1 Sparsest Cut and Metric Embeddings

In this lecture, we will describe an $O(\log n)$-approximation algorithm for the sparsest cut problem. We note that the best known approximation guarantee for this problem is currently $O(\sqrt{\log n})$. (See [ARV04].) Here, we will see an algorithm whose main tool is metric embeddings.

Given a graph $G = (V, E)$ and a partition of the vertices $(S, V \setminus S)$, we define:

$$\Phi_G(S) := \frac{\frac{1}{|E|} E(S, V \setminus S)}{\frac{2}{V^2} |S| \cdot |V \setminus S|}, \tag{1}$$

where $E(S, V \setminus S)$ indicates the number of edges crossing the partition. Note that the numerator is the total fraction of the edges that cross the cut, and the denominator is the total fraction of pairs of vertices that cross the cut. The sparsity $\Phi(G)$ of a graph G is given by the minimum sparsity over all possible cuts:

$$\Phi(G) := \min_{S \subset V} \Phi_G(S). \tag{2}$$

Let A denote the adjacency matrix for the graph G. The quantity $\Phi_G(S)$ can be rewritten using an indicator function of a cut

$$1_S(i) = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{otherwise,} \end{cases}$$

as

$$\Phi_G(S) = \frac{|V|^2}{2|E|} \cdot \frac{\sum_{i,j} A_{ij} |1_S(i) - 1_S(j)|}{\sum_{i,j} |1_S(i) - 1_S(j)|}.$$

Note that for any fixed graph, the quantity $|V|^2/(2|E|)$ is fixed. Thus, we could also try to approximate the quantity:

$$\frac{\sum_{i,j} A_{ij} |1_S(i) - 1_S(j)|}{\sum_{i,j} |1_S(i) - 1_S(j)|}.$$

This will not change any of the results we present in this lecture. Computing $\Phi_G(S)$ is known to be NP-hard. We will always use $n = |V|$.

1.1 LP Relaxation

A semimetric on a set X is a function $f : X \times X \to \mathbb{R}$, such that for all $x, y, z \in X$:

- $f(x, y) \geq 0$,
- $f(x, x) = 0$,
- $f(x, y) = f(y, x)$,
- $f(x, z) \leq f(x, y) + f(y, z)$.

Note that the cut metric $d_S(i,j) = |1_S(i) - 1_S(j)|$ is a semimetric on V. We can therefore relax the problem of computing $\Phi(G)$ by replacing the cut metric with a semimetric. We obtain the following relaxation (originally due to Leighton and Rao), which we will refer to as $LR(G)$.

$$LR(G) = \min \frac{|V|^2}{2|E|} \cdot \frac{\sum_{i,j} A_{ij} \cdot d(i,j)}{\sum_{i,j} d(i,j)} \tag{3}$$

$$d : V \times V \to \mathbb{R}, \tag{4}$$

$$d \text{ is a semimetric.} \tag{5}$$

We can see that $LR(G)$ provides a lower bound on $\Phi(G)$ since, as previously stated, a cut metric is a semimetric. We will also see that $\Phi(G) \leq LR(G) \cdot O(\log n)$ and we will give an algorithm to find a cut demonstrating this upper bound.

The current formulation of $LR(G)$ is not linear. However, it is possible to rewrite it as an equivalent linear program. That is, a solution d will be optimal and feasible for both the following LP and $LR(G)$.

$$\begin{aligned} &\min \sum_{i,j} A_{ij} d(i,j) \\ \text{subject to: } &\sum_{i,j} d(i,j) = \frac{|V|^2}{2|E|}, \\ &d(i,k) \leq d(i,j) + d(j,k), \\ &d(i,j) \geq 0. \end{aligned}$$

To see why the two are equivalent, note that for any optimal solution d, we can assume that the equality $\sum_{i,j} d(i,j) = \frac{|V|^2}{2|E|}$ holds. If it did not hold, it could be increased by scaling d, such that it holds and this would not change the value of $LR(G)$. Thus, the normalization factor and the denominator will cancel each other out, so that the value of $LR(G)$ equals the objective function of our LP. We can therefore find an solution (semimetric) for $LR(G)$ in polynomial time.

1.2 l_1-Embedding

We have relaxed the distance function in the form of $d_S(i,j) = |1_S(i) - 1_S(j)|$ to an arbitrary distance function. We will now consider an intermediate relaxation, in which we allow distance functions that can be realized by an embedding of vertices into an ℓ_1-space.

