Lecturer: Robert Krauthgamer Scribe by: Shiri Chechik Updated: April 12, 2009

1 Sparsest Cut

1.1 Approximation Algorithm for Sparsest Cut

Stage 0: Solve the relaxation of the sparsest cut problem. As it a linear program, it can be solved in polynomial time.
Stage 1: Embed the metric d found by the LP into l_1 with distortion $O(\log n)$. This also can be done in polynomial time by Bourgains theorem. So exists $f : V \to l_1$ such that for every $u, v \in V$

$$d(u,v) \leq ||f(u) - f(v)|| \leq O(\log n) d(u,v).$$

Therefore, $\sum\limits_{(u,v)\in E} \|f(u) - f(v)\|_1 \leq O(\log n)LP$ and $\sum_i^k \|f(s_i) - f(t_i)\|_1 = 1$ where LP is the solution found by the linear programming relaxation.

Lemma 1 *Every n-point metric $\tilde{d}$ that embeds isometrically into l_1 can be written as a positive combination of cut metrics τ_i. I.e., there exists $\alpha_i > 0$ such that $\tilde{d}(x,y) = \sum \alpha_i \tau_i(x,y)$ for every $x, y \in V$. Furthermore, such α_i can be found in polynomial time and the number of $\alpha_i > 0$ is at most $\binom{n}{2}$.*

We now use Lemma 1 for the second stage of the approximation algorithm.
Stage 2: Write the distance $\tilde{d}$ from the embedding as $\tilde{d} = \sum \alpha_i \tau_i(x,y)$ for τ_i cut metrics.
We now show that at least one of the cut metrics τ_i yields the desired approximation.

Claim 2 *There exists j^* such that the objective $OBJ|_{d=\tau_{j^*}} \leq OBJ|_{d=\tilde{d}}$, i.e., $\frac{\sum \tau_{j^*}(u,v)}{\tau_{j^*}(s_i,t_i)} \leq \frac{\sum \tilde{d}(u,v)}{\tilde{d}(s_i,t_i)}$.*

Proof The proof is a generalization of the following. $\forall a_1, \ldots, a_m, b_1, \ldots, b_m > 0$ $\min\limits_l \frac{a_l}{b_l} \leq \frac{\sum a_l}{\sum b_l}$. Assume, towards contradiction, that $\min\limits_l \frac{a_l}{b_l} > \frac{\sum a_l}{\sum b_l}$. We get that $b_1 \frac{\sum a_l}{\sum b_l} < a_1$. Similarly $b_i \frac{\sum a_l}{\sum b_l} < a_i$ for every $1 \leq i \leq m$. Therefore, $\sum\limits_l a_l = \sum\limits_i b_i \frac{\sum a_l}{\sum b_l} < \sum\limits_l a_l$, a contradiction.
■

Finally, τ_{j^*} gives us $S_G^* \subseteq V$ whose value $= OBJ|_{d=\tau_{j^*}} \leq OBJ|_{d=d^*} \leq O(\log n)LP \leq O(\log n)OPT$.

Refinement: We can improve the approximation ratio to $O(\log k)$ by having one side of distortion guarantee only for demand pairs, i.e., $d(u,v) \leq \|f(u) - f(v)\|$ only for the k demand pairs and $\|f(u) - f(v)\| \leq Dd(u,v)$ for all pairs.

Theorem 3 (Aumann-Rabani, Linial-London-Rabinovich95) *Sparsest cut can be approximated in poly-time within factor $O(\log k)$.*

2 Minimum Bisection

The input of the *Minimum Bisection* problem is a graph $G = (V, E)$ such that $|V| = n$. The goal is to find a cut $(S, \bar{S})$ such that $|S| = |\bar{S}| = n/2$ so as to minimize $e(S, \bar{S})$. This problem is known to be NP-hard.

Recall that the sparsest-cut problem with uniform demands is the search of a cut S that minimizes $\frac{e(S,\bar{S})}{|S||\bar{S}|}$. Note that $\frac{e(S,\bar{S})}{|S|\cdot|\bar{S}|} \cong \frac{e(S,\bar{S})}{min\{|S|,|\bar{S}|\}\cdot n}$, up to a factor of 2.

We now show a poly-time algorithm that finds a $\frac{2}{3}$-balanced cut S of cost $e(S, \bar{S}) \leq O(\log n) \cdot b^*_G$ where b^*_G is the optimal cost of the minimum bisection problem on the graph G.

Algorithm $\boldsymbol{2/3 - balanced - cut(G = (V, E))}$

1. Set $G_{alg} \leftarrow (V, E)$, denote by V_{alg} the set of vertices of the graph G_{alg}
2. While $|V_{alg}| \leq \frac{2n}{3}$
   - use $O(\log n)$ approximation algorithm for the sparsest-cut with uniform demands problem on G_{alg} to find a cut $(S, \bar{S})$, where $|S| \leq |\bar{S}|$.
   - remove S from G_{alg}.
3. return V_{alg}, the vertices of G_{alg}.

Claim 4 *The set of vertices returned by the algorithm V_{alg} satisfies $\frac{n}{3} \leq |V_{alg}| \leq \frac{2n}{3}$.*

Proof At the beginning of the last iteration $|V_{alg}| \geq \frac{2n}{3}$ and we remove at most half the vertices from V_{alg} (since we remove the smaller side of the cut). ■

Denote by b^*_G the optimal cost of the minimum bisection problem and by S^*_G the optimal cut that achieves the cost b^*_G where $|S^*_G| = n/2$.

