Lecture 6: Linear Programming for Sparsest Cut

Sparsest Cut and SOS

- The SOS hierarchy captures the algorithms for sparsest cut, but they were discovered directly without thinking about SOS (and this is how we'll present them)
- Why we are covering sparsest cut in detail:
  1. Quite interesting in its own right
  2. Illustrates the kinds of things SOS can capture
  3. Determining if SOS can do better is a major open problem on SOS.

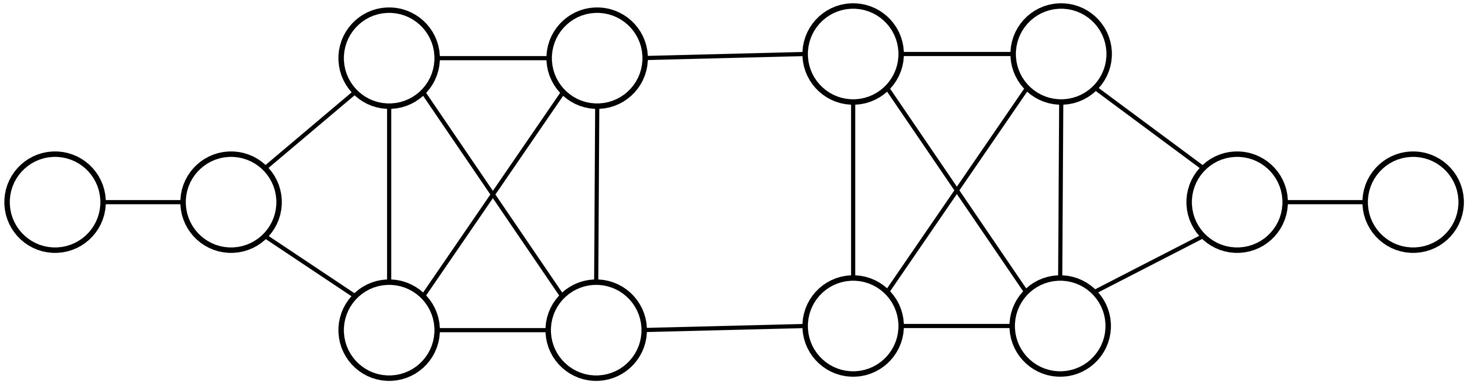
Lecture Outline

- Part I: Sparsest cut
- Part II: Linear programming relaxation and analysis via metric embeddings
- Part III: Bourgain’s Theorem
- Part IV: Tight example: expanders

Part I: Sparsest Cut

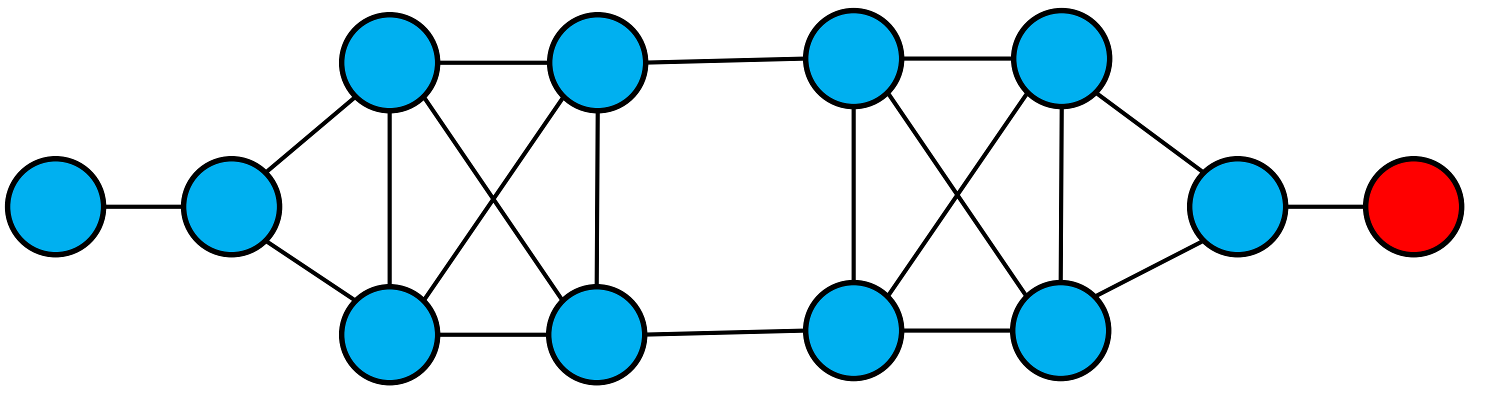
Flaw of Minimum Cut

- We've seen that MIN-CUT can be solved efficiently
- However, MIN-CUT may not be the best way to decompose a graph
- Example:

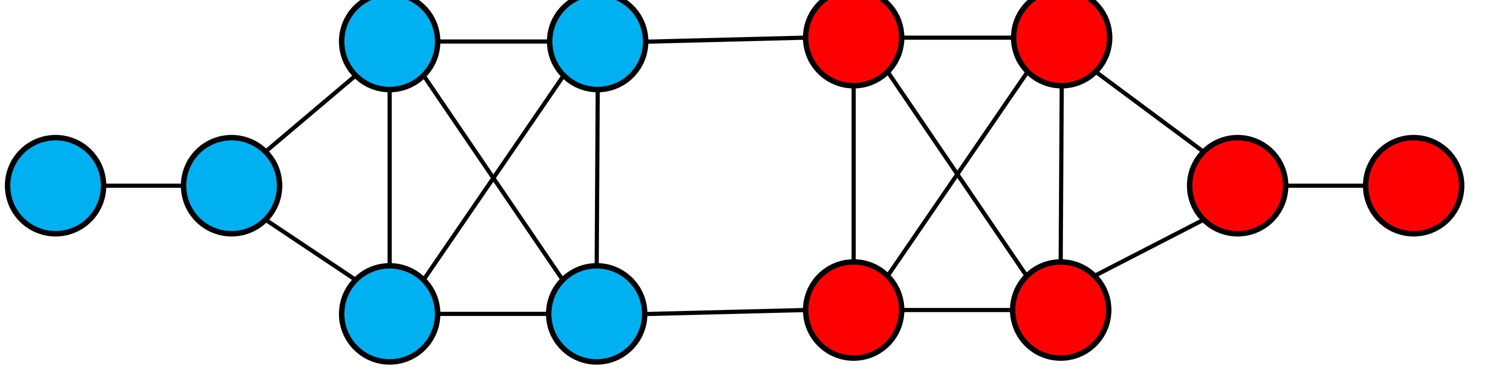


Flaw of Minimum Cut

- MIN-CUT:



- Desired Cut:



Sparsest Cut Problem

- Idea: Divide # of cut edges by # of possible which could have been cut
- Definition: Given a cut $C = (S, \bar{S})$, define

$$\phi(C) = \frac{\#\ of\ edges\ cut}{|S| \cdot |\bar{S}|}$$

- Sparsest cut problem: Minimize $\phi(C)$
- Can also have a weighted version:

$$\phi(C) = \frac{\sum_{i,j:i \in S, j \in \bar{S}, (i,j) \in E(G)} w(i,j)}{\sum_{i,j:i \in S, j \in \bar{S}} w(i,j)}$$

Linear Programming for Sparsest Cut

- Theorem [LR99]: There is a linear programming relaxation for sparsest cut which gives an $O(\log n)$ approximation.

Part II: Linear Programming Relaxation and Analysis via Metric Embeddings

Metric and Pseudo-metric Spaces

- Definition: A metric space (X, d) is a set of points X and a distance function $d: X \times X \to \mathbb{R}_{\geq 0}$ where
  1. $\forall x_1, x_2 \in X, d(x_1, x_2) = d(x_1, x_2)$
  2. $\forall x_1, x_2 \in X, d(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2$
  3. $\forall x_1, x_2, x_3 \in X, \mathrm{d}(\mathrm{x}_1, \mathrm{x}_3) \leq d(x_1, x_2) + d(x_2, x_3)$
- Example 1: Euclidean Space: $d(x, y) = \|y - x\|$
- Example 2: L^1 distance: $d(x, y) = \sum_i |y_i - x_i|$
- Without the second condition, this is called a pseudo-metric space

Cut Spaces

- A cut $C = (S, \bar{S})$ induces a pseudo-metric space on a graph G: Take $d(u, v) = 0$ if $u, v \in S$ or $u, v \in \bar{S}$ and otherwise take $d(u, v) = c$ for some $c > 0$.
- We call this a cut space.

