# 1 Sparsest Cut Problem and LP Formulation

Recall the Sparsest Cut problem is defined as follows. Given an undirected graph $G = (V, E)$ with an edge-cost function $c : E \to \mathbb{R}^+$, and $k$ pairs of vertices $(s_1, t_1), \ldots, (s_k, t_k)$, each of which is associated with a demand value $\text{dem}(i) > 0$ for $i = 1, \ldots, k$, for a set of edges (a cut) $E' \subseteq E$, let

$$\text{dem}(E') = \sum_{i:\ s_i, t_i \text{ separated by } E'} \text{dem}(i), \qquad c(E') = \sum_{e \in E'} c_e,$$

and

$$\text{sparsity}(E') = \frac{c(E')}{\text{dem}(E')}.$$

The goal is to find $E' \subseteq E$ to minimize $\text{sparsity}(E')$.

Consider the following formulation of the Sparsest Cut Problem.

$$\text{(SC-LP)} \quad \min \sum_{e \in E} c_e x_e \tag{1}$$

$$\text{subject to} \quad \sum_{i=1}^{k} \text{dem}(i) y_i \geq 1, \tag{2}$$

$$\sum_{e \in p} x_e \geq y_i, \quad \text{for } p \in P_{s_i t_i},\ i = 1, \ldots, k \tag{3}$$

$$x, y \geq 0, \tag{4}$$

where $P_{s_i t_i}$ is the set of all $s_i$-$t_i$ paths. (SC-LP) can be interpreted as a normalized version of the Sparsest Cut LP: in the original LP, we actually want to minimize

$$\text{sparsity} = \frac{\sum_{e \in E} c_e x_e}{\sum_{i=1}^{k} \text{dem}(i) y_i} \quad (x_e, y_i \in \{0, 1\});$$

and in (SC-LP), the denominator is normalized to be 1, as in constraint (2).

## 1.1 Approximation Algorithms via Rounding LP

We will introduce two ways to rounding (SC-LP) to derive approximation algorithms for the Sparsest Cut problem.

The first one uses a relatively simple reduction to the MULTICUT problem, but illustrates the relationship between the two cut problems and a general technique. The approximation ratio $O(\log k \log D)$ (or $O(\log^2 k)$) is not optimal, where $D = \sum_{i=1}^{k} \mathrm{dem}(i)$.

The second one uses a sophisticated connection to embedding metric spaces into real normed spaces, and leads to the optimum approximation ratio $O(\log k)$.

# 2 Rounding Sparsest-Cut via Multicut Relationship

## 2.1 Minimum Multicut Problem

Recall the MINIMUM MULTICUT problem. Given a graph $G = (V, E)$ with an edge-cost function $c : E \to \mathbb{R}^+$, and $k$ pairs of vertices $(s_1, t_1), \ldots, (s_k, t_k)$, the goal is to find a set of edges $E' \subseteq E$ that separate all pairs and minimize $c(E')$. Consider its LP formulation.

$$\text{(MC-LP)} \quad \min \sum_{e \in E} c_e x'_e \tag{5}$$

$$\text{subject to} \quad \sum_{e \in p} x'_e \geq 1, \quad \text{for } p \in P_{s_i t_i},\ i = 1, \ldots, k \tag{6}$$

$$x' \geq 0. \tag{7}$$

In Lecture 18, we have the following result.

**Lemma 2.1** *For any feasible solution $x'$ to (MC-LP), we can find a multicut $E' \subseteq E$ s.t. $c(E') \leq O(\log k) \sum_{e \in E} c_e x'_e$.*

## 2.2 Rounding Algorithm

Suppose $(x, y)$ is a feasible solution to (SC-LP), we will show how to round it via the multicut relationship to obtain an approximation algorithm for the SPARSEST CUT problem.

