.2 is actually that the probability an arc e is cut equals l_e^*. So following our previous analysis, we know that the convex combination of cuts in which each cut is weighted by its probability of being chosen by the algorithm actually equals l^*. Since l^* is a basic feasible solution this is a contradiction unless there is only one cut that the algorithm outputs with nonzero probability. This can only be the case if l^* is integral, thus proving the theorem. ■

19.2 Multiway Cut

The multiway cut problem, which we have seen before, is a generalization of the $s-t$ mincut problem. In this problem we are given an undirected graph $G=(V,E)$, costs (or capacities) c_e on the edges, and a specified set of terminals $T=\{s_1,\ldots,s_k\}\subseteq V$. We want to find a subset $E'\subseteq E$ of edges of minimum cost such that $G-E'$ has no s_i-s_j path for all $i,j\le k,\ i\ne j$. A linear programming relaxation of this problem is:

$$\begin{array}{llll}
\min & \sum_{e\in E} c_e l_e & & \\
\text{subject to} & & & \\
 & l(P)=\sum_{e\in P} l_e & \geq & 1 \quad \forall i\neq j,\ \forall s_i-s_j \text{ paths } P \\
 & l_e & \geq & 0 \quad \forall e\in E
\end{array}$$

Note that if we restrict $l_e\in\{0,1\}$ then this is an exact ILP formulation of the problem. We will prove that by solving the LP relaxation and using randomized rounding we can get a 2-approximation.

Theorem 19.2.1 *If l is any feasible solution to the LP and $Z=\sum_{e\in E}c_e l_e$ then there is an integral solution (i.e. a multiway cut E') with $c(E')\le 2Z$*

Corollary 19.2.2 *If l^* is an optimal LP solution with cost Z^* then $OPT\le c(E'(l^*))\le 2Z^*\le 2OPT$*

Proof of Theorem: We give an algorithm that rounds the LP solution to an integral one. The algorithm is as follows: pick a random $R\in_R[0,\frac{1}{2}]$ uniformly at random. Let $E_i=\delta(B_{s_i}(R))$ and output $\cup_i E_i$.

In order to prove the theorem we will prove the following, which clearly implies the theorem:

$$\mathbf{E}[c(\cup_i E_i)]\le\mathbf{E}\left[\sum_i c(E_i)\right]\le 2Z$$

Let the volume $V(s_i,r)=\sum_{u,v\in B_{s_i}(R)}c_{uv}l_{uv}$. In order to make the exposition simpler we assume that there are "fake" vertices at the intersection of each edge and a border of a ball, so that every edge with one end in the ball has the other end also in the ball. Clearly $\sum_i V(s_i,\frac{1}{2})\le Z$ since the balls of radius $\frac{1}{2}$ around the terminals do not intersect and the total volume of the graph is Z. Also, note that for any edge $e\in B_{s_i}(\frac{1}{2})$ we have that

$$\mathbf{Pr}[e\text{ is cut}]\le\frac{|d(s_i,u)-d(s_i,v)|}{1/2}\le 2l_{uv}$$

where d is the shortest path distance function under edge lengths l. Then we get that

$$\mathbf{E}[c(E_i)]\le\sum_{e\in B_{s_i}(1/2)}\mathbf{Pr}[e\text{ is cut}]\times c_e\le\sum_{e\in B_{s_i}(1/2)}2l_e c_e=2\times V(s_i,\frac{1}{2})$$

Now by linearity of expectations we get that

$$\mathbf{E}\left[c(\bigcup_i E_i)\right] \le \sum_i \mathbf{E}[c(e_i)] \le 2\sum_i V(s_i, \frac{1}{2}) \le 2Z$$

which completes the proof of the theorem. ■

We can easily derandomize this rounding method by simply ordering the vertices in the balls around the terminals by their distances from the center of the ball and then checking the cuts corresponding to balls around the terminals of those radii. Also, by simply not including the ball with the greatest volume we can improve the approximation ratio to $2(1 - \frac{1}{k})$.

19.2.1 Integrality Gap

Consider the star graph in which k terminals are all adjacent to a central nonterminal and every edge hast cost 1. An optimal multiway cut contains all but one edge, for a total cost of $k - 1$. In the LP relaxation, though, we can set $l_e = \frac{1}{2}$ for all edges, which gives a cost of $\frac{k}{2}$. Thus the integrality gap of the LP is $2(1 - \frac{1}{k})$, so our rounding algorithm is tight. There is a way to add extra valid inequalities to the LP to reduce the integrality gap and get a $\frac{3}{2}$-approximation, but we do not discuss it here.