College of Computer & Information Science
Northeastern University
CS G399: Algorithmic Power Tools I

Fall 2007
19 October 2007
Scribe: Zhifeng Sun

Lecture Outline:

- Leighton-Rao sparsest cut theorem

1 Leighton-Rao's sparsest cut theorem

In the previous lecture, we already saw that Max-Flow f^* could be much smaller than Min-Cut η for the Uniform Multicommodity Flow problem, $f^* \le O\left(\frac{\eta}{\log n}\right)$. In this lecture, we are going to show this inequality is tight up to constant factors in Uniform Multicommodity Flow problem. This was proved by Leighton-Rao in 1990. The result was proved for multicommodity flow with general demands later by Linial-London-Rabinovich and Aumann-Rabani.

First, let's recall the LP for this problem. Let P_i be the set of paths between pair (s_i, t_i), p_j^i be the j^{th} path in P_i, $dem(i)$ be the demands of pair (s_i, t_i), and f_j^i be the amount of commodity i sent on path p_j^i. Then we get the following LP for Demands Multicommodity Flow problem.

$$\begin{aligned} &\max f \\ \text{s.t.} &\textstyle\sum_{p_j^i} f_j^i \ge f \cdot dem(i) \quad \forall i \\ &\textstyle\sum_{p_j^i : e \in p_j^i} f_j^i \le c_e \qquad \forall e \end{aligned}$$

And its dual is

$$\begin{aligned} &\min \textstyle\sum_e c_e d_e \\ \text{s.t.} &\textstyle\sum_{e \in p_j^i} d_e \ge l_i \quad \forall i, j \\ &\textstyle\sum_i l_i \cdot dem(i) \ge 1 \end{aligned}$$

In the remainder, we assume uniform demands. That is, we assume we have a commodity for each pair (u, v) with $demand = 1$. For simplicity, we also assume in this lecture that $c_e = 1$. The arguments can be easily generalized to arbitrary c_e's. (We could also replace the edge with capacity c_e with c_e parallel edges of capacity 1. While this will suffice for proving the max-flow-min-cut relationship in the above theorem, it would not establish the polynomial-time computability.)

Theorem 1. *For any n, there is an n-node uniform multicommodity flow problem with Max-Flow f^* and Min-Cut η for which $\eta = O(f^* \log n)$. And a cut of value $O(f^* \log n)$ can be found in polynomial time. In other words, this is an $O(\log n)$-approximation for sparsest cut.*

To provide some intuition, let us consider the following proof of the max-flow min-cut theorem for the single commodity flow problem.

Single Commodity Flow: After solving the dual LP, we have d_e values. Treat these values as the distance assignments of each edge. From the dual LP constraint, we know that the distance between the source and the sink is 1. Now taking the source as center, begin to grow a ball continously with radius t, t ranging from 0 to 1. The set of nodes inside the ball are all nodes that are within distance t from the source, according to the d_e labels. From the complimentary slackness, we know if $d_e > 0$ then the corresponding edge is saturated, so the min-cut is some collection of this saturated edges.

Let $C(t)$ denote the set of edges that cross the ball of radius t. At time t $(t : 0 \to 1)$, let $x(t)$ be the value of cut $C(t)$. Then during the period $(t, t + dt)$, the cut $C(t)$ contributes $x(t) \cdot dt$ to $\sum_e c_e d_e$. So $\int_0^1 x(t)dt = \sum_e c_e d_e$. From an averaging argument, we know $\exists t$ at which $C(t) \leq \sum_e c_e d_e = f^*$.

Multicommodity Flow: We generalize the idea for single commodity flow to uniform multicommodity flow. After solving the dual LP, we know $f^* = \sum_e d_e$. The average d_e value $\bar{d} = f^*/m$ (here m is the number of edges whose $d_e > 0$). Now replace every edge e in the original graph G by a path of $\lceil d_e/\bar{d} \rceil$ edges, each with distance $\bar{d}$. We could have a new graph, G^+. And the total distance values of edges in this graph won't exceed $2f^*$. Furthermore, the total number of edges in G^+ does not exceed $\sum_{e:d_e>0} \lceil d_e/\bar{d} \rceil$, which is at most $2m$.

$$\sum_{all\ original\ edges} \lceil \frac{d_e}{\bar{d}} \rceil \leq \sum \frac{d_e}{\bar{d}} \bar{d} + \sum \bar{d} \leq f^* + f^* = 2f^*$$

Our goal is to find a cut of value $O(f^* \log n)$. We define set S as balanced if $\frac{2n}{3} \geq |S| \geq \frac{n}{3}$. Otherwise, S is unbalanced. The following is a region growing approach similar to the single commodity flow. Set

$$\alpha = \frac{O\left(n^2 \cdot f^* \log n\right)}{m}$$

Pick a source in graph G^+, grow the region hop by hop. Each hop we grow the region by $\bar{d}$ which is the distance weight of edges in G^+. After each hop, we compare the ratio $\frac{\#crossing\ edges}{\#internal\ edges}$ with α. If the ratio is greater then α, we keep on growing. If not, we stop. Pick a new source, which is not covered by any region, using the same approach to grow, until we all the sources are covered by a region. After the region growing step, there are two possible cases.