For a vector $x \in \mathbb{R}^m$, the ℓ_1-norm is defined as the following mapping from $\mathbb{R}^m$ to $\mathbb{R}$: $||x||_1 = \sum_i^m |x_i|$. This norm makes $\mathbb{R}^m$ into a metric space with the ℓ_1-distance function: $||x - y||_1 = \sum_i^m |x_i - y_i|$. It is a simple exercise to show that this distance function satisfies the requirements of a semimetric.

We will redefine $LR(G)$ as an optimization problem of finding an embedding $f : V \to \mathbb{R}^m$ such that:

$$\Phi'(G) = \inf_{f:V\to\mathbb{R}^m} \frac{|V|^2}{2|E|} \cdot \frac{\sum_{i,j} A_{ij} ||f(i) - f(j)||_1}{\sum_{i,j} ||f(i) - f(j)||_1}.$$

Theorem 1 *For every G, $\Phi'(G) = \Phi(G)$. Also, there exists a polynomial time algorithm that, given a mapping $f : V \to \mathbb{R}^m$, finds a cut $S \subset V$ such that:*

$$\frac{\sum_{u,v} A_{uv} |1_S(u) - 1_S(v)|}{\sum_{u,v} |1_S(u) - 1_S(v)|} \leq \frac{\sum_{u,v} A_{uv} ||f(u) - f(v)||_1}{\sum_{u,v} ||f(u) - f(v)||_1}.$$

Note that $\Phi'(G) \leq \Phi(G)$ because if we let $S^* \subset V$ denote the sparsest cut, we can set $f(i) = 1_{S^*}(i)$ for all $i \in V$. Now we will proceed to give an algorithmic proof of Theorem 1. We will use the following fact.

Fact 2 *For $a_i, b_i \geq 0$:*

$$\frac{\sum_{i=1}^{m} a_i}{\sum_{i=1}^{m} b_i} \quad \geq \quad \min_i \frac{a_i}{b_i}. \tag{6}$$

Proof Rewrite each a_i term as $a_i = x_i b_i$. (If $b_i = 0$, then set x_i to be arbitrarily large. In this case, if $a_i \neq 0$, then $a_i > x_i b_i$.) Clearly, the i^* minimizing the fraction will also minimize x_i. Thus, $\forall j : x_{i^*} \leq x_j$. We can rewrite the ratio of the sums as $\frac{\sum_j a_j}{\sum_j b_j} \geq \frac{\sum_j x_j b_j}{\sum_j b_j} \geq \frac{\sum_j x_{i^*} b_j}{\sum_j b_j}$. The two sides of the inequality are equal iff $\forall j : x_{i^*} = x_j$, otherwise a_{i^*}/b_{i^*} is strictly smaller than the ratio of the sums. ■

Let us apply Fact 2 to the inequality from Theorem 1.

$$\frac{\sum_{u,v} A_{u,v}||f(u) - f(v)||_1}{\sum_{u,v} ||f(u) - f(v)||_1} = \frac{\sum_i \sum_{u,v} A_{u,v}|f_i(u) - f_i(v)|}{\sum_i \sum_{u,v} |f_i(u) - f_i(v)|} \geq \min_i \frac{\sum_{u,v} A_{u,v}|f_i(u) - f_i(v)|}{\sum_{u,v} |f_i(u) - f_i(v)|}.$$

Let i^* denote the index that minimizes this fraction. Define a function $g : V \to \mathbb{R}$ such that $g()$ is a scaled/shifted version of $f_i()$ in which $\max_v g(v) - \min_v g(v) = 1$.[1]

Next we pick a threshold t uniformly at random from the interval $[\min_v g(v), \max_v g(v)]$. We define the set $S_t = \{v : g(v) \leq t\}$. Note that the probability of r being included in the set S_t is $\Pr(r \in S_t) = \max_v g(v) - g(r)$. It follows that $E[1_{S_t}(u)] = 1 \cdot Pr(u \in S_t)$. Thus,

$$E[|1_{S_t}(u) - 1_{S_t}(v)|] = |(max_w g(w) - g(v)) - (max_w g(w) - g(u))| = |g(u) - g(v)|.$$

