EECS 495: Randomized Algorithms	Lecture 11
Metric Embeddings	

Reading: Text: Vazirani, Chapter 21

Started with min-cut, saw max-cut last lecture, now for another important cut problem, sparsest cut. We use this problem to demonstrate embedding techniques in randomized approximation algorithms.

Multi-commodity flow/sparest cut

Problem: Multi-commodity flow.

Given

- graph $G = (V, E)$
- capacities $c : E \rightarrow \Re^+$
- commodities $\{1, \ldots, k\}$ with
  - demands $\text{dem}(i)$
  - source/sink pairs (s_i, t_i)

Find

- max throughput f^* s.t. for each commodity, $f^* \times \text{dem}(i)$ flow can be routed simultaneously

Note: Given cut $(S, \bar{S})$, let

- $c(S)$ be total capacity of $e \in \delta(S)$
- $\text{dem}(S)$ be total demand separated by cut

Fact: $f^* \leq \min_{S \subseteq V} \frac{c(S)}{\text{dem}(S)}$ (one direction of min-cut/max-flow thm).

Def: The *sparsity* of a cut $(S, \bar{S})$ is $\frac{c(S)}{\text{dem}(S)}$.

Problem: Sparest cut. Given above setup, find cut S with minimum sparsity.

Note: Unlike single-commodity, bound is *not* tight!

Example: $K_{3,2}$, unit capacity, unit demand between each pair of non-adjacent vertices.

- sparsest cut: remove a node on lhs, sparsity is 1
- flow: can only route all lhs demands by sending 1/2 on each edge, but this saturates edges, so rhs can't route demand.

LP Formulation

Primal

Let $\mathcal{P}_i = \{q_j^i\}$ be paths from s_i to t_i.

$$\begin{array}{ll} \max & f \\ s.t. & \sum_j f_j^i \geq f \cdot \text{dem}(i), \ 1 \leq i \leq k \\ & \sum_{q_j^i : e \in q_j^i} f_j^i \leq c_e, \ e \in E \end{array}$$

where f_j^i is flow from commodity i on path q_j^i.

Dual

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e d_e \\ s.t. \quad & \sum_{e \in q^i_j} d_e \geq l_i \\ & \sum_{i=1}^{k} l_i \cdot \mathrm{dem}(i) \geq 1 \end{aligned}$$

where d_e is "distance" of edge e.

[[*Don't actually need exponential LP, and/or can use separation oracle that looks at shortest path from s_i to t_i.*]]

Fact: There is an optimal soln. in which the d_e form a *metric* and second ineq. is tight.

Def: A *metric space* is a pair (V, d) where $d : V \times V \rightarrow \Re^+$ s.t.

- $d(x, y) = 0 \leftrightarrow x = y$
- symmetry: $d(x, y) = d(y, x)$
- triangle inequality: $d(x, y) + d(y, z) \geq d(x, z)$

Proof: If not, can massage it to one:

- if for some u, v, w, $d_{uv} + d_{vw} < d_{uw}$, can set $d_{uw} = d_{uv} + d_{vw}$ without changing shortest path length, so still feasible
- set $l_i = d(s_i, t_i)$ where $d(\cdot)$ is dist. according to metric d_e
- if second ineq. isn't tight, scale all d_e to make it tight

Hence optimal throughput f^* equals

$$\min_{\text{metric } d} \frac{\sum_e c_e d_e}{\sum_i \mathrm{dem}(i) d(s_i, t_i)}.$$

Idea: Cuts are metrics too!

[[*To find sparest cut, write optimal metric in terms of cut metrics and hope it suggests a good one.*]]

Cut packings

Def: A *cut-packing* is a function $y : 2^V \rightarrow \Re^+$ assigning value y_S to cut $(S, \bar{S})$ s.t. for each edge e, the amount of cut e "feels" is at most d_e:

$$\sum_{S : e \in \delta(S)} y_S \leq d_e$$

- *exact* if $\sum_{S:e \in \delta(S)} y_S = d_e$
- *β-approximate* for $\beta \geq 1$ if $\frac{d_e}{\beta} \geq \sum_{S:e \in \delta(S)} y_S \leq d_e$

Claim: Given a β-approx. cut packing for (V, d), let S^* be the sparsest cut among those with $y_S > 0$. Then sparsity of this cut is at most $\beta \cdot f^*$.

[[*Hence if we can find β-approximate cut packings we can get β-approximations to sparsest cut.*]]

Proof: Let y be a β-approx. cut packing. Then

$$\begin{aligned} f^* &= \frac{\sum_e c_e d_e}{\sum_i \mathrm{dem}(i) d(s_i, t_i)} \\ &\geq \frac{\sum_e c_e \sum_{S:e \in \delta(S)} y_S}{\sum_i \mathrm{dem}(i) \sum_{S:(s_i,t_i) \in \delta(S)} \beta y_S} \\ &= \frac{\sum_S c(S) y_S}{\beta \sum_S \mathrm{dem}(S) y_S} \\ &\geq \frac{1}{\beta} \left(\frac{c(S^*)}{\mathrm{dem}(S^*)} \right) \end{aligned}$$

(recall ratio of sums at least min ratio).