case 1: For all the regions $R_i's$, the number of nodes from graph G is smaller or equal to $\frac{2n}{3}$. In this case, we can group the regions into balanced set in the following way. If there is a region that is larger than $\frac{n}{3}$, then we can just pick this region as S and group the other regions as $\bar{S}$. If all of the regions are smaller than $\frac{n}{3}$, then we can greedily pack regions to S until the size of S is larger than $\frac{n}{3}$. Because each region has the property that $\frac{\#crossing\ edges}{\#internal\ edges} \leq \alpha$, we know if we group some regions together to a bigger region R, it still have this property, which means

$$\frac{\#crossing\ edges}{\#internal\ edges} \leq \alpha = \frac{O\left(n^2 \cdot f^* \log n\right)}{m}$$

So,

$$\delta(R) = \#crossing\ edges \leq \frac{O\left(n^2 \cdot f^* \log n\right)}{m} \cdot (\#internal\ edges) \leq O\left(n^2 \cdot f^* \log n\right)$$

And $|R| \cdot |\bar{R}| = O\left(n^2\right)$. We know this cut

$$\eta = \frac{\delta(R)}{|R| \cdot |\bar{R}|} = O\left(f^* \log n\right)$$

This means we find a cut of value $O\left(f^* \log n\right)$.

case 2: There is a region R that contains more than $\frac{2n}{3}$ nodes from graph G. Each step we grow the region R, its ratio $\frac{\#corssing\ edges}{\#internal\ edges} \geq \alpha$ until i^{th} step. So the total number of edges included in this region after each step is at least 1, $(1+\alpha)$, $(1+\alpha)^2, \ldots, (1+\alpha)^i$. From $(1+\alpha)^i \leq m$, we have

$$i \leq \frac{log 2m}{\log(1+\alpha)} \approx \frac{\log 2m}{\alpha} = \frac{m \cdot \log 2m}{n^2 \cdot f^* \log n} = O\left(\frac{m}{f^* n^2}\right)$$

which mean the number of hops of this region is at most $\frac{m}{f^* n^2}$. Then the radius r of this region

$$r \leq \bar{d} \cdot \frac{m}{f^* n^2} = \frac{f^*}{m} \cdot \frac{m}{f^* n^2} = O\left(\frac{1}{n^2}\right)$$

We can have $r \leq \frac{1}{2n^2}$ by setting appropriate α value. By metric property, we know for any nodes u and v,

$$dist(u,v) \leq dist(u,R) + dist(v,R) + \frac{1}{2n^2}$$

From the constraint $\sum_i l_i \geq 1$ in the dual, we have

$$1 \leq \sum_{u,v} dist(u,v) \leq 2n \sum_u dist(u,R) + \frac{1}{2}$$

So,

$$\sum_{u \notin R} dist(u,R) \geq \frac{1}{4n}$$

Look at figure 1. Let n_i be the number of nodes that are at least i hops away from the region R, and r_i be the ratio of #edges crossing cut i to the demand crossing cut i. Then $\#crossing\ edges \geq r_i \cdot \frac{2n}{3} \cdot n_i$ since the demand crossing cut is $\leq \frac{2n}{3} \cdot n_i$. And its contribution to the left hand side of the above inequality is $n_i \cdot \frac{f^*}{m}$. By average argument (contribution per edge), there is at least one cut i for which

$$\frac{n_i \cdot \frac{f^*}{m}}{r_i \cdot \frac{2n}{3} \cdot n_i} \geq \frac{1/4n}{m}$$

So we have

$$r_i \leq 6f^*$$

Combining the above two cases, we can conclude that $\eta \leq O\left(f^* \log n\right)$.

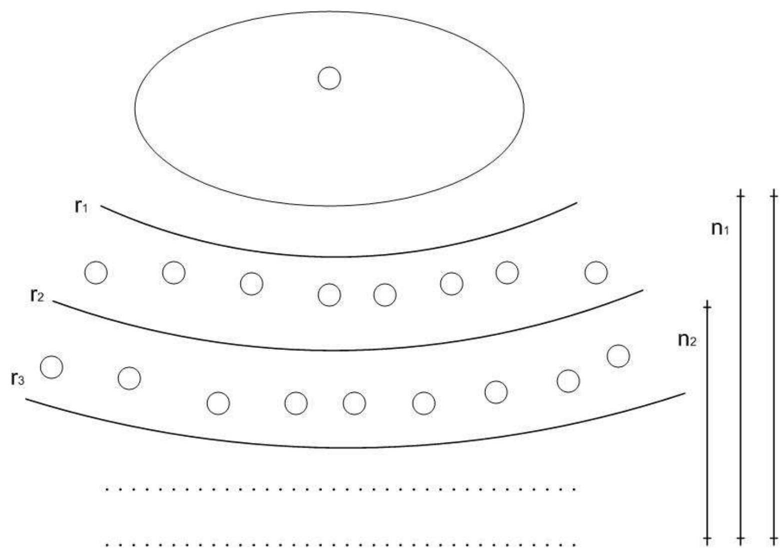

Figure 1: One region has more than $2n/3$ nodes