Claim 5 *The cost of V_{alg} is at most $O(\log n)b^*_G$.*

Proof Let S_ℓ be the set removed in iteration ℓ. Let S^*_ℓ be the set that best minimizes $\frac{e(S'_\ell,\bar{S}'_\ell)}{|S'_\ell|}$ in iteration ℓ. Where $\bar{S}'_\ell$ is the complement of the cut S'_ℓ in the graph of iteration ℓ. The set V_{alg} in iteration ℓ, denoted by V^ℓ_{alg}, contains at least $\frac{2n}{3}$ nodes, therefore $S^*_G \cap V^\ell_{alg} > \frac{n}{2} - \frac{n}{3} = \frac{n}{6}$ and also $\bar{S}^*_G \cap V^\ell_{alg} > \frac{n}{6}$. We get that $\frac{e(S^*_\ell,\bar{S}^*_\ell)}{|S^*_\ell|} \leq \frac{b^*_G}{n/6}$ and as we use a $\log n$-approximation algorithm we now get $\frac{e(S_\ell,\bar{S}_\ell)}{|S_\ell|} \leq O(\log n)\frac{e(S_{\ell*},\bar{S}^*_\ell)}{|S^*_\ell|} \leq O(\log n)\frac{b^*_G}{n/6}$. Hence $e(V_{alg}, \bar{V}_{alg}) \leq \sum_\ell e(S_\ell, \bar{S}_\ell) \leq O(\log n)\frac{b^*_G}{n/6} \cdot \sum_\ell |S_\ell| \leq O(b^*_G \log n)$.

Theorem 6 (Leighton-Rao 88) *There is a poly-time algorithm that finds a $\frac{2}{3}$-balanced cut S of cost $e(S, \bar{S}) \leq O(\log n) \cdot b^*_G$ where b^*_G is the optimal cost of the minimum bisection problem on the graph G.*

3 Distortion Lower Bounds

We now show a specific n-point space such that embedding this space to l_2 requires distortion of at least $\sqrt{\log n}$.

Lemma 7 (Short diagonals) *Let x_1, x_2, x_3, x_4 be points in l_2. Then $\|x_1 - x_3\|^2 + \|x_2 - x_4\|^2 \leq \|x_1 - x_2\|^2 + \|x_2 - x_3\|^2 + \|x_3 - x_4\|^2 + \|x_4 - x_1\|^2$.*

Proof Observe that is suffices to prove it for $x_1, x_2, x_3, x_4 \in R$. For points x_i in some R^d, simply apply the inequality on each coordinate and then add these inequalities together. So consider $x_1, x_2, x_3, x_4 \in R$, $\|x_1 - x_2\|^2 + \|x_2 - x_3\|^2 + \|x_3 - x_4\|^2 + \|x_4 - x_1\|^2 - \|x_1 - x_3\|^2 - \|x_2 - x_4\|^2 = |x_1 - x_2 + x_3 - x_4|^2 \geq 0$. ∎

Theorem 8 (Enflo69) *Let $G = (V, E)$ be the discrete cube $\{0,1\}^m$ and shortest-path distance $d_G(x, y) = \#(\text{bits } i \text{ such that } x_i \neq y_i)$. Then embedding d_G into l_2 requires distortion $\geq \sqrt{m} = \sqrt{\log |V|}$.*

Remark: The above is optimal. The identity mapping: $x \to x$ has distortion $\sqrt{m}$.
Proof Consider $V = \{0,1\}^m$. For $x \in V$, let $\bar{x} \in \{0,1\}^m$ be the complement of x. We will show that for every $f : V \to l_2$:

$$E_{x \in V}[\|f(x) - f(\bar{x})\|^2] \leq m \cdot E_{(x,y) \in E}[\|f(x) - f(y)\|^2]. \tag{1}$$

This would be enough to prove the lemma. By 1,

$$E_{x \in V}[\|f(x) - f(\bar{x})\|^2] \leq m \cdot E_{(x,y) \in E}[\|f(x) - f(y)\|^2] \leq m \cdot E_{(x,y) \in E}[d_G(x,y)^2] \leq m \cdot 1 = m.$$

So there exists a point x such that $\|f(x) - f(\bar{x})\| \leq \sqrt{m} = \frac{d_G(x,\bar{x})}{\sqrt{m}}$.

We now prove equation 1. We prove it by induction. For $m = 2$, use the short diagonal Lemma (divided by 2). Assume the claim holds for $m' < m$ and consider $m' = m$. Let x be a point in $\{0,1\}^{m-1}$. Apply the short diagonals lemma to $x0, x1, \bar{x}0, \bar{x}1$. We get $\|f(x0) - f(\bar{x}1)\|^2 + \|f(x1) - f(\bar{x}0)\|^2 \leq \|f(x0) - f(x1)\|^2 + \|f(x1) - f(\bar{x}1)\|^2 + \|f(\bar{x}1) - f(\bar{x}0)\|^2 + \|f(\bar{x}0) - f(x0)\|^2$. Finally, by summing on all $x's$ and using the induction hypothesis we get the desired inequality.

∎