Lecture 26: The Sparsest Cut Problem

We will now look at another relaxation for the sparsest cut problem: ($G = (V, E)$ is the i/p)

$$\min \sum_{(i,j) \in E} d(i,j)$$

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

and d is an ℓ_2^2-metric

That is, we want $d(i,j) = \| f(i) - f(j) \|^2 \;\; \forall i,j \in V$ for some mapping $f : V \to \mathbb{R}^n$.

Let us show that the above program is a relaxation of the sparsest cut problem. Let $(S^*, V-S^*)$ be a sparsest cut in G. Let $f(i) = (0, \ldots, 0) \;\forall i \in S^*$ and $f(j) = \left(\frac{1}{\sqrt{|S^*| \cdot |V-S^*|}}, 0, \ldots, 0\right) \;\; \forall j \in V-S^*$.

Easy to see that $\sum_{i,j \in V} \| f(i) - f(j) \|^2 = 1$.

Moreover, $\sum_{(i,j) \in E} \| f(i) - f(j) \|^2 = \frac{|\delta(S^*)|}{|S^*| \cdot |V-S^*|} = \rho(G)$.

Let us look at the following program where we make the constraint $\sum_{i,j \in V} d(i,j) \geq 1$ tight.

It will also be useful to scale distances up by n^2 and the objective function down by n^2.

So our program is as given below:

$$\min \frac{1}{n^2} \sum_{(i,j)\in E} d(i,j)$$

s.t.

$$\sum_{i,j\in V} d(i,j) = n^2$$

d is an ℓ_2^2-metric.

The above program is a relaxation of the sparsest cut problem and we can rewrite it as the following vector program:

$$\min \frac{1}{n^2} \sum_{(i,j)\in E} \| v_i - v_j \|^2$$

s.t.

$$\sum_{i,j\in V} \| v_i - v_j \|^2 = n^2$$

$$\| v_i - v_j \|^2 + \| v_j - v_k \|^2 \geq \| v_i - v_k \|^2 \quad \forall i,j,k$$

$$v_i \in \mathbb{R}^n \quad \forall i.$$

This vector program can be written and solved as an SDP in polynomial time to obtain an almost optimal solution. We want to use the optimal soln. to this SDP to find a sparse cut.

Claim. If there is some $i \in V$ such that $|\{ j \in V : \| v_i - v_j \|^2 \leq 1/4 \}| \geq n/4$ then we can efficiently find a subset S of V such that

$$\rho(S) \leq \frac{16}{n^2} \sum_{(i,j)\in E} \| v_i - v_j \|^2 \leq 16 \cdot \rho(G).$$

For simpler notation, let us use $d(i,j)$ to denote $\| v_i - v_j \|^2$.

Proof. Let $C = \{ j \in V : \|v_i - v_j\|^2 \leq 1/4 \}$.

We have $n^2 = \sum_{j,k \in V} d(j,k) \leq \sum_{j,k \in V} \left(d(j,i) + d(i,k) \right)$

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

So $n^2 \leq 2n \sum_{j \in V} d(j,C) + n^2/2$.

Thus $\sum_{j \in V} d(j,C) \geq n/4$.

Then $\sum_{j,k \in V} |d(j,C) - d(k,C)| \geq \sum_{\substack{j \notin C \\ k \in C}} d(j,C)$

$$= |C| \cdot \sum_{j \notin C} d(j,C) \geq \frac{n}{4} \cdot \frac{n}{4} = \frac{n^2}{16}.$$

Let us order the vertices as $j_1, \dots, j_n$ in decreasing order of $d(j,C)$. Let $S_k = \{j_1, \dots, j_k\}$ for $k = 1, \dots, n-1$.

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

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

$$= \frac{\sum_{\substack{(j_p, j_q) \in E \\ p > q}} \left(d(j_p, C) - d(j_q, C) \right)}{\sum_{\substack{j_p, j_q \in V \\ p > q}} \left(d(j_p, C) - d(j_q, C) \right)} \leq \frac{\sum_{e \in E} d(e)}{n^2/16}$$

$$\leq 16 \cdot \rho(G).$$

Let $B(i, r) = \{ j \in V : d(i,j) \leq r \}$.
Given the above claim, we'll assume from now on that $|B(i, 1/4)| < n/4 \ \forall i \in V$. Otherwise we can easily find a cut $(S, V-S)$ with $\rho(S) \leq 16 \cdot \rho(G)$.

