Lecture 8: Sept 24, 2003

Lecturer: Yuval Rabani *Scribe: Zoya Svitkina*

1 Integrality Gap for Multicut LP

Last time we presented an $O(\log k)$-approximation algorithm for the multicut problem, where k is the number of terminal pairs to be separated. How good is this bound? In general, this is not known. However, the LP that we used has integrality ratio of $\Omega(\log k)$.

Theorem 1. For all sufficiently large n and $k < \binom{n}{2}$, there exists an instance of the multicut problem with integrality gap $\Omega(\log k)$.

Proof: We will prove the theorem for a special case. The example will be a bounded-degree expander graph on n nodes. Recall that an (α, n)-expander is a graph G such that for any node set U with $|U| \leq \frac{n}{2}$, the number of edges leaving U is at least $\alpha \cdot |U|$. Usually what is meant is a family of graphs with constant expansion and bounded degree. They are often useful for proving bounds on integrality gaps.

It's relatively easy to see that bounded-degree expanders have diameter $\Theta(\log n)$ (e.g., imagine performing a breadth-first search). Also, there are $\Theta(n^2)$ pairs of nodes separated by distance $\Omega(\log n)$ each. This follows from the fact that the graph is of bounded degree, so for each node there are $\Theta(n)$ other nodes at distance $\Omega(\log n)$. All these pairs of nodes will be commodities in our example.

No feasible solution to the problem can contain components of size more than $\frac{n}{2}$, because then this component would have to include nodes which are far from each other, and which therefore form sourse-sink pairs. But if a component has size $x \leq \frac{n}{2}$, by definition of expander the number of edges leaving this component is $\Theta(x)$. So the total number of cut edges in a solution is $\Theta(n)$.

As a fractional solution, assign a weight of $O(\frac{1}{\log n})$ to each edge. Since the number of edges between s_i and t_i is $\Omega(\log n)$, the total weight on them will be at least 1, making the solution feasible. The cost is then $O(\frac{|E|}{\log n}) = O(\frac{n}{\log n})$, which makes the integrality gap of $\Omega(\log k)$.

□

Remark: To get a bound that depends only on k, one can replace edges of the expander graph with paths, thus increasing n, but leaving k and the solution costs unchanged.

2 Multiway cut

Multiway cut is a special case of the multicut problem, for which significantly better approximation guarantees are known. The input is a graph $G = (V, E)$ and a set of terminals

$T \subseteq V$, $|T| = k$. The commodities are all pairs of terminals, i.e. each terminal has to be separated by the cut from all others. In this special case of multicut problem, a graph that connects each $s - t$ pair by an edge would be a clique.

Remark: Given an instance of multicut problem, if we know which sets of terminals are in the same connected components in the optimal solution, then we can use a multiway cut algorithm on those groups of terminals, and get a constant-factor approximation. Thus, trying all combinations of terminals gives us a constant-factor approximation algorithm for multicut problem, with running time which is exponential in k.

2.1 A combinatorial algorithm [1]

This algorithm is very simple. For each terminal t, it finds a minimum cut C_t separating t and $T - \{t\}$. The output is the union of $(k-1)$ cheapest such cuts, and has cost C. Clearly, no two terminals stay in the same connected component, so the output is a feasible multiway cut. To bound the cost of this solution, consider an optimal solution, and let C_t^* be the total capacity of edges that separate t's component from others, for any $t \in T$. Since C_t was a min-cut, $C_t^* \geq C_t$. If we add all C_t^*'s, we would count each edge uv that was cut in the optimal solution twice, once from the side of u, and once from the side of v. So we have:

$$C \leq \frac{k-1}{k} \cdot \sum_t C_t \leq \frac{k-1}{k} \cdot 2 \cdot OPT$$

For example, if $k = 3$, the algorithm gives a $\frac{4}{3}$-approximation.

2.2 The previous LP

Let's look again at the LP relaxation that we used for the multicut problem. What caused the $O(\log k)$ bound in the rounding there was the fact that edges repeatedly had a chance of being cut, i.e., if an edge is not cut by one ball, it can still be cut by another ball. But since the balls that we use are of radius $< \frac{1}{2}$, and the distances between terminals are ≥ 1, in the case of multiway cut no edge can be cut in the same place by more than one ball[1]. In the case that an edge uv can be cut by a ball around terminal t, $d(u,v) \geq |d(t,u) - d(t,v)|$, and since the radius is randomly chosen in the interval $[0, \frac{1}{2})$, $Prob[uv\ is\ cut] = 2 \cdot |d(t,u) - d(t,v)| \leq 2d(u,v)$. So this LP and rounding procedure give a $2 \cdot \frac{k-1}{k}$ approximation algorithm.

The integrality gap for this LP is also $2 \cdot \frac{k-1}{k}$. For the case $k = 3$, an example is a graph of three terminals t_i and one node r, with three edges rt_i of capacity 1. The optimal integer solution has to cut two of the edges. A fractional solution assigns $d(r, t_i) = \frac{1}{2}$, with the other distances induced. This solution is feasible because $d(t_i, t_j) = 1$ for $i \neq j$, and has cost $\frac{3}{2}$. Thus the integrality gap is $\frac{4}{3}$.