Problem Reformulation

- Reformulation: Minimize $\frac{\sum_{i,j:i<j,(i,j)\in E(G)} d(i,j)}{\sum_{i,j:i<j} d(i,j)}$ over all cut spaces
- First issue: Objective function is nonlinear
- Fix: Set denominator equal to 1.
- Modified Reformulation: Minimize $\sum_{i,j:i<j,(i,j)\in E(G)} d(i,j)$ over all cut spaces normalized so that $\sum_{i,j:i<j} d(i,j) = 1$

Problem Relaxation

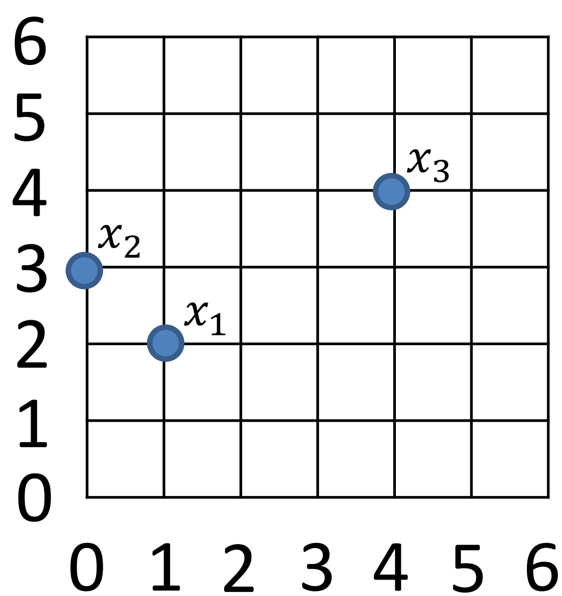
- Want to minimize $\sum_{i,j:i<j,(i,j)\in E(G)} d(i,j)$ over all cut spaces normalized so that $\sum_{i,j:i<j} d(i,j) = 1$
- Relaxation: Minimize $\sum_{i,j:i<j,(i,j)\in E(G)} d(i,j)$ over all pseudo-metrics normalized so that $\sum_{i,j:i<j} d(i,j) = 1$. Linear program constraints:
  1. $\forall i,j, d(i,j) = d(j,i) \geq 0$
  2. $\forall i,j,k, d(i,k) \leq d(i,j) + d(j,k)$
  3. $\sum_{i,j:i<j} d(i,j) = 1$

L^1 Spaces

- Definition: We say that a pseudo-metric (X, d) is an L^1 space if there is a mapping f: $X \rightarrow \mathbb{R}^{\mathrm{n}}$ such that $\forall x, y \in X$,
$$d(x, y) = \sum_i |f(y)_i - f(x)_i|$$
- In this case, we may as well pretend we are already in $\mathbb{R}^n$ with the L^1 distance function
- Lemma: For the sparsest cut relaxation, there is no gap between L^1 spacs and cut spaces!

