Lecture 6: Sparsest Cut Problem

Previous lectures: Max Coverage & Set Cover. (Algos, Integrality gaps, Hardness)

Max Cut (Algos, Int gaps).

Today: third problem, sparsest cut: see Algos, Int Gaps, Hardness

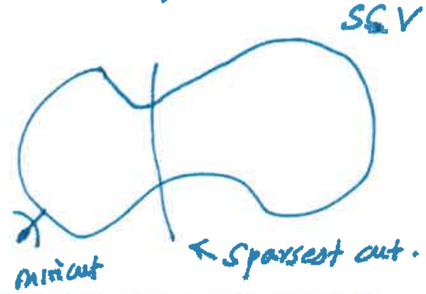
↑ and next 2-3 lectures. LPs Spectral SDPs.

Def: Graph $G=(V,E)$ edges have weights "capacities" $\{c_{uv}\}_{uv \in E}$

Find a bottleneck in the graph " $\min_{S \subseteq V} \dfrac{cap(S,\bar{S})}{\min\{|S|,|\bar{S}|\}}$.

" $\sum_{e \in E} c_e$ e has one end in S and other in $\bar{S}$.

(edge) expansion of cut S.



$$\boxed{\text{sparsity of cut } (S,\bar{S}) \quad \Phi(S) = \frac{cap(S,\bar{S})}{|S| \cdot |\bar{S}|}}$$ ← product of two sides.

Want to find min sparsity cut. $\Phi(G) = \min_S \Phi(S)$.

Fact: sparsity of cut $\in$ edge expansion of cut $\times \left[\frac{2}{n}, \frac{1}{n}\right]$

Pf: for any cut $(S,\bar{S})$ $|S| \cdot |\bar{S}| = |S| \cdot (n - |S|) = |S| \times$ (some factor between $n/2$ and n).

↑ say $|S| \leq |\bar{S}|$ ∎

⇒ sparsity of graph captures edge expansion of graph to within constant factor!

Want to compute the (uniform) sparsest cut $\min_S \Phi(S) = \min_S \frac{cap(S,\bar{S})}{|S|.|\bar{S}|}$.

- [Shahrokhi Matula] NP hard.

So how to approximate it?

[Leighton Rao 88] $O(\log n)$ approximation, $\Omega(\log n)$ int gap via LPs.
[Arora Rao Vazirani 03] $O(\sqrt{\log n})$ apx. via SDPs.
[Naor Young 18] $\Omega(\sqrt{\log n})$ int gap.

Several Interesting ideas here...

Start from the beginning.

- Leighton Rao LP relaxation.

Set edge lengths y_e for each edge $y_e \in [0,1]$. as in max cut case.

"Ideal setting" $y_e \in \{0,1\}$ captures if edge is uncut or cut.

triangle inequality: $y_{ij} \le y_{ik} + y_{kj}$ $\forall k, i, j$.

"if $y_{ij} = 1$ $\Rightarrow y_{ik}$ or $y_{kj} = 1$ in the $\{0,1\}$ case"

Now: objective function is $\min \frac{\sum_{ij \in E} c_{ij} y_{ij}}{\sum_{i,j \in V} y_{ij}}$

subject to y is a metric (so $y_{ij} \ge 0$, $y_{ij} \le y_{ik} + y_{kj}$)

Fact: if $y_{ij} \in \{0,1\}$ then this is the sparsest cut problem.

But now relaxation so $LP_{obj} \le OPT$.

Btw: how to handle ratio objective?

Easy: $\min \frac{a^T x}{b^T x}$ st $x \in K$, $x \ge 0$ cone $\iff$ $\min a^T x$ st $b^T x = 1$, $x \in K$, $x \ge 0$. (at least when $b \ge 0$).

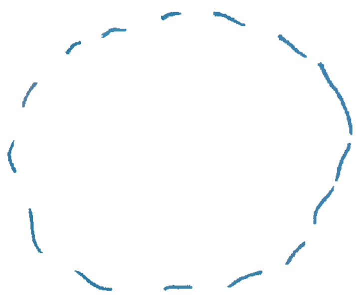
Theorem: Given any solution y to the LP (also called "metric relaxation")
$\exists$ a ~~set~~ cut with sparsity $\leq O(\log n) \cdot$ LP value.

