Lecture 25 : The Sparsest Cut Problem

We saw the following relaxation of the sparsest cut problem (in Lect. 24 notes):

$$\min \sum_{e \in E} x_e$$

$$\text{s.t.} \quad \sum_{i,j \in V} d_x(i,j) \geq 1$$

$$x_e \geq 0 \text{ for all } e \in E,$$

where $d_x(i,j)$ is the shortest path distance between i and j, using x_e as length of edge $\forall e \in E$.

Let us introduce new variables y_{ij} for all $i,j \in V$ along with the constraints $y_{ij} \leq x_{e_1} + x_{e_2} + \dots + x_{e_k}$ for all possible simple paths $e_1 - e_2 - \dots - e_k$ between i and j. We also have the constraint $\sum_{i,j \in V} y_{ij} \geq 1$.

We also have the compact formulation (note that the above formulation has exponentially many constr.) $y_{ij} \leq y_{jk} + y_{ik}$ where ~~$(i,k) \in E$ and $j \in V$~~ $i, j, k \in V$ (please check this) and $y_{ij} \leq x_{ij}$ $\forall (i,j) \in E$.

- Let x^* be an optimal solution of the above LP.
- Partition G into clusters such that
  - (i) each cluster has diameter (wrt edge lengths as given by x^*) at most $D = 1/4n^2$;
  - (ii) there are at most $\alpha \sum_{e \in E} x^*_e$ intercluster edges with $\alpha = \frac{4 \ln n}{D}$.

Lemma 1. Either there is a cluster C with $|C| \geq \frac{2n}{3}$ or we can efficiently find $S \subseteq V$ s.t. $\rho(S) = O(\log n) \cdot \sum_{e \in E} x_e^* \leq O(\log n) \cdot \rho(G)$.

Proof. Suppose there is no cluster of size $\geq \frac{2n}{3}$. Order the clusters by increasing size and add clusters in this order to S until $|S| \geq n/3$. Then it must be the case that $|V-S| \geq n/3$ as well. So we have:

$$\rho(S) = \frac{|\delta(S)|}{|S| \cdot |V-S|} \leq \frac{\alpha \cdot \sum_{e \in E} x_e^*}{n^2/9}$$

(since $|\delta(S)| \leq$ no. of intercluster edges)

$$\leq \frac{4n^2 \cdot 4 \ln n \cdot \rho(G)}{n^2/9} = O(\log n) \cdot \rho(G).$$

So when there is no cluster of size $\geq 2n/3$, we have found an $O(\log n)$-approximation for sparsest cut. What about the case when there is such a cluster?

Lemma 2. If there exists a cluster C such that $|C| \geq 2n/3$ and with diameter $\leq 1/4n^2$, then we can efficiently find S s.t. $\rho(S) \leq 6 \cdot \sum_{e \in E} x_e^*$.

Proof. Let $d(i, C) = \min_{j \in C} d_{x^*}(i, j)$. Order the vertices as $i_1, \ldots, i_n$ in decreasing order of $d(i, C)$. Let $S_k = \{i_1, \ldots, i_k\}$ for $k = 1, 2, \ldots, n-1$.

Claim. $\sum_{i,j\in V} |d(i,C) - d(j,C)| \geq 1/6.$

Let us assume the above claim and finish the proof of the lemma. Then we'll prove this claim. The intuition behind the claim is that since C has many vertices and a small diameter, for the constraint $\sum_{i,j} d_{x^*}(i,j) \geq 1$ to hold, it has to be the case that the total of the difference between the distance to C for each pair of vertices must be large.

Assuming the above claim, we have:

$$\min_{1\leq k\leq n-1} \rho(S_k) = \min_{1\leq k\leq n-1} \frac{|\delta(S_k)|}{|S_k|\cdot|V-S_k|}$$

$$\leq \frac{\sum_{k=1}^{n-1} \left(-d(i_{k+1},C) + d(i_k,C)\right)\cdot|\delta(S_k)|}{\sum_{k=1}^{n-1} \left(-d(i_{k+1},C) + d(i_k,C)\right)\cdot|S_k||V-S_k|}$$

$$= \frac{\sum_{\substack{(i_\ell,i_h)\in E\\ 1\leq \ell<h\leq n}} d(i_\ell,C) - d(i_h,C)}{\sum_{\substack{i_\ell,i_h\\ 1\leq \ell<h\leq n}} d(i_\ell,C) - d(i_h,C)}$$

$$\leq \sum_{e\in E} x^*_e \Big/ 1/6$$

(since $d(i_\ell,C) - d(i_h,C) \leq x^*_{(i_\ell,i_h)}$)

→ due to the claim

$$= 6 \sum_{e \in E} x_e^* \leq 6 \cdot \rho(G).$$

It remains to prove the claim. Pick some $i' \in C$.
For any $i \in V$, there is some $j \in C$ such that
$d(i, C) = d_{x^*}(i,j)$. So $d_{x^*}(i, i') \leq d_{x^*}(i,j) + d_{x^*}(j, i') \leq d(i, C) + \frac{1}{4n^2}$.

$$\text{Then } 1 \leq \sum_{i,j} d_{x^*}(i,j) \leq \sum_{i,j} \left(d_{x^*}(i,i') + d_{x^*}(i',j) \right)$$

$$= 2n \sum_{i \in V} d_{x^*}(i, i')$$

$$\leq 2n \sum_{i \in V} \left(d(i, C) + \frac{1}{4n^2} \right)$$

$$= 2n \sum_{i \in V} d(i, C) + 1/2.$$

$$\text{Therefore } \sum_{i \in V} d(i, C) = \sum_{i \notin C} d(i, C) \geq 1/4n$$

$$\text{Then } \sum_{i,j} \left| d(i, C) - d(j, C) \right| \geq \sum_{i \notin C,\, j \in C} d(i, C)$$

$$= |C| \cdot \sum_{i \notin C} d(i, C) \geq \frac{2n}{3} \cdot \frac{1}{4n} = \frac{1}{6}.$$

Thus we have an $O(\log n)$-approximation algorithm for the sparsest cut problem.

To improve this result, let us look at another relaxation of the sparsest cut problem.

$$\min \sum_{e \in E} x_e$$

$$\text{s.t. } \sum_{i,j} d_x(i,j) \geq 1$$

$$d_x \text{ is an } \ell_2\text{-squared metric.}$$

We can write this as a vector program (next lecture), which can be solved as an SDP in polynomial time. We'll use the almost optimal soln. to this SDP to find a sparse cut.