Lecture 7: Sparsest Cut & Metric Embeddings

Last time considered $\text{sparsity}(S) = \phi(S) = \dfrac{\text{cap}(S,\bar{S})}{|S|\cdot|\bar{S}|}$ for a graph with edge-capacity $c_e \ \forall e \in E$.

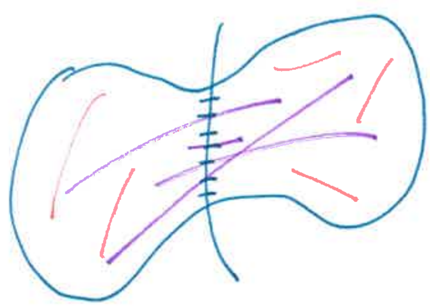
Wanted to find $\min_{S \subseteq V} \phi(S)$. (Sparsest Cut)

Let's generalize a bit: sps. demands on vertex pairs

$D_{ij} \geq 0$ for $i,j \in V$.

generalized sparsity(S) $\dfrac{\text{cap}(S,\bar{S})}{\text{dem}(S,\bar{S})} = \dfrac{\sum_{e \in E(S,\bar{S})} \text{cap}(e)}{\sum_{i \in S,\, j \in \bar{S}} D_{ij}}$

counted here



: demands separated

: demands not separated

: edges cut

When nonzero demand between single pair s,t $\quad D_{st} = 1$ (say) $\quad D_{ij} = 0 \ \forall$ all other ij

$\Rightarrow$ min s-t-cut problem

poly time solvable. [Ford Fulkerson, Edmonds-Karp, etc]

When non-zero demands b/w $D_{s_1 t_1}, D_{s_2 t_2}$ all others zero

$\Rightarrow$ still poly time solvable [T.C. Hu, Rothschild-Winston, Seymour]

But more general demand patterns: NP-hard.

In fact, reduction from MaxCut hardness, see HW #2.

Today: $O(\log k)$ approximation for (generalized) sparsest cut

$\Rightarrow$ another proof for theorem from last time.

for (uniform) sparsest cut where $D_{ij} = 1 \ \forall ij$.

$k = \#$ of pairs $s_i t_i$ st $D_{s_i t_i} \neq 0$.

Using embeddings of metrics into geometric spaces

Metric Relaxation

$$\min \sum_{ij \in E} c_{ij} y_{ij}$$

$$\text{s.t.} \sum_{ij \in V \times V}' D_{ij} y_{ij} = 1$$

y is a metric, $y_{ij} \geq 0$, $y_{ij} \leq y_{ik} + y_{kj}$ $\forall k, ij$

Ideal Formulation: y is a metric that is an indicator of a cut

this is what we want!

i.e. $\exists$ Set $S \subseteq V$ st
$$y_{ij} = \begin{cases} 1 & \text{if } |\{i j\} \cap S| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note: y is a metric (δ_S)

δ_S: Call it the cut metric corresponding to set S.

We're just dropping the "cut" requirement in the metric relaxation.

So sparsest cut $= \min\limits_{\substack{y \in \text{cut} \\ \text{metrics}}} \dfrac{c^T y}{D^T y}$

think of c, D, y all being vectors in $\mathbb{R}^{\binom{n}{2}}$.

LP relaxation $= \min\limits_{y \in \text{metrics}} \dfrac{c^T y}{D^T y}$.

It's annoying to deal with discrete objects, so let's convexify the set of cut metrics.

$$K_n = \text{"cut cone"} = \sum_{S \subseteq [n]} \underbrace{\alpha_S}_{\text{non negative values}} \cdot \underbrace{\delta_S}_{\text{cut metrics.}}$$

Fact: K_n is generated by cut metrics, so $K \subseteq$ metric cone$_n$ = set of all metrics on n points

$\begin{cases} x \in \text{cone} \Rightarrow \lambda x \in \text{cone}\ \forall \lambda \geq 0 \\ x, y \in \text{cone} \Rightarrow x+y \in \text{cone}. \end{cases}$

Fact: K_n = the set of n point sub-metrics of $(\mathbb{R}^{f(n)}, \ell_1)$

$\|x-y\|_1 = \sum_t |x_t - y_t|$

"the cut cone is exactly all n-point submetrics of ℓ_1, the Manhattan metric".

Pf: ($\Rightarrow$) sps. $d \in K_n \Rightarrow d = \sum_{S \subseteq [n]} \alpha_S \delta_S$. $\alpha_S \geq 0$.