Thm: $\exists$ instances where integrality gap of LP $= \Omega(\log n)$
so. all cuts have sparsity $\geq \Omega(\log n) \cdot$ LP value.

Easiness Proof:

given metric $\{y_{ij}\}_{ij}$ s.t. $\sum_{ij}' y_{ij} = 1$ ($\Rightarrow$ avg distance $= 1/\binom{n}{2}$)

have $LP = \sum' c_{ij} y_{ij}$.



Want to find a good sparsity cut.

Algorithm:

(1) Find a "Low Diameter Decomposition" with radius $R = \frac{1}{4n^2}$ ← smaller than avg distance.
that cuts "few" edges. $\leq O(\log n) \cdot \frac{LP}{R}$

(2) if each piece has $\leq n/3$ nodes then collect pieces into
two parts $L, \bar{L}$ each having $\geq n/3$ nodes greedily.
output L. $\Rightarrow$ sparsity $\leq \dfrac{O(\log n) \cdot \frac{LP\ value}{R}}{(n/3)(2n/3)} = O(\log n) \cdot LP$. ☺

(3) if $\exists$ piece that has $> n/3$ nodes ("dense" piece, all these nodes at dist $\leq 2R$ from each other)
then contract this set of nodes into "root"
Lay out vertices in order of distance from this root (according to y).

Find best "level cut". Claim: cut $\leq O(LP)$!

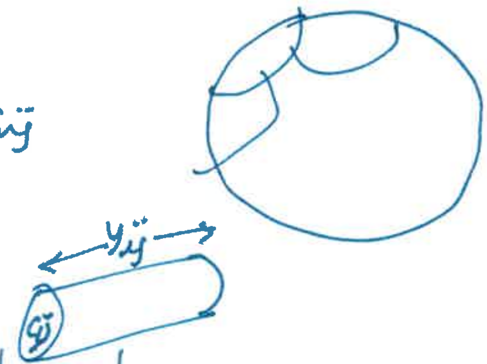
So two things to show:

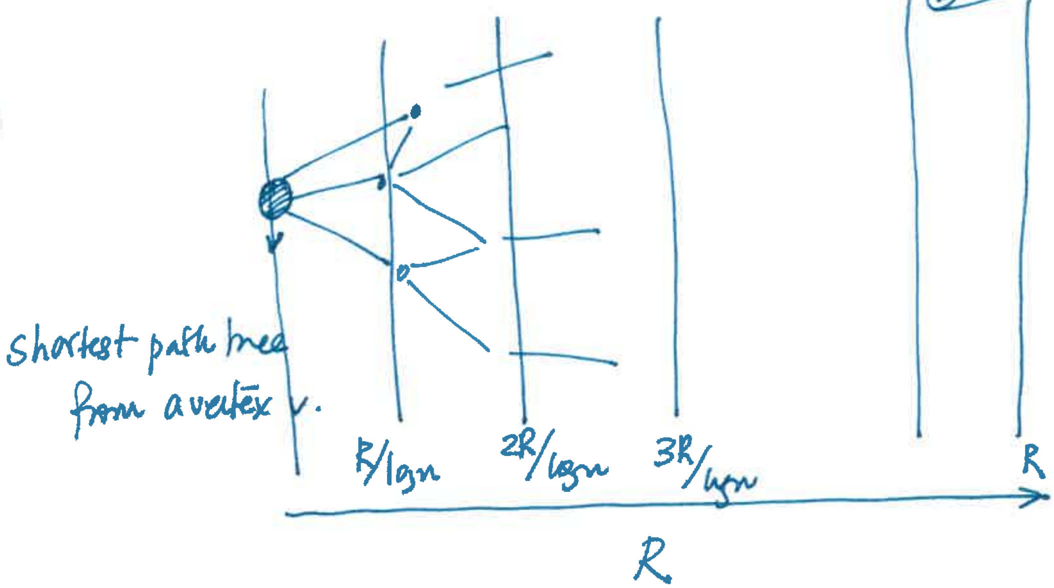
1. that given any metric y, can do this Low Diam Decomp.
2. and that dense case is easy.

<u>Low Diam Decomposition</u> Many ways of doing this
Let's see a "region growing" argument.