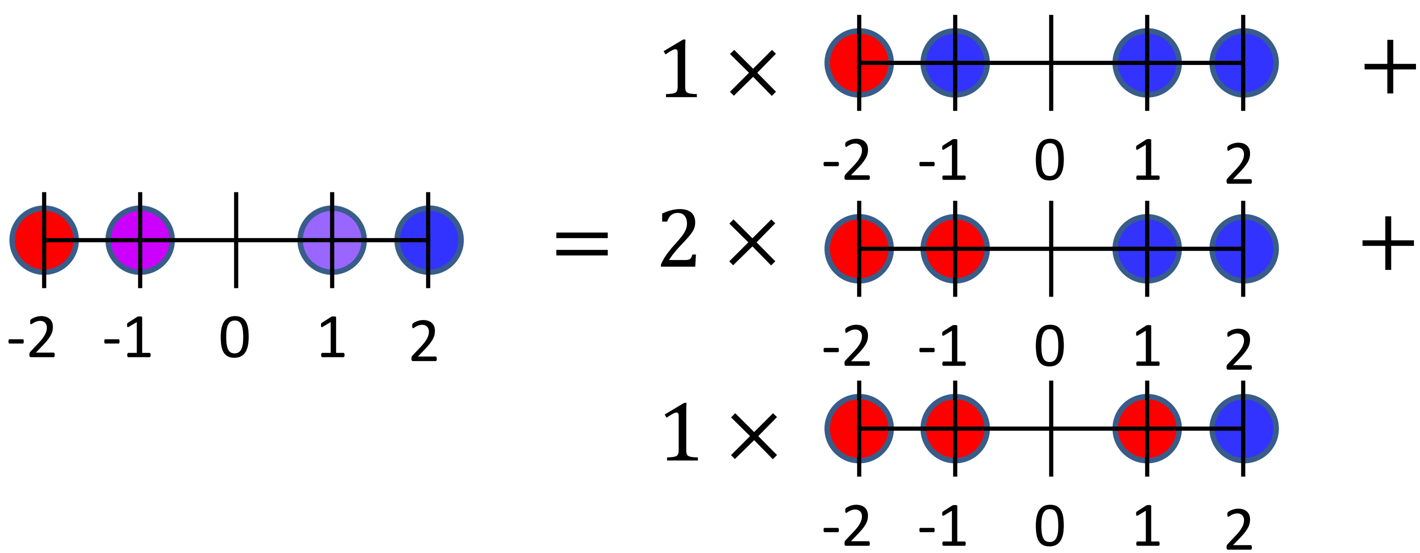
L^1 Space Example

- If $x_1 = (1,2)$, $x_2 = (0,3)$, and $x_3 = (4,4)$, then in the L^1 metric, $d(x_1, x_2) = 2$, $d(x_1, x_3) = 5$, and $d(x_2, x_3) = 5$



Decomposing L^1 Pseudo-metrics

- Lemma: Any finite L^1 space can be decomposed as a linear combination of cut spaces.
- Proof sketch: We can work coordinate by coordinate. For a single coordinate, here is the picture:



Useful Lemma

- Lemma: If $a, b \geq 0$ and $c, d > 0$ then
$$\min\left\{\frac{a}{c}, \frac{b}{d}\right\} \leq \frac{a+b}{c+d} \leq \max\left\{\frac{a}{c}, \frac{b}{d}\right\}$$
- Proof: Without loss of generality, assume that $\frac{a}{c} \leq \frac{b}{d}$. Take $a' = \frac{bc}{d} \geq a$ and take $b' = \frac{da}{c} \leq b$. Now $\frac{a}{c} = \frac{a+b'}{c+d} \leq \frac{a+b}{c+d} \leq \frac{a'+b}{c+d} = \frac{b}{d}$
- Together with the previous decomposition, this shows that for any L^1 space, there's always a cut spacec which is as good or better.

Metric Embeddings and Distortion

- Often want to embed a more complicated metric space into a simpler one. This embedding won't be perfect, but may still be useful
- Given metric spaces $(X, d), (Y, d')$ and a map $f: X \to Y$:
  1. Define the expansion of f to be $\max_{u,v \in X} \frac{d'(f(u),f(v))}{d(u,v)}$
  2. Define the contraction of f to be $\max_{u,v \in X} \frac{d(u,v)}{d'(f(u),f(v))}$
  3. Define the distortion of f to be the product of the expansion and the contraction of f

Metric Embeddings into L^1

- If the pseudo-metric given by our linear program can be embedded into L^1 with distortion α, this gives an α-approximation for the value of the sparsest cut.
- Question: How well can general finite pseudo-metric spaces be embedded into L^1?