Let $y_{\max} = \max_{1 \leq i \leq k} y_i$. Let

$$A_j = \left\{ i \mid y_i \in \left( \frac{y_{\max}}{2^{j+1}}, \frac{y_{\max}}{2^j} \right] \right\},$$

and $\mathrm{dem}(A_j) = \sum_{i \in A_j} \mathrm{dem}(i)$, for $j = 0, 1, 2, \ldots$. We first prove the following lemma.

**Lemma 2.2** *Let $D = \sum_{i=1}^{k} \mathrm{dem}(i)$. There exists $h$ s.t.*

$$\mathrm{dem}(A_h) \frac{y_{\max}}{2^{h+1}} \geq \frac{1}{8 \log D}.$$

**Proof:** Note $A_0, A_1, A_2, \ldots$ are disjoint. So $\sum_{i=1}^{k} \mathrm{dem}(i) y_i = \sum_{j \geq 0} \sum_{i \in A_j} \mathrm{dem}(i) y_i$. From the definition of $A_j$, for any $i \in A_j$, we have $y_i/2 \leq y_{\max}/2^{j+1}$. Therefore,

$$\sum_{j \geq 0} \mathrm{dem}(A_j) \frac{y_{\max}}{2^{j+1}} \geq \sum_{j \geq 0} \sum_{i \in A_j} \mathrm{dem}(i) \frac{y_i}{2} = \frac{1}{2} \sum_{i=1}^{k} \mathrm{dem}(i) y_i \geq \frac{1}{2}. \tag{8}$$

Note the last inequality above is from (2) in (SC-LP). Let $t = 2\log D - 1$. We split (8) into two parts.

$$\sum_{j\geq 0} \mathrm{dem}(A_j)\frac{y_{\max}}{2^{j+1}} = \sum_{0\leq j\leq t} \mathrm{dem}(A_j)\frac{y_{\max}}{2^{j+1}} + \sum_{j>t} \mathrm{dem}(A_j)\frac{y_{\max}}{2^{j+1}}. \quad (9)$$

Note we can assume $\min_i \mathrm{dem}(i) = 1$ and thus $y_{\max} \leq 1$ (in the optimal solution to (SC-LP)). Then

$$\sum_{j>t} \mathrm{dem}(A_j)\frac{y_{\max}}{2^{j+1}} \leq \frac{1}{2^{t+1}} \sum_{j\geq 0} \mathrm{dem}(A_j) = \frac{1}{D^2} D = \frac{1}{D}.$$

Also, assume $D \geq 4$, for otherwise, we can get a 4-approximation (just connecting one pair $(s_i, t_i)$). So from (8), (9), and the above inequality, we have

$$\sum_{j\leq t} \mathrm{dem}(A_j)\frac{y_{\max}}{2^{j+1}} \leq \frac{1}{4}.$$

Recall $t = 2\log D - 1$. From the Pigeonhole principle, the proof is completed. □

Now suppose $A_h$ is derived as in Lemma 2.2. We will construct a feasible solution $x'$ to multicut (MC-LP) with $A_h$ and $(x, y)$: let $\alpha = 2^{h+1}/y_{\max}$, and $x'_e = \alpha x_e$

**Lemma 2.3** *$x'$ constructed above is a feasible solution to the multicut LP (MC-LP) on $A_h$.*

**Proof:** From (MC-LP), we just need to verify that for any $i \in A_h$ and $p \in P_{s_i t_i}$, $\sum_{e\in p} x'_e \geq 1$. In fact,

$$\sum_{e\in p} x'_e = \alpha \sum_{e\in p} x_e \geq \alpha y_i \geq \alpha \frac{1}{\alpha} = 1, \quad \text{where}$$

the first inequality is from (3) in (SC-LP) and the second is from the definition of $A_h$. □

From Lemma 2.1, we can find a multicut $E'$ (which separates pairs in $A_h$) based on $x'$, s.t. $c(E') \leq O(\log k) \sum_{e\in E} c_e x'_e$. We are ready to prove the main lemma of this section.