Thus, we obtain:

$$\frac{\sum_{u,v} A_{u,v}||f(u) - f(v)||_1}{\sum_{u,v} ||f(u) - f(v)||_1} \geq \min_i \frac{\sum_{u,v} A_{u,v}|f_i(u) - f_i(v)|}{\sum_{u,v} |f_i(u) - f_i(v)|} = \frac{\sum_{u,v} A_{u,v}|g(u) - g(v)|}{\sum_{u,v} |g(u) - g(v)|}$$

$$= \frac{\sum_{u,v} A_{u,v} E[|1_{S_t}(u) - 1_{S_t}(v)|]}{\sum_{u,v} E[|1_{S_t}(u) - 1_{S_t}(v)|]} = \frac{E\left[\sum_{u,v} A_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|\right]}{E\left[\sum_{u,v} |1_{S_t}(u) - 1_{S_t}(v)|\right]}.$$

The last steps follow from linearity of expectation. We note that we can actually find a set S^* such that that:

$$\frac{\sum_{u,v} A_{u,v}|1_{S^*}(u) - 1_{S^*}(v)|}{\sum_{u,v} |1_{S^*}(u) - 1_{S^*}(v)|} \leq \frac{E\left[\sum_{u,v} A_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|\right]}{E\left[\sum_{u,v} |1_{S_t}(u) - 1_{S_t}(v)|\right]} \leq \frac{\sum_{u,v} A_{u,v}||f(u) - f(v)||_1}{\sum_{u,v} ||f(u) - f(v)||_1}. \tag{7}$$

To see this, let $a_t = \sum_{u,v} A_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|$ and let $b_t = \sum_{u,v} |1_{S_t}(u) - 1_{S_t}(v)|$. Then we have:

$$\frac{E[a_t]}{E[b_t]} \quad \leq \quad C, \Rightarrow \tag{8}$$

$$E[a_t] \quad \leq \quad C \cdot E[b_t]. \tag{9}$$

[1]Note that the fraction is invariant under scaling and shifting $f_i()$ by non-zero factors α and β, respectively:

$$\frac{\sum_{u,v} A_{u,v}|f_i(u) - f_i(v)|}{\sum_{u,v} |f_i(u) - f_i(v)|} = \frac{\sum_{u,v} A_{u,v}|\alpha f_i(u) + \beta - \alpha f_i(v) - \beta|}{\sum_{u,v} |\alpha f_i(u) + \beta - \alpha f_i(v) - \beta|} = \frac{\alpha \sum_{u,v} A_{u,v}|f_i(u) - f_i(v)|}{\alpha \sum_{u,v} |f_i(u) - f_i(v)|}.$$

We note that $E[b_t]$ is positive. Equation (9) implies that there is some set S_t for which (9) holds. We note that there are actually only n choices for t that give different cuts. (For example, choosing any t between $g(u)$ and $g(w)$, where $g(w)$ is the closest value to $g(u)$, yields the same cut.) Thus, if we try n values of t, we will find at least one set S_t for which (9) holds. Setting $S^* = S_t$ for this value of t, it follows that (7) holds for this set S^*, and this concludes the proof of Theorem 1.

1.3 Bourgain's Theorem

Theorem 3 *(Bourgain) Let $d : V \times V \to \mathbb{R}$ be a semimetric on V. Then there exist a mapping $f : V \to \mathbb{R}^m$ s.t. $\forall u, v \in V : ||f(u) - f(v)||_1 \leq d(u, v) \leq ||f(u) - f(v)||_1 \cdot c \cdot \log n$, where $c > 0$ is an absolute constant. Moreover, mapping f can be found efficiently with high probability.*

Theorem 3 implies $\Phi(G) \leq \log n \cdot LR(G)$:

$$LR(G) = \frac{|V|^2}{2|E|} \cdot \frac{\sum_{u,v} A_{uv} d(u,v)}{\sum_{u,v} d(u,v)} \geq \frac{|V|^2}{2|E|} \cdot \frac{\sum_{u,v} A_{uv} ||f(u) - f(v)||_1}{c \cdot \log n \sum_{u,v} ||f(u) - f(v)||_1} \geq \frac{1}{c \cdot \log n} \Phi(G).$$

Before we prove Theorem 3, we note the following fact, which will be useful. Embeddings of finite sets into ℓ_1 can be equivalently characterized as probabilistic embeddings into the real line.