[[*Exact cut packings means all cuts with non-zero weight are sparest cuts.*]]

Question: How small can β be in general?

Metric Embeddings

Idea: To find good cut packing, embed arbitrary metric into "simple" metric without too much distortion, then cut-pack simple metric.

l_1 metrics and their packing

Def: A *norm* on a vector space $\Re^m$ is a function $||\cdot|| : \Re^m \to \Re^+$ s.t. for any $x, y \in \Re^m$ and $\lambda \in \Re$,

- $||x|| = 0 \leftrightarrow x = 0$
- $||\lambda x|| = |\lambda| \cdot ||x||$
- $||x + y|| \leq ||x|| + ||y||$

Def: For $p \geq 1$, the l_p-*norm* is

$$||x||_p = \left(\sum_{k=1}^{m} |x_k|^p\right)^{1/p}$$

Def: The associated metric is defined by $d_{l_p}(x, y) = ||x - y||_p$.

[[*l_2 is what you think of as Euclidian distance; l_1 is grid distance in 2-d. Here we only care about l_1.*]]

Def: A *β-distortion l_1-embedding* for metric (V, d) is a mapping $\sigma : V \to \Re^m$ for some m such that for all $u, v \in V$,

$$\frac{1}{\beta} d(u, v) \leq ||\sigma(u) - \sigma(v)||_1 \leq d(u, v).$$

Claim: Given β-distortion l_1 embedding σ with dimension m, can construct β-approximate cut packing y with at most mn non-zero y_S.

Proof: If $m = 1$, suppose vertices mapped to $u_1 \leq \ldots \leq u_n$. Define

$$y_{1,\ldots,t} = u_{t+1} - u_t.$$

Then

$$\sum_{S:(i,j)\in\delta(S)} y_S = \sum_{t=i}^{j-1} y_{1,\ldots,t} = u_t - u_s.$$

If $m > 1$, do this in each dimension, get mn cuts, l_1 norm additive, so works.

[[*Other direction true too.*]]

Embedding into l_1

Idea: Use distances to sets:

- Pick $S_1, \ldots, S_m$
- Define $\sigma_i(u) = \min_{t \in S_i} d(u, t)/m$
- No stretch since for dim. i and edge (u, v),

$$d(u, t_u) \leq d(u, t_v) \leq d(v, t_v) + d(u, v).$$

Question: How to choose S_i for to not overshrink?

Idea: Choose randomly!

Algorithm: For $1 \leq i \leq \log n$, let S_i include each $t \in V$ with prob. $1/2^i$.

Claim: $O(\log n)$-distortion embedding.

Let $B(t, r)$ be ball of radius r around t. Bound $d(u, v)$:

- Let $r_1 \geq r_2 \geq 0$
- Suppose $S_i \cap B(u, r_1) = \emptyset$ and $(S_i \cap B(v, r_2) \neq \emptyset)$
- Then $|\sigma_i(u) - \sigma_i(v)| \geq (r_1 - r_2)/l$

Claim: Let A and B be disjoint subsets s.t. $|A| < 2^i$ and $|B| \geq 2^{i-1}$. Then $\Pr[(S_i \cap A = \emptyset) \wedge (S_i \cap B \neq \emptyset)] \geq c$ for some constant c.

Proof: Sets disjoint so independent, easy to calculate.

Let ρ_i be min radius s.t. both $B(u, \rho_i)$ and $B(v, \rho_i)$ have at least 2^i points.

- Suppose ρ_i limited by u
- Suppose $\rho_i < d(u,v)/2$
- $A = B^o(u, \rho_i)$, $B = B(v, \rho_{i-1})$
- Then S_i contributes $c(\rho_i - \rho_{i-1})/l$ to i'th coordinate in expectation
- Summing over coordinates up to $\rho_i = d(u,v)/2$ gives result for (u,v)

Get for all edges whp by running $\log n$ times and using Chernoff, hence dimension of embedding is $O(\log^2 n)$.

[[*Really only wanted to make sure the k demand pairs not overshrunk. Can save by only selecting subsets of vertices that are source/sink pairs and re-running above analysis to conclude demand pairs not overshrunk. Gives a* $(\log k)$*-approx.*]]

[[*Also gives approx. multi-commodity max-flow/min-cut, i.e., sparsity at most* $O(\log k)$ *times throughput.*]]

[[*Improved to* $O(\sqrt{\log n})$ *with SDPs by Arora, Rao, Vazirani in 2004 and via a faster algorithm using "expander flows" by Arora, Hazan, and Kale.*]]