Claim. There exists $o \in V$ with $|B(o, 4)| \geq 3n/4$.
Furthermore, let $\mathcal{U} = B(o, 4) - B(o, 1/4)$.
Then $|\mathcal{U}| \geq n/2$ and for all $i \in \mathcal{U}$, there exist $\geq n/4$ vertices $j \in \mathcal{U}$ s.t. $d(i,j) > 1/4$.

Proof. Suppose no such o exists. Then for all $i \in V$, we have more than $n/4$ vertices at distance > 4. So:

$$\sum_{i,j \in V} d(i,j) = \sum_{i \in V} \left(\sum_{j \in V} d(i,j) \right) > n \cdot 4 \cdot \frac{n}{4} = n^2.$$

This contradicts the feasibility of v to our vector program. So there exists some $o \in V$ s.t. $|B(o, 4)| \geq \frac{3n}{4}$. Since $|B(o, 1/4)| \leq n/4$ by our initial assumption, for $\mathcal{U} = B(o, 4) - B(o, 1/4)$, $|\mathcal{U}| \geq |B(o,4)| - |B(o, 1/4)|$

$$\geq 3/4 \ n - n/4 = \frac{n}{2}$$

Finally, pick any $i \in \mathcal{U}$. Since $|B(i, 1/4)| < n/4$ while $|\mathcal{U}| \geq n/2$, there are $\geq n/4$ $j \in \mathcal{U}$ s.t. $d(i,j) > 1/4$. ∎

Since solutions to vector programs are invariant under translation, we can move o to the origin. This allows us to simplify notation and say that $\|v_i - v_o\|^2 = \|v_i\|^2$ for all i.

The ARV Algorithm

0. Let $v_1, \dots, v_n$ be the vectors obtained by solving the SDP.

1. Pick a random vector $r \in \mathbb{R}^n$ s.t. $r_i \sim N(0,1)$.

2. Let $\mathcal{L} = \{ i \in V : v_i^T r \leq -1 \}$
and $R = \{ i \in V : v_i^T r \geq 1 \}$.

3. Find a maximal matching $M \subseteq \{ (i,j) \in \mathcal{L} \times R : d(i,j) \leq \Delta \}$.

4. Let $\mathcal{L}' \subseteq \mathcal{L}$ and $R' \subseteq R$ be the sets of unmatched vertices.

5. Sort vertices by increasing distance to $\mathcal{L}'$ to get $i_1, i_2, \dots, i_n$.

6. Let $S_k = \{ i_1, \dots, i_k \}$ for $1 \leq k \leq n-1$.
Return $S = \arg\min_{1 \leq k \leq n-1} f(S_k)$.

A simple observation is that for any $i \in \mathcal{L}'$ and $j \in R'$ we have $d(i,j) > \Delta$.

The following theorems can be shown.

Theorem 1. $\Pr[|\mathcal{L}| > cn \text{ and } |R| > cn] \geq c$ for some positive constant c.

Theorem 2. (Structure theorem) For $\Delta \geqslant \dfrac{c_0}{\sqrt{\log n}}$ for some constant c_0, we have

$$E[|M|] \leqslant \left(\frac{c}{2}\right)^2 n.$$

Using Theorem 2 and Markov Inequality, we have

$$Pr\left[|M| \geqslant \frac{c}{2}\cdot n\right] \leqslant \frac{E[|M|]}{\frac{c}{2}\cdot n} \leqslant \frac{c}{2}.$$

If the size of the matching $\leqslant \frac{cn}{2}$ (which happens with probability $\geqslant 1 - c/2$) while the sizes of $\mathcal{L}$ and R are at least cn (which happens with prob. $\geqslant c$), we can conclude that $|\mathcal{L}'| \geqslant \frac{cn}{2}$ and $|R| \geqslant \frac{cn}{2}$ with constant probability.

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

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

$$= \frac{\sum_{(i,j)\in E}|d(i,\mathcal{L}') - d(j,\mathcal{L}')|}{\sum_{i,j\in V}|d(i,\mathcal{L}') - d(j,\mathcal{L}')|} \leqslant \frac{\sum_{(i,j)\in E}\|v_i - v_j\|^2}{\sum_{i\in\mathcal{L}'}\sum_{j\in R'} d(j,\mathcal{L}')}$$

$$\leqslant \frac{\sum_{(i,j)\in E}\|v_i - v_j\|^2}{|\mathcal{L}'|\cdot|R'|\cdot\Delta} = O(\sqrt{\log n})\cdot\frac{1}{n^2}\cdot\sum_{(i,j)\in E}\|v_i - v_j\|^2$$

$$\leqslant O(\sqrt{\log n})\cdot\rho(G).$$