Fact 4 *For every finite set V, dimension m, and mapping $F : V \to \mathbb{R}^m$, there is a finitely supported distribution D over functions $f : V \to \mathbb{R}$, such that $\forall u, v \in V$:*

$$E_{f \sim D}[|f(u) - f(v)|] = ||F(u) - F(v)||.$$

Conversely, for every finitely supported distribution D over function $f : V \to \mathbb{R}$, there is a dimension m and a mapping $F : V \to \mathbb{R}^m$ such that $\forall u, v \in V$:

$$E_{f \sim D}[|f(u) - f(v)|] = ||F(u) - F(v)||.$$

Proof For the first part, consider $F(v) = (F_1(v), F_2(v), \ldots, F_m(v))$. Define D to be the uniform distribution over the m functions of the form $x \to m \cdot F_i(x)$. It clearly fulfills the desired property: $E_{f \sim D}[|f(u) - f(v)|] = \sum_i 1/m \cdot |mF_i(u) - mF_i(v)| = ||F(u) - F(v)||_1$.

For the second part, suppose the distribution D is composed of the functions f_i, each occuring with probability p_i. Then we define a function $F(u) = (p_1 f_1(u), p_2 f_2(u), \ldots, p_m f_m(u))$. $F(u)$ fulfills the required property: $||F(u) - F(v)||_1 = \sum_i |p_i f_i(u) - p_i f_i(v)| = \sum_i p_i |f_i(u) - f_i(v)| = E_{f \sim D}[|f(u) - f(v)|]$. ■

Thus, it is sufficient to construct probabilistic embeddings into the real line. To apply Fact 4 constructively, we require that the support of the distribution D has a polynomial size. One approach to finding such a probabilistic embedding $f : V \to \mathbb{R}$ is to choose $r \in V$ uniformly at random and for each $u \in V$, let $f(u) = d(r, u)$. However, there are simple examples in which for some pairs of vertices u, v, $E[|f(u) - f(v)|] = d(u, v)/n$, while for other pairs of vertices u, v, $E[|f(u) - f(v)|]\ d(u, v)$. For instance, consider a set of vertices in which eat pair is at a distance 1. Then add a single additional vertex that is at distance 2 from all other vertices.

One way around this problem is to consider embeddings that are defined using distances to *subsets* of elements. The next fact pertains to such embeddings.

Fact 5 *Let $d : V \times V \to \mathbb{R}$ be a semimetric, and $A \subseteq V$ a non-empty subset. Define a mapping $f_A : V \to \mathbb{R}$ as $f_A(u) = \min_{r \in A} d(r, u)$. Then $\forall u, v \in V : |f_A(u) - f_A(v)| \leq d(u, v)$.*

Proof Let a and b be points s.t. $d(a, u) = f_A(u)$, $d(b, v) = f_A(v)$. By the minimality of $d(b, v)$ and triangle inequality: $d(b, v) \leq d(a, v) \leq d(u, v) + d(u, a)$. Thus, $|f_A(u) - f_A(v)| = |d(a, u) - d(b, v)| \leq |d(a, u) - d(u, v) - d(u, a)| = d(u, v)$. ■

If we define a function $f_A(u) = \min_{r\in A} d(r,u)$, Facts 4 and 5 imply that $E_{A\sim D}[|f_A(u) - f_A(v)|]$ lower bounds $d(u,v)$. In order upper bound $d(u,v)$ by $c \cdot \log n \cdot E_{A\sim D}[|f_A(u) - f_A(v)|]$, we want that $E_{D \sin A}[|f_A(u) - f_A(v)|]$ is not too much smaller than $d(u,v)$. What properties do we want from the subset $A \subset V$? For each u, v such that $d(u,v)$ is "large", let a, b denote the closest elements in A to u, v, respectively. We want that a is close to u and b is far from v, or vice-versa. This will ensure that the distance between u and v after embedding is also large. In fact, roughly speaking, if we could ensure that, for each u, v, this happens with some probability, we would have a good embedding. This is not always possible, but we can show that the following method of choosing subsets is sufficient to prove Bourgain's Theorem.