**Lemma 2.4** *For $E'$ found above, we have*

$$\mathrm{sparsity}(E') \leq O(\log k \log D) \sum_{e\in E'} c_e x_e.$$

**Proof:**

$$\begin{aligned}
\mathrm{sparsity}(E') &= c(E')/\mathrm{dem}(A_h) \\
&\leq O(\log k) \sum_{e\in E} c_e x'_e/\mathrm{dem}(A_h) \qquad \text{(from Lemma 2.1 and 2.3)} \\
&\leq O(\log k) \frac{\alpha}{\mathrm{dem}(A_h)} \sum_{e\in E} c_e x_e \\
&\leq O(\log k) 8\log D \sum_{e\in E} c_e x_e. \qquad \text{(from Lemma 2.2)}
\end{aligned}$$

□

**Theorem 2.5** *We can obtain an $O(\log k \log D)$-approximation to the* Sparsest Cut *problem.*

**Proof:** Now suppose $(x, y)$ is the *optimal solution* to (SC-LP). Note that the optimal value of (SC-LP) $\mathrm{OPT_{LP}} = \sum_{e \in E} c_e x_e \le \min_{E' \subseteq E} \mathrm{sparsity}(E') = \mathrm{OPT}$. So from Lemma 2.4, we have

$$\mathrm{sparsity}(E') \le O(\log k \log D) \sum_{e \in E} c_e x_e \le O(\log k \log D)\mathrm{OPT}.$$

□

Note: a sophisticated argument can be used to reduce the ratio to $O(\log^2 k)$.

# 3 Rounding Sparsest-Cut via Metric Embedding

We will show that we can obtain the optimum approximation ratio $O(\log k)$ for the Sparsest Cut problem via embedding metric spaces into real normed spaces.

## 3.1 Introduction to Metric Embeddings

We first give a brief introduction to metric embeddings here. Metric embeddings are a powerful tool in variety of settings and they got their impetus in computer science with the application to the Sparsest Cut problem.

Let $(V, d)$ and $(V', d')$ be two *metric spaces*. An *embedding* of $(V, d)$ into $(V', d')$ is a one-to-one map $f : V \to V'$. $f$ is an *expansion* if for any $u, v \in V$, $d'(f(u), f(v)) \ge d(u, v)$; $f$ is a *contraction* if for any $u, v \in V$, $d'(f(u), f(v)) \le d(u, v)$; $f$ is an *isometric* embedding if for any $u, v \in V$, $d'(f(u), f(v)) = d(u, v)$. The *distortion* of $f$ is defined to be

$$\mathrm{distortion}(f) = \max_{u,v \in V} \left\{ \frac{d'(f(u), f(v))}{d(u, v)}, \frac{d(u, v)}{d'(f(u), f(v))} \right\}.$$

Specifically, if $f$ is an expansion, then

$$\mathrm{distortion}(f) = \max_{u,v \in V} \frac{d'(f(u), f(v))}{d(u, v)};$$

if $f$ is a contraction, then

$$\mathrm{distortion}(f) = \max_{u,v \in V} \frac{d(u, v)}{d'(f(u), f(v))}.$$

The following results due to Bourgain. We will apply the $l_1$ version (Theorem 3.3) to get an $O(\log k)$ approximation for the Sparsest Cut problem.

**Theorem 3.1** *A finite metric space on $n$ points can be embedded into $l_2$ metric space in $\mathbb{R}^{O(\log n)}$ with distortion $O(\log n)$.*

**Theorem 3.2** *A finite $l_2$ metric space on $n$ points in $\mathbb{R}^d$ can be embedded isometrically into $l_1$ metric space in $\mathbb{R}^{f(d)}$, for some function $f$.*

**Theorem 3.3** *A finite metric space on $n$ points can be embedded into $l_1$ metric space in $\mathbb{R}^{O(\log^2 n)}$ with distortion $O(\log n)$.*

A metric space $(V, d)$ is $l_1$-*embeddable* if there exists an isometric embedding of $(V, d)$ into $l_1$.