Part III: Bourgain’s Theorem

Bourgain's Theorem

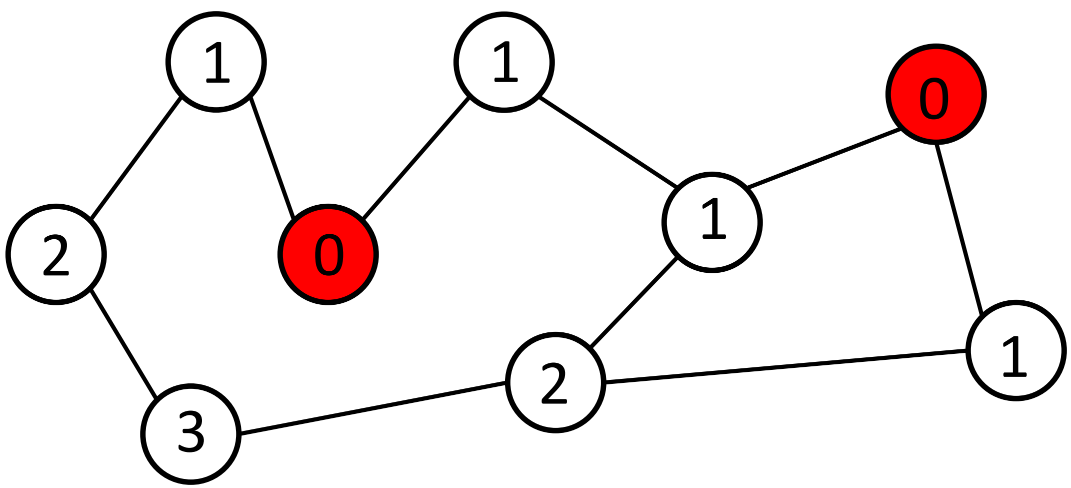
- Theorem [Bou85]: Every metric on n points can be embedded into an L^1 metric with distortion $O(\log n)$. Moreover, $O((logn)^2)$ coordinates are sufficient
- Note: the bound on the number of coordinates is due to Linial, London, and Rabinovich [LLR95]

Fréchet Embeddings

- Def: Given a set of points S, define
$$d(x, S) = \min_{s \in S} d(x, s)$$
- Fréchet embedding: Gives a value to each point based on its distance from some subset S of points and takes the distance between. In other words,
$$d_S(x, y) = |d(y, S) - d(x, S)|$$
- Proposition: For any S, $d_S(x, y) \leq d(x, y)$

Fréchet Embedding Example

- Start with the distance metric $d(u, v) =$ length of the shortest path from u to v on the graph shown. If we take S to be the set of red vertices, we get the values shown for $d(v, S)$.



Fréchet Embeddings Bound

- $d(x, S) = \min_{s \in S} d(x, s)$
- $d_S(x, y) = |d(y, S) - d(x, S)|$
- Proposition: For any S, $d_S(x, y) \leq d(x, y)$
- Proof: Let s be the point in S of minimal distance from x.

  $$d(y, S) \leq d(y, s) \leq d(x, s) + d(x, y) = d(x, y) + d(x, S)$$
- By symmetry, $\mathrm{d}(x, S) \leq d(x, y) + d(y, S)$ so $\mathrm{d_S(x, y)} = |d(y, S) - d(x, S)| \leq d(x, y)$, as needed.

Bourgain's Theorem Proof Idea

- Proof idea: Choose many Fréchet embeddings, have a coordinate for each one.
- Resulting expansion is at most the sum of the weights on the embeddings (this will be $O(logn)$ for us)
- Challenge: Ensure that the contraction is $O(1)$. In other words, ensure that some of the Fréchet embeddings preserve some of the distance between each pair of points x and y.

Bad Case #1

- Issue: Could have that $f_S(x, y) \ll d(x, y)$. In fact, $f_S(x, y)$ can easily be zero!
- Case 1: All points in S are far from x and y and $d(x, S) = d(y, S)$.
- Example:

Bad Case #2

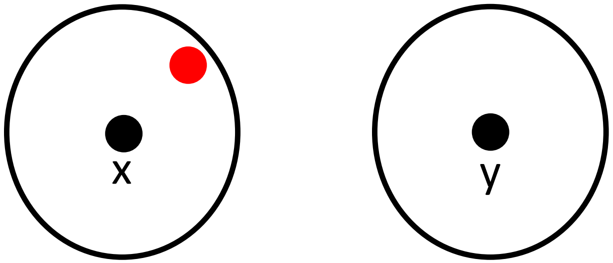
- Case 2: There two points s_x and s_y in S where s_x is very close to x and s_y is very close to y. If so, can have that

$$\mathrm{d}(\mathrm{x}, \mathrm{S}) = d(x, s_x) = d(y, s_y) = d(y, S)$$

- Example:

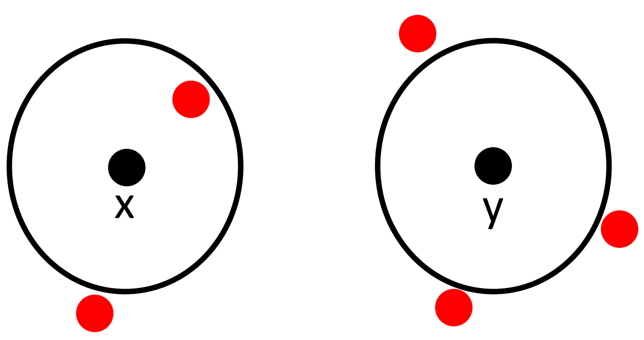
Attempt #1

- Want S to contain exactly one point p which is very close to x or y.
- Let $d = d(x, y)$. Pick S so that S has precisely one point p which is within distance $\frac{d}{3}$ of either x or y.
- Can be accomplished with constant probability by taking a random S of the appropriate size.



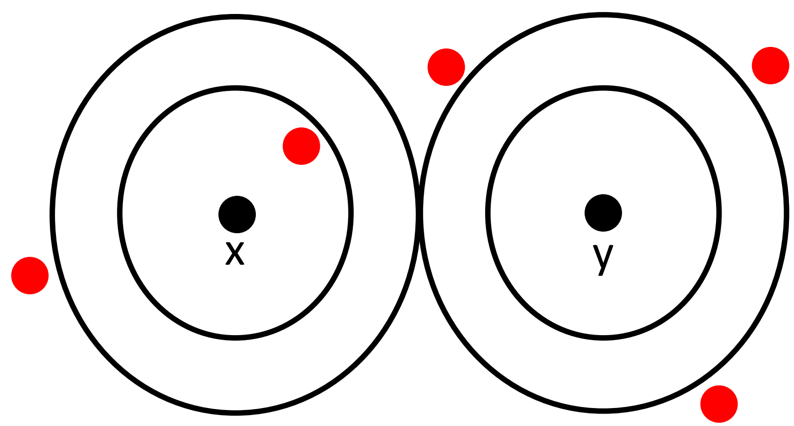
Attempt #1

- Attempt #1: Pick S so that S has precisely one point p which is within distance $\frac{d}{3}$ of either x or y.
- Danger: S also contains point(s) of distance slightly more than $\frac{d}{3}$ from the other point.



Attempt #1

- Possible fix: Require that S contains exactly one point within distance $\frac{d}{3}$ of x or y and no other points within distance $\frac{d}{2}$ of x or y
- This implies $d_S(x, y) \geq \frac{d}{6}$
- However, may be too much to ask for…



Actual Analysis

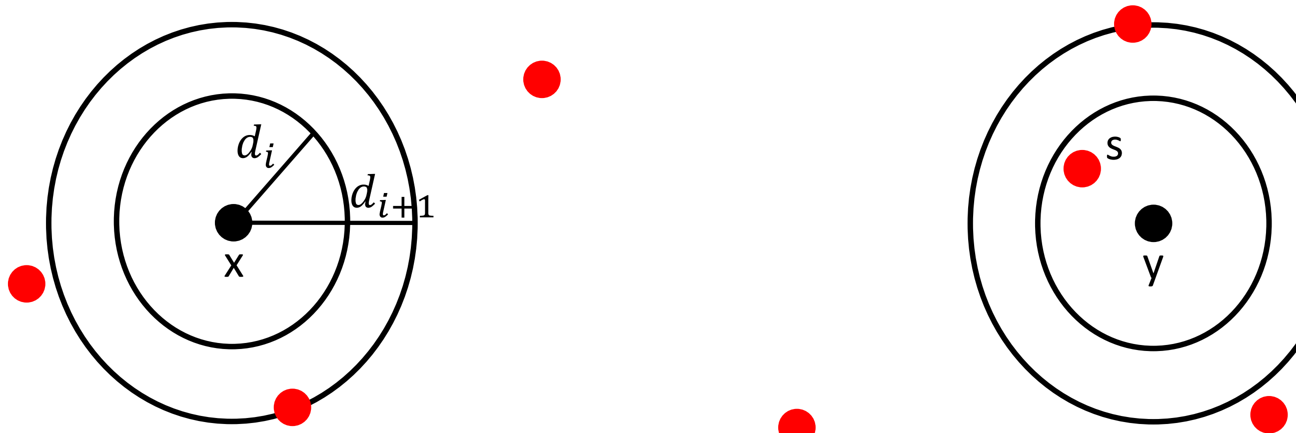
- Def: Given r, p, define $B_r(p) = \{x: d(x,p) \leq r\}$
- For each $i \in [1, \lceil \log_2 n \rceil]$, define d_i to be

$$d_i = \min\left\{ \min\{r : |B_r(x) \cup B_r(y)| \geq 2^i\}, \frac{d}{3} \right\}$$