Theorem 6 *For a finite set of points V, consider a distribution D over subsets of V sampled by uniformly picking a scale $t \in \{0, 1, \ldots, \log n\}$. Each $v \in V$ is then picked to be in A with probability 2^{-t}. Let $d : V \times V \to \mathbb{R}$ be a semimetric. Then, the following holds for some constant $c > 0$:*

$$\forall u, v \in V : \ E_{A\sim D}[|f_A(u) - f_A(v)|] \geq \frac{1}{c \cdot \log n} d(u,v).$$

Proof For each t, let ru_t be the distance from vertex u to its 2^t-th closest point with respect to distances $d(u,v)$ (u is the $1 = 2^0$-th closest such vertex). So, the number of vertices having a distance less than ru_t from u is strictly less than 2^t and the number of vertices having a distance less than or equal to ru_t is greater than or equal to 2^t. In other words,

$$|\{w : d(u,w) < ru_t\}| < 2^t, \tag{10}$$
$$|\{w : d(u,w) \leq ru_t\}| \geq 2^t. \tag{11}$$

Let t^* be the scale such that both $ru_{t^*} < d(u,v)/3$ and $rv_{t^*} < d(u,v)/3$, and at least one of $rv_{t^*+1} \geq d(u,v)/3$, $rv_{t^*+1} \geq d(u,v)/3$ holds.

Now we define $ru'_t = \min\{ru_t, d(u,v)/3\}$ and $rv'_t = \min\{rv_t, d(u,v)/3\}$.

Claim 7 *There exists a constant c such that for every scale $t \in \{0, 1, \ldots, t^*\}$,*

$$E_{A\sim D_t}[|f_A(u) - f_A(v)|] \geq c \cdot (ru'_{t+1} + rv'_{t+1} - ru'_t - rv'_t).$$

Proof We will show that there are two disjoint events and each one occurs with probability at least c.

(i) $|f_A(u) - f_A(v)| \geq ru'_{t+1} - rv'_t$,

(ii) $|f_A(u) - f_A(v)| \geq rv'_{t+1} - ru'_t$.

Event (i) occus when A avoids set $S_1 = \{z : d(u,z) < ru'_{t+1}\}$ and intersects $S_2 = \{z : d(v,z) \leq rv'_t\}$. The set S_1 has size $|S_1| < 2^{t+1}$ and the set S_2 has size $|S_2| \leq 2^t$. Note that the sets S_1 and S_2 are disjoint.

What is the probability that A avoids S_1? It is at most $(1 - \frac{1}{2t})^{2^t} \sim \frac{1}{e}$. What is the probability that A intersects S_2? It is at most $1 - (1 - \frac{1}{2^t})^{2t} \sim 1 - \frac{1}{e}$. So with probability at least some constant c, event (i) occurs:

$$|f_A(u) - f_A(v)| \geq f_A(u) - f_A(v) \geq ru'_{t+1} - rv'_t.$$

Using an analogous argument, we can see that Event (ii) occurs if A avoids set $T_1 = \{z : d(u,z) \leq rv'_{t+1}\}$ and A intersects $T_2 = \{z : d(u,z) \leq ru'_t\}$. Since $|T_1| < 2^{t+1}$ and $|T_2| \leq 2^t$, we can see that with some constant probability, event (ii) occurs:

$$|f_A(u) - f_A(v)| \geq f_A(v) - f_A(u) \geq rv'_{t+1} - ru'_t.$$

As the two events are disjoint, and both happen with some probability at least c,

$$E_{A\sim D_t}[|f_A(u) - f_A(v)|] \geq c \cdot (ru'_{t+1} + rv'_{t+1} - ru'_t - rv'_t)$$

must hold. ■

Using our claim and averaging over all scales, we have:

$$\begin{aligned} E_{A\sim D}[|f_A(u) - f_A(v)|] &\geq \frac{1}{t^*}\sum_{i=0}^{t^*} c \cdot (ru'_{i+1} + rv'_{i+1} - ru'_i - rv'_i) && (12)\\ &\geq c \cdot (ru'_{t^*+1} + rv'_{t^*+1} - ru'_0 - rv'_0) && (13)\\ &\geq \frac{c}{\log n + 1} \cdot \frac{d(u,v)}{3}. && (14) \end{aligned}$$

This concludes the proof of Theorem 6. ■

Note that there are 2^n possible subsets of V that will each occur as A with some probability. We require that the number of subsets we actually use is only polynomial in n. However, using a Chernoff Bound, we can show that it suffices to sample $O(\log^3 n)$ subsets of V using the procedure described in Theorem 6, and use these sets as the support of the probability distribution D.

References

[Tre12] Luca Trevisan. Spectral graph theory and graph partitioning. ADFOCS Lecture Notes, 2012.

This lecture was based mainly on Chapter 7 of the Lecture Notes on Expander Graphs by Luca Trevisan. [Tre12].