Imagine each edge ij has cross section = capacity c_{ij}
length = y_{ij}

$\Rightarrow$ volume = $c_{ij}\, y_{ij}$





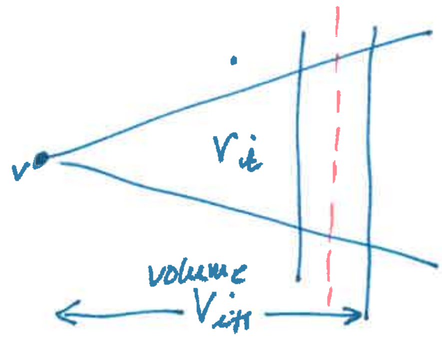
Let V_i = volume in layers between v and $\frac{iR}{\log n}$

Assume $V_0 = LP/n$. (imaginary volume).

Look for layer i s.t. $V_{i+1} \le 2V_i$

<u>Fact</u>: $\exists$ such a layer, else each time volume doubles (and more!)
starts at LP/n, so after $\log n$ levels, must be $> LP$,
impossible!

So $\exists$ layer i s.t. $V_{i+1} \le V_i \times 2$.



~~Find a random cut in~~

Find the cheapest cut in $\left[\frac{iR}{\log n}, \frac{(i+1)R}{\log n}\right]$

no more expensive than random ~~cut~~ level cut

But each edge in that range cut w.p. $\dfrac{\text{length of edge within that range}}{(R/\log n)}$.

$$\Rightarrow E[\text{cut value}] = \sum_{\substack{ij \\ \text{in that} \\ \text{range}}} \frac{c_{ij} \times y_{ij}(\text{in that range})}{(R/\log n)}$$

$$= \frac{V_{i+1} - V_i}{R/\log n}$$

$$\le \frac{V_{i+1}}{R} \log n.$$

$$\le \frac{2V_i}{R} \cdot \log n.$$

V_i = is volume that is definitely cut away.

$\Rightarrow$ we are paying at most $\dfrac{2V_i}{R} \log n$

but cutting away V_i volume. (So we can "charge" to this removed volume).

Eventually when process stops, total edges cut

fake volume added to roots.

$$\le \frac{2.(\text{total volume cut away})}{R} \cdot \log n \le \frac{2(LP + n\frac{LP}{n})}{R} \cdot \log n$$

$= 4LP . \log n / R$.

This proves the Low diam decomp property

cut graph into pieces of radius $\leq R$

and total edges cut $\leq \frac{LP}{R} \cdot O(\log n)$.

Notes: Can give randomized versions too

see [Bartal, Calinescu Karloff Rabani, Miller Peng Xu]

this is a "local" version, "cut and charge to the left, to the volume removed".

Called "region growing", used widely in algorithm design for approx algos, distributed computation, etc.

Originally due to [Awerbuch Peleg], [Linial Saks].

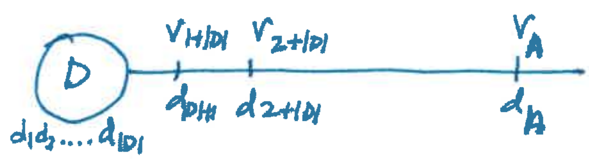
Finally: Handle dense cluster case.

Sps. $\exists$ set $D \subseteq V$ of size $\geq n/3$, such that all distances in D are $\leq 2R$.

"dense"

$\leq \frac{2}{4n^2}$ (say).

then



- Arrange vertices according to distance from D. $d_1 \leq d_2 \leq \ldots \leq d_n$.

$$d_v = \min_{u \in D} y_{uv} = \text{distance of } v \text{ to closest vertex in } D.$$

- Consider sets $S_i = \{v_1, v_2 \ldots v_i\}$.

Output set S_i with least sparsity. $\phi(S_i) = \frac{cap(S_i, \bar{S_i})}{|S_i| \cdot |\bar{S_i}|}$

- Claim: if $|D| \geq n/3$, there is a set S_i with sparsity $\leq O(LP)$.