- Lemma: If S consists of $\left\lceil \frac{n}{2^i} \right\rceil$ points chosen at random then $\mathrm{P}[f_S(x,y) \geq d_{i+1} - d_i]$ is $\Omega(1)$
- Proof: With probability $\Omega(1)$,
  1. $\exists p \in S: p \in B_{d_i}(x) \cup B_{d_i}(y)$
  2. $\nexists p': p' \in S, p' \neq p, min\{d(x,p'), d(y,p')\} < d_{i+1}$

Actual Analysis Picture

- If S consists of $\left\lceil \frac{n}{2^i} \right\rceil$ points chosen at random then with probability $\Omega(1)$:



Actual Analysis Continued

- Lemma: If S consists of $\left\lceil \frac{n}{2^i} \right\rceil$ points chosen at random then with constant probability, $f_S(x,y) \geq d_{i+1} - d_i$
- Corollary: Averaging over all $i \in [1, \lceil logn \rceil]$, the expected value of $f_S(x,y)$ is at least $\Omega\left(\frac{d}{logn}\right)$
- For each $i \in [0, \lceil logn \rceil]$, take $O(logn)$ S of size 2^i at random. This ensures that everything is close to its expectation with high probability.

Actual Analysis Continued

- Full embedding procedure: For each $i \in [0, \lceil logn \rceil - 1]$, take $\mathrm{m} = O(logn)$ S of size 2^i at random. For each such S, create a coordinate where each point x has value $\frac{1}{m} d(x, S)$.
- Averaging over many subsets of each size ensures that everything is close to its expectation with high probability.

Part IV: Tight Example: Expanders

Expander Graphs

- A vertex/edge expander is a graph G where every subset of G has a lot of neighbors/outgoing edges
- Definition: The vertex expansion of a graph G is $\min_{S:0<|S|\leq\frac{n}{2}} \frac{|N(S)|}{|S|}$ where
$$N(S) = \{v: \exists u \in S: (u,v) \in E(G)\}$$
- Definition: The edge expansion of a graph G is $\min_{S:0<|S|\leq\frac{n}{2}} \frac{|\delta(S)|}{|S|}$ where
$$\delta(S) = \{(u,v): u \in S, v \notin S, (u,v) \in E(G)\}$$

Observations on Expander Graphs

- Expander graphs are extremely useful in complexity theory.
- Derandomization: random walks mix well
- Here: Edge expanders have no sparse cuts.
- Proposition: If G has edge expansion c then for all cuts $\mathrm{C} = (S, \bar{S})$, $\phi(C) = \frac{\#\, of\ edges\, cut}{|S| \cdot |\bar{S}|} \geq \frac{c}{n}$
- Proof: By definition, $\#\, of\ edges\, cut \geq c|S|$ and $|\bar{S}| \leq n$

Constructing Expanders

- With high probability, random graphs are excellent expanders.
- Constructing expanders explicitly is more challenging and is an entire field of research on its own.

$\Omega(\log n)$ gap with expanders

- Use the distance metric d_{ij} = smallest length of a path from i to j.
- For a d-regular expander with edge expansion $\frac{d}{4}$:
  1. $\sum_{i,j:i<j,(i,j)\in E(G)} d_{ij} = |E(G)|$ which is $O(nd)$
  2. $\sum_{i,j:i<j} d_{ij}$ is $\Omega(n^2 \log(n))$ as most pairs of vertices are logarithmic distance apart
- Linear programming relaxation value: $O\left(\frac{d}{nlogn}\right)$
- Actual value is $\Omega\left(\frac{d}{n}\right)$

References

- [Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. Israel J. Math., 52(1–2), p. 46–52. 1985.
- [LR99] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. Journal of the ACM (JACM) 46(6), p. 787–832. 1999
- [LLR95] N. Linial, E. London, Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. Combinatorica 15(2),p. 215–245. 1995