Now use a coordinate for each S such that $\alpha_S > 0$.

$$\text{vertex } i \xmapsto{\varphi} (\alpha_{S_1} \mathbb{1}(i \in S_1), \alpha_{S_2} \mathbb{1}(i \in S_2) \ldots \ldots)$$

$$j \xmapsto{\varphi} (\alpha_{S_1} \mathbb{1}(j \in S_1), \ldots \ldots \ldots)$$

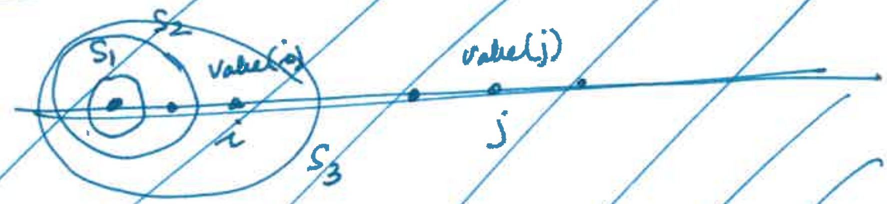
$$\|\varphi(i) - \varphi(j)\|_1 = \sum \alpha_S \cdot |\mathbb{1}(i \in S_1) - \mathbb{1}(j \in S_1)|$$

$$= \sum \alpha_S \cdot \delta_S(i,j) = d(i,j).$$

$\Rightarrow d \in \ell_1$.

(⇐). Want to show: $d \in \ell_1 \Rightarrow d$ = sum of cut metrics. ④

~~Sps. 1-dim ℓ_1, so. $d(i,j) = |\text{value}(i) - \text{value}(j)|$~~



~~Now: define cut metrics δ_{S_i} S_i = points taking on 1st i values.~~

By example: $\begin{cases} \text{1 dim} \\ \text{sps. 4 points} \end{cases}$

sorted order.

$$\begin{aligned} i &= 0{\cdot}3 \\ j &= 0{\cdot}7 \\ k &= 0{\cdot}7 \\ \ell &= 1{\cdot}4. \end{aligned}$$

$d(i,j) = |0{\cdot}3 - 0{\cdot}7| = 0{\cdot}4$

$d(\ell,j) = |1{\cdot}4 - 0{\cdot}7| = 0{\cdot}7$

etc.

define $S_1 = \{i\}$

$S_2 = \{i,j,k\}$

$$d = (0{\cdot}7 - 0{\cdot}3)\,\delta_{S_1} + (1{\cdot}4 - 0{\cdot}7)\,\delta_{S_2}.$$

Now: do each coordinate separately $\|x-y\|_1 = \sum_i |x_i - y_i|$

and each coordinate embedded like above

Note: this is algorithmic. So given d, can get $\sum \alpha_S \delta_S$ in time $\tilde{O}(n \text{ \#dims})$

So. $$\text{Sparsest cut} = \min_{y \in \ell_1} \frac{c^T y}{D^T y}$$

$$\text{LP relaxation} = \min_{y \in \text{metrics}} \frac{c^T y}{D^T y}.$$

what's the gap?

Thm [Bourgain]: "every finite metric is close to being in ℓ_1"

Every n-point metric embeds into ℓ_1 with distortion $O(\log n)$.

Namely: sps. y is a metric on n points V

then $\exists$ map $\varphi : V \to \mathbb{R}^K$ s.t. ← some dimension, we will see

$$\frac{y_{ij}}{O(\log n)} \leq \| \varphi(i) - \varphi(j) \|_1 \leq y_{ij}$$

[Linial London Rabinovich] And we can compute in poly time.

Corollary: $\exists$ an $O(\log n)$ approximation to (generalized) sparsest cut.

Pf: $$LP = \frac{c^T y^*}{D^T y^*} \geq \frac{c^T d}{(D^T d)} \cdot \frac{1}{O(\log n)} = \frac{c^T (\sum_i \alpha_S \delta_S)}{D^T (\sum_i \alpha_S \delta_S)} \cdot \frac{1}{O(\log n)}$$

say y^* = optimal solution of LP.

say d is the ℓ_1 distance given by Bourgain

But $d \in$ cut cone $\Rightarrow d = \sum_i \alpha_S \delta_S$ for some cut metrics.

$$= \frac{\sum_i \alpha_S . (c^T \delta_S)}{\sum_i \alpha_S (D^T \delta_S)} \cdot \frac{1}{O(\log n)} \geq \min_{S^* : \alpha_{S^*} > 0} \frac{c^T \delta_{S^*}}{D^T \delta_{S^*}} \cdot \frac{1}{O(\log n)}$$

linearity of the vector multiplication ~~inner product~~

Choose the best cut from conic combo.