[1] In the case of an edge uv such that for two terminals t and t', $d(t,u) < \frac{1}{2}$ and $d(t', v) < \frac{1}{2}$, we can insert a new node w between u and v such that each of uw and wv can now be cut by only one ball.

2.3 A new LP [2]

The approximation factor for multiway cut can be improved by using a different LP. The idea is to restrict the class of metrics so that, in particular, the previous example would not be allowed. A way to view the solution to multiway cut problem is as an assignment of nodes to terminals. The cut would then consist of all edges between nodes that have been assigned to different terminals.

Let the set of terminals be $T = \{1, 2, ..., k\}$. For each node u, we will define a vector x^u such that $x^u_i = \begin{cases} 1 & if\ u\ is\ assigned\ to\ i \\ 0 & otherwise \end{cases}$ and $x^u \in \{e^1, e^2, ..., e^k\}$, where $e^i_j = \begin{cases} 1 & i = j \\ 0 & o.w. \end{cases}$.

With the additional constraint that $x^i = e^i$ (i.e., a terminal is assigned to its own component), we want to minimize the capacity of cut edges, which is $\sum_{uv \in E} \frac{1}{2} ||x^u - x^v||_1 \cdot c_{uv}$.

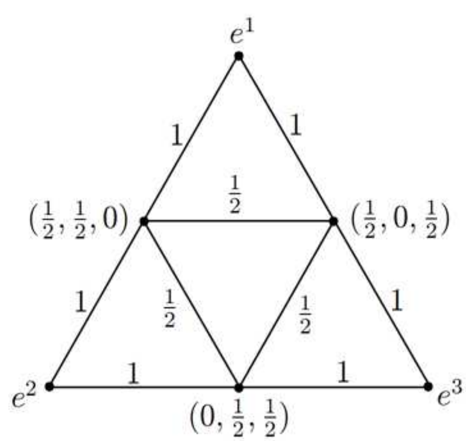
Then the LP relaxation is:

$$
\begin{array}{lll}
min & \frac{1}{2} \sum_{uv} ||x^u - x^v||_1 \cdot c_{uv} & \\
 & & \\
s.t. & x^i = e^i & \\
 & \sum_j x^u_j = 1 & \forall u \\
 & x^u \geq 0 & \forall u
\end{array}
$$

The objective function may not seem linear, but it can be expressed in a linear way. $||x^u - x^v||_1 = \sum_j |x^u_j - x^v_j|$, and the absolute value $|x^u_j - x^v_j|$ can be replaced by a variable z^{uv}_j with constraints $z^{uv}_j \geq x^u_j - x^v_j$ and $z^{uv}_j \geq x^v_j - x^u_j$.

A solution to this LP corresponds to embedding the nodes of the graph into a $(k-1)$-dimensional simplex. It can be verified that the fractional solution to previous LP for the 3-terminal example that we presented before cannot be embedded into a simplex, and therefore is not a feasible solution to this improved LP.

2.3.1 A bad example for this LP[4]



In this example, the optimal integral solution has value 4, and the fractional solution has value 3.75, giving an lower bound on the integrality gap of $\frac{16}{15}$.

The graph consists of 3 terminals and 3 non-terminal nodes, with edge capacities shown in the figure. The coordinates of the nodes shown in the figure represent the embedding produced by the LP. Since the length of edges of capacity $\frac{1}{2}$ turns out to be $\frac{1}{2}$, the objective function value is $3 \cdot \frac{1}{2} \cdot \frac{1}{2} + 3 = 3\frac{3}{4}$. For the optimal solution, one option is to assign all 3 non-terminals to the same terminal, which gives a cost of $4 \cdot 1 = 4$, and the other option is to assign two of them to one terminal, and one to another terminal, with the cost of $3 + 2 \cdot \frac{1}{2}$, which is also 4.

For the special case of $k = 3$, both an algorithm and a tight example are known showing that the integrality gap of this LP is exactly $\frac{12}{11}$ [3, 5].

References

[1] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The Complexity of Multiterminal Cuts. SIAM Journal on Computing, Volume 23, Number 4, pp. 864-894.

[2] G. Calinescu, H. Karloff, and Y. Rabani. An improved approximation algorithm for multiway cut. STOC 1998.

[3] W. H. Cunningham and L. Tang. Optimal 3-terminal cuts and linear programming. IPCO '99.

[4] A. Freund and H. Karloff. A lower bound of 8/(7+1/(k-1)) on the integrality ratio of the Calinescu-Karloff-Rabani relaxation for multiway cut. Information Processing Letters 75(1-2):43-50, 2000

[5] D. R. Karger, P. Klein, C. Stein, M. Thorup, and N. E. Young. Rounding algorithms for a geometric embedding of minimum multiway cut. STOC 1999