Lecture 7: Sparsest Cut & Metric Embeddings

Last time considered sparsity \( s = \frac{\text{cap}(S, \bar{S})}{\phi(S)} \) for a graph with edge-capacity \( c_{e \in E} \).

Wanted to find \( \min \sum_{e \in E} \phi(e) \) (Sparsest Cut).

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

\( \Delta_{ij} \geq 0 \) for \( i, j \in V \).

Generalized sparsity \( s = \frac{\text{cap}(S, \bar{S})}{\text{dem}(S, \bar{S})} \) = \( \frac{\sum_{e \in E(S, \bar{S})} \text{cap}(e)}{\sum_{j \in S} \Delta_{ij}} \)

\( \text{dem}(S, \bar{S}) \) = \( \sum_{j \in S} \Delta_{ij} \)

\( \Delta_{ij} \) : demands separated
\( \Delta_{ij} \) : demands not separated
\( e \) : edges cut

When demand between single pair \( s, t \)

\( D_{st} = 1 \) (say) \( D_{ij} = 0 \) \( i \neq j \)

\( \Rightarrow \) min s-t-cut problem

poly time solvable. \cite{Fulkeron,Edmonds-Karp,ek}

When non-zero demands \( \times \) \( D_{s, t_1}, D_{s, t_2} \) all others zero

\( \Rightarrow \) still poly time solvable \cite{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

\[ \rightarrow \text{another proof for theorem from last time.} \]

\[ \text{for (uniform) sparsest cut where} \]

\[ D_{ij} = 1 \text{ if } i \neq j. \]

Using embeddings of metrics into geometric space.

Metric Relaxation

\[ \min \sum_{ij \in E} C_{ij} y_{ij} \]

\[ \text{s.t. } \sum_{ij \in V \times V} D_{ij} y_{ij} = 1 \]

\[ y_i \text{ is a metric, } y_{ij} > 0, y_{ij} \leq y_{ik} + y_{kj} \]

Ideal Formulation, \( y \) is a metric that is an indicator of a cut

\[ \text{i.e. } \exists S \subseteq V \text{ s.t. } y_{ij} = \begin{cases} 1 & \text{if } |S \cap i \cap S \cap j| = 1 \\ 0 & \text{otherwise.} \end{cases} \]

Note: \( y \) is a metric (\( S \))

\( S \) call it the cut metric corresponding to set \( S \).

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

K = \# of pairs \( i \neq j \) s.t. \( D_{ij} \neq 0 \).
So span over cut = \[ \min_{y \in \text{cut metric}} \frac{c^T y}{B^T y} \]

LP relaxation = \[ \min_{y \in \text{metrics}} \frac{c^T y}{B^T y} \]

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

\[ K_n = \text{"cut cone"} = \sum_{S \subseteq \mathcal{A}} \alpha_S \cdot S \]

Fact: \( K_n \) is generated by cut metrics, so \( K \leq \text{metric cone}_n = \text{set of all metrics on } n \text{ points} \)

Fact: \( K_n = \) the set of \( n \) point submetrics of \((\mathbb{R}, l_1)\) for \( \|x - y\|_1 = \sum_e |x_e - y_e| \)

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

Pf: \( (\Rightarrow) \) sps. \( d \in K_n \Rightarrow d = \sum_{S \subseteq \mathcal{A}} \alpha_S \cdot S. \quad \alpha_S \geq 0. \)

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

\[ \text{vertex } i \mapsto (\alpha_S \cdot 1(i \in S_1), \alpha_S \cdot 1(i \in S_2), \ldots) \]
\[ j \mapsto (\alpha_S \cdot 1(j \in S_1), \ldots) \]

\[ \|\phi(i) - \phi(j)\|_1 = \sum_{i \in S_1} \alpha_s |1(i \in S_1) - 1(j \in S)| \]

\[ = \sum_{i \in S} \alpha_S \cdot S(i,j) = d(i,j). \]

\[ \Rightarrow d \in l_1. \]
\(\leq\). Want to show: \(x \in \mathbb{E}_1 \Rightarrow d = \sum_{i}^n \text{ cut metrics.}\)

\[
\text{Sp. 1-dim } \mathbb{E}_1, \text{ so. } d(i,j) = |\text{value}(i) - \text{value}(j)|
\]

Now: define \(\text{cut metrics } S_{i,j}\).

\[
\text{By example: } \{ \text{sp. } \& \text{ point } \}
\]

\[
\begin{align*}
&i = 0.3 \\
&j = 0.7 \\
&k = 0.7 \\
&l = 1.4.
\end{align*}
\]

\[
\text{define } S_1 = \{ i, j, k \} \\
S_2 = \{ i, j, k \} \\
S_3 = \{ i, j, k \}
\]

\[
d = (0.7 - 0.3) S_1 + (1.4 - 0.7) S_2.
\]

Now: do each coordinate separately:

\[
\|x - y\|_1 = \sum_{i}^n |x_i - y_i|
\]

Note: this is algorithmic. So given \(d\), we can get indices \(i^*\) in the \(B_1\) (in #dim) above.