Note: you can do this decomposition $d = \sum_i \alpha_S \delta_S$ using the equivalence proof above

$$\Rightarrow \frac{c^T \delta_{S^*}}{D^T \delta_{S^*}} \leq O(\log n) . LP$$

☺

Bourgain's theorem says: $\forall y \in \text{metric}_n$, $\exists d \in \ell_1$

s.t. d and y are close (multiplicatively)

And then sparsest cut can be rounded using this "embedding" of $y \longrightarrow d$.

Only piece remains: How to prove Bourgain's thm (for ℓ_1)?

Several ways. Here's his original proof (almost)

- Create $O(\log n)$ ~~coords~~ groups of $O(\log n)$ coordinates each.

  for $g = 1$ to $\log_2 n$.

  for rep $= 1$ to $\Theta(\log n)$

  $E|R_{g,rep}| = n/2^g$.

  pick set $R_{g,rep}$ = pick each point w.p. $1/2^g$ indep

  set $\varphi_{g,rep}(i)$ = distance of i to closest point in $R_{g,rep}$.

Claim: w.h.p. $\|\varphi(i)-\varphi(j)\|_1 \geq \cancel{y_{ij}/\Theta(\log n)}\ y_{ij} \cdot \Omega(\log n)$ (*)

w.p. 1 $\|\varphi(i)-\varphi(j)\|_1 \leq \Theta(\log^2 n) \cdot y_{ij}$ (**)

Pf: (**) follows from the triangle inequality that in each coord $|\varphi_{g,r}(i) - \varphi_{g,r}(j)| \leq d(i,j)$

and $\exists$ $\Theta(\log^2 n)$ coords.

So remains to show: that we do get "reasonable" contribution to the distance from this embedding.

<u>Define</u>: r_t = smallest radius so that $|B(i,r)|, |B(j,r)| \geq 2^t$

$$r_0 = 0 \leq r_1 \leq r_2 \leq \dots \leq r_{f-1} < \frac{y_{ij}}{4} \leq r_f$$

Consider some r_t in early part.

So that $B(i, r_t)$ $B(j, r_t)$ disjoint

Say r_t was "defined" by i s.t. $\overset{\text{open}}{B}(i, r_t) < 2^t$

~~but~~ then $B(j, r_{t-1}) \geq 2^{t-1}$

Now if we sample R at rate $1/2^t$,

- $\Pr[R \text{ misses } B(i, r_t) \text{ and } R \text{ hits } B(j, r_{t-1})] \geq$
  $$\left(1 - \frac{1}{2^t}\right)^{2^t} \cdot \left(1 - \left(1 - \frac{1}{2^t}\right)^{2^{t-1}}\right) \geq \Omega(1)$$

random set

- If R satisfies this, then distance$(i, R) \geq r_t$
  distance$(j, R) \leq r_{t-1}$
  $\Rightarrow$ coordinate for R gives $|\varphi(i) - \varphi(j)| \geq r_t - r_{t-1}$

"good" event

- "good" event happens for rate $1/2^t$ with constant probability.
  $\Rightarrow$ a constant fraction of those coordinates are "good."
  $\Rightarrow$ give $\Omega(\log n) \cdot (r_t - r_{t-1})$ contribution.

Summing over all sampling scales gives

$$\Omega(\log n)\left[(r_1-r_0)+(r_2-r_1)+\cdots+(r_{f-1}-r_{f-2})+\left(\frac{y_{ij}}{4}-r_{f-1}\right)\right]$$

$$\geq \Omega(y_{ij}\cdot \log n).$$

as claimed, proves Bourgain's Thm. (Scaledown by $O(\log^2 n)$) ☺

— x —

- Similar argument shows:- $\|\varphi(i)-\varphi(j)\|_p$ is also a $O(\log n)$- ~~log n~~ distortion embedding.

  for other norms p.

  (use Cauchy-Schwarz or Hölder).

- A different argument for ℓ_1 follows from approximating metrics by random distribution over trees. [Bartal, Fakcharoenphol Rao Talwar].

- In fact, you need ~~to~~ only ensure that $\|\varphi(i)-\varphi(j)\| \geq \dfrac{y_{ij}}{\Omega(\log n)}$

  only for ij st $D_{ij}>0$

  so reduce $\log n \to \log k$.

- Integrality gap? The last lecture we saw a gap of $\Omega(\log n)$ for uniform sparsest cut. That is a special case so the lower bound on the gap still holds here.

- What next? Better LP? SDP?

  Let's see SDPs and eigenvalue/eigenvector (spectral) ideas next!