**Exercise:** (i) Any tree-metric is $l_1$-embeddable. (ii) Any ring metric is $l_1$-embeddable.

## 3.2 Cut-Metrics and $l_1$ Embeddings

We establish an important relationship between *cut-metric* and $l_1$-embedding in this section.

**Remark:** *Note the metric discussed here may not satisfy the axiom "$d(u, v) = 0 \Leftrightarrow u = v$", but satisfies the rest three axioms of a metric.*

Given a set of points $V$, and $S \subseteq V$, the *cut-metric* $d_S$ induced by $S$ is

$$d_S = \begin{cases} 1, & \text{if } |S \cap \{u, v\}| = 1; \\ 0, & \text{otherwise.} \end{cases}$$

A metric space $(V, d)$ is *in the cone of cut-metrics* on $V$ if $\exists \lambda : 2^V \to \mathbb{R}^+$ s.t.

$$d(u, v) = \sum_{S \subseteq V} \lambda_S d_S(u, v).$$

It is not hard to verify the following claim.

**Claim 3.4** *For any $S_1, S_2 \subseteq V$ and $\lambda_1, \lambda_2 \geq 0$, $d' = \lambda_1 d_{S_1} + \lambda_2 d_{S_2}$ is a metric.*

**Theorem 3.5** *$(V, d)$ is $l_1$-embeddable iff $(V, d)$ is in the cone of cut-metrics on $V$.*

**Proof:** "$\Leftarrow$": Suppose $(V, d)$ is in the cone of cut-metrics on $V$ and $d = \sum_{S \in V} \lambda_S d_S$, we can embed any $u \in V$ into $f(u) \in \mathbb{R}^k$: $f(u) = (\lambda_1 I_{u \in S_1}, \lambda_2 I_{u \in S_2}, \ldots, \lambda_k I_{u \in S_k})$, where $S_1, S_2, \ldots, S_k$ are all subsets of $V$, and $I_{u \in S_i}$ is the indicator of whether $u \in S_i$.

"$\Rightarrow$": Now suppose $(V, d)$ is $l_1$-embeddable, we will show $(V, d)$ is in the cone of cut-metrics on $V$.

Suppose $f$ is an embedding of $V$ into $\mathbb{R}^h$, s.t. $u \in V$ is embedded into $\langle u(i) \rangle$, and

$$d(u, v) = ||f(u) - f(v)||_1 = \sum_{i=1}^{h} |u(i) - v(i)| = \sum_{i=1}^{h} d_i(u, v),$$

for any $u, v \in V$. Let $d_i(u, v) = |u(i) - v(i)|$. To prove $(V, d)$ is in the cone of cut-metrics on $V$, it is sufficient to prove $(V, d_i)$ is in the cone of cut-metrics on $V$, for $i = 1, \ldots, h$.

Let $V = \{v_1, v_2, \ldots, v_n\}$. W.l.o.g., assume that $v_1(i) \leq v_2(i) \leq \ldots \leq v_n(i)$. Let $S_j = \{v_1, v_2, \ldots, v_j\}$ and $\lambda_{S_j}(i) = v_{j+1}(i) - v_j(i)$, for $j = 1, 2, \ldots, n-1$. Also, let $\lambda_S(i) = 0$ if $S \subseteq V$ is not one of $S_1, \ldots, S_{n-1}$. It is easy to verify that $d_i = \sum_{S \subseteq V} \lambda_S(i) d_S$, i.e., $(V, d_i)$ is in the cone of cut-metrics on $V$. To prove $(V, d)$ is in the cone of cut-metrics, simply let $\lambda_S = \sum_{i=1}^{h} \lambda_S(i)$, since $d = \sum_{i=1}^{h} d_i = \sum_{i=1}^{h} \sum_{S \subseteq V} \lambda_S(i) d_S = \sum_{S \subseteq V} \left( \sum_{i=1}^{h} \lambda_S(i) \right) d_S$. □