So: sparsest cut = \(\min_{y \in \mathbb{E}_1} \frac{c^T y}{D^T y}\)

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

What's the gap?
Theorem [Bourgain]: "Every metric is close to being in \( L^1 \)"

Every \( n \)-point metric embeds into \( L^1 \) with distortion \( O(\log n) \).

Namely: \( \forall \) metric \( n \) point \( V \) then \( \exists \) map \( \phi: V \to \mathbb{R}^k \) such that

\[
\forall i, j \in V \quad d_{ij} \leq \| \phi(i) - \phi(j) \|_1 \leq 2d_{ij} / \log n
\]

[Linial London Rabnovich] and we can compute in poly time.

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

Proof: \( LP = \sum^k \frac{d_{ij}}{\phi(i)} \geq \frac{c^T d}{(D^T d) \cdot O(\log n)} \geq \frac{c^T \sum \delta_{s_1} \delta_{s_2}}{D^T (\sum \delta_{s_1} \delta_{s_2}) \cdot O(\log n)} \).

Say \( \delta \) is optimal solution to \( LP \).

\( d \) is the \( L_1 \) distance given by Bourgain.

Choose the best cut from conic combo.

\( \Rightarrow \frac{c^T \delta_{s_1} \delta_{s_2}}{D^T \delta_{s_1} \delta_{s_2} \cdot O(\log n)} \).

Note: you can do this decomposition \( d = \sum \delta_{s_1} \delta_{s_2} \) using the equivalence proof above
Bourgain's theorem says: \[ y \in \mathbb{R}^m, \quad f \in L_2, \]

st. \( d \) and \( y \) are close (multiplicatively)

And then sparsest cut can be rounded using this 
"embedding" of \( y \rightarrow d \).

Only piece remains: How to prove Bourgain's thm \( \Omega(\log n) \) (for \( \ell_1 \))?

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

- Create \( O(\log n) \) group of \( O(\log n) \) coordinates each.
  
  \[
  \text{for } g = 1 \text{ to } \log_2 n, \\
  \text{for } \text{rep} = 1 \text{ to } \Theta(\log n), \\
  \text{pick set } R_{g, \text{rep}} = \text{pick each point rep. } \frac{1}{2^g} \text{ indep} \\
  \text{set } \rho(i) = \text{distance of } i \text{ to closest point in } R_{g, \text{rep}}.
  \]

Claim: \[
\begin{align*}
\| \rho(i) - \rho(j) \|_2 \leq \frac{\log n}{2^g} & \quad \forall i, j \in \mathbb{R}^{\log n} \quad (\ast) \\
\| \rho(i) - \rho(j) \|_2 \leq \Theta(\log^2 n) & \quad \forall i, j \quad (\ast\ast)
\end{align*}
\]

Pf: \((\ast\ast)\) follows from the triangle inequality that

in each word \( |C_{\rho(i)} - C_{\rho(j)}| \leq d(i, j) \)

and \( \Omega(\log n) \) words.
So remains to show that we do get "reasonable" contribution to the distance from this embedding.

Define: $r_e$ = smallest radius so that $|B(i, r)|, |B(j, r)| \geq 2^e$

$r_0 = 0 \leq r_1 \leq r_2 \leq \ldots \leq r_{e-1} < r_e = \frac{r_0}{4} \leq r_f$

Consider some $r_e$ in early part. So that $B(i, r_e) \cap B(j, r_e)$ disjoint.

Say $r_e$ was "defined" by $i$ s.t. $B(i, r_e) \leq 2^b$ but then $B(j, r_e) \geq 2^{e-1}$

Now if we sample at rate $\frac{1}{2^e}$,

\[ Pr[R \text{ selects } B(i, r_e) \text{ and } R \text{ does not } B(j, r_e)] \geq \left(1 - \frac{1}{2^e}\right)^2 \cdot \left(1 - (1 - \frac{1}{2^e})^{2^{e-1}}\right) \geq 2b(1).\]

If $R$ satisfies this, then distance $(i, R) \geq r_e$

distance $(j, R) \leq r_{e-1}$

$\Rightarrow$ coordinate for $R$ gives $|q(i) - q(j)| \geq r_e - r_{e-1}$

"good" event happens for rate $\frac{1}{2^e}$ with constant probability.

$\Rightarrow$ a constant fraction of these coordinates are "good.

$\Rightarrow$ give $\log n \cdot (r_e - r_{e-1})$ contribution.
Summing over all sample scales gives

\[ \Omega(\text{log} n) \left[ (x_1-x_0) + (x_2-x_1) + \ldots + (x_{i-1}-x_{i-2}) + (x_i-x_{i-1}) \right] \]

\[ \geq \Omega(\text{log} n). \]

As claimed, proves Bourgain's Thm.

(Scaled down by \(\Theta(\text{log} n)\))

- Similar argument shows: \( \| \phi(i)-\phi(j) \|_p \) is also a 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

  \( \| \phi(i)-\phi(j) \| \geq \frac{\text{Yij}}{\text{log} n} \)

  only for \( i \neq j \) and \( D_{ij} > 0 \)

  so reduce \( \text{log} n \rightarrow \text{log} k \).

- Integrality gap? The last lemma we saw a gap \( \Omega(\text{log} n) \) for

  uniform spansest 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!