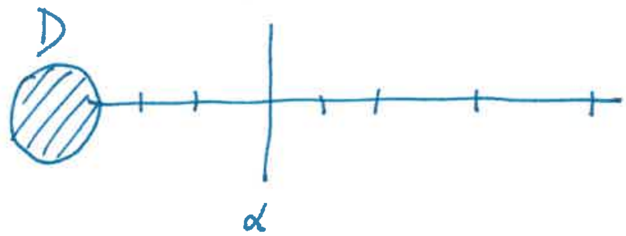
Proof:

First consider $1 = \sum_{ij}' y_{ij} \leq \sum_{ij}' (d_i + d_j + 2R)$ ← Δ-ineq and D has small diam.

$$= n \cdot \sum_i' d_i + n \cdot \sum_j' d_j + \underbrace{n^2 \cdot 2R}_{= \frac{1}{2} \text{ by our choice}}$$

$$\Rightarrow \sum_i' d_i \geq \frac{1}{4n}.$$

Now:



Pick a radius α, $\in [0,1]$, look at ball $B_\alpha = \{$all vertices at distance $\leq \alpha$ from $D\}$.

$\bar{B}_\alpha$ = complement.

$$LP = \int_{\alpha=0}^{1} \text{cap}(B_\alpha, \bar{B}_\alpha)\, d\alpha \geq \int \text{Alg} \cdot |B_\alpha| \cdot |\bar{B}_\alpha| \, d\alpha$$ ← We chose best sparsity set

$$\geq \text{Alg} \cdot \frac{n}{3} \cdot \underbrace{\int |\bar{B}_\alpha| \, d\alpha}_{= \sum_v' d_v}$$

$$\geq \text{Alg} \cdot \frac{n}{3} \cdot \frac{1}{4n} = \frac{\text{Alg}}{12}.$$

$$\Rightarrow \text{Alg} \leq 12\, LP.$$

How good is the LP?

Sadly, integrality gap = $\Omega(\log n)$.

Examples come from the other important class of counterexamples

constant degree expander graphs.

Expander Graphs: $G=(V,E)$ expander if $\forall$ sets S that are "small enough", the neighborhood of S is "large".

Every small set "expands".

let's make it d-regular.

technically we will give graph families

Formally: an α-edge expander G is graph s.t.

$$\forall |S| \le |V|/2, \quad |E(S,\bar{S})| \ge \alpha \cdot d \cdot |S|.$$

Fact/Thm: α-expander graphs exist ~~where~~ with $\alpha =$ constant.

when we say graph is expander without specifying α, usually mean $\alpha = \Theta(1)$ universal constant.

Examples: ~~random~~ take constant # of random matchings.

But even deterministic constructions are known.

- Say vertex set is $\mathbb{F}_p$, p prime

  each i connected to $i+1, i-1, 1/i$

- Cayley graphs of certain groups. [Margulis, Gabber-Galil]

- And zig-zag product (maybe discuss later).

Thm: Integrality gap of metric relaxation $= \Omega(\log n)$.

Pf: take constant-degree expander with $\alpha = \Theta(1)$.
$d = 3$ (say). ↑ expansion.

Set $c_e = 1 \quad \forall e \in E$

Fact: $\forall$ vertex in expander, at most 3^k vertices at hop distance $\leq k$.

$\Rightarrow$ at least $n/2$ vertices at distance $\Omega(\log n)$ from any vertex.

Set $\text{len}(e) = \frac{c}{n^2 \log n}$. And define y_{ij} = shortest path distance b/w i & j according to these edge lengths.

↑ "length of edge"

$\Rightarrow \sum_i y_{ij} \geq 1$ for suitable constant, since $\Omega(n^2)$ vertex pairs at hop-distance $\Omega(\log n)$ from each other.

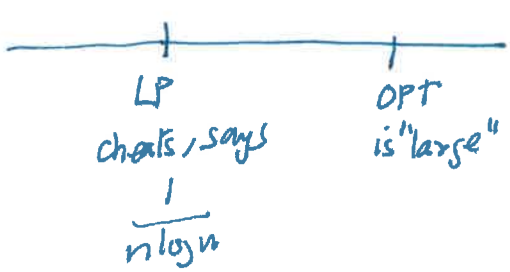
Now: LP value $= \sum_{e \in E} c_e y_e = O\left(\frac{1}{n^2 \log n}\right) \cdot O(n) = O\left(\frac{1}{n \log n}\right)$.

But expansion of all cuts $= \Omega(1)$

$\Rightarrow$ sparsity of all cuts $= \Omega\left(\frac{1}{n}\right)$.

sparsity $\approx$ expansion $\times \Theta\left(\frac{1}{n}\right)$.
from page 1 of these notes

$\Rightarrow$ gap of $\Omega(\log n)$ between LP and OPT.



Next Lecture(s): Make connection to

- metric embeddings and
- eigenvalues/vectors