Lecture 10: Mar 13, 2015

Lecturer: Deeparnab Chakrabarty *Scribe: Sayan Biswas & Indranil Bhattacharya*

10.1 Sparsest Cut in a Graph

Given a graph (V, E), in the sparsest cut problem our goal is to find a subset of vertices S, which minimizes the ratio $\frac{C(\delta S)}{|S|.|\bar{S}|}$ which we'll call $sp(S)$. Here δS denotes the cut-edges between S and $\bar{S}$, where $\bar{S}$ is precisely $V \setminus S$ and $C(\delta S)$ denotes the total cost of the edges in δS. This problem is equivalent to finding a set of edges $F \subseteq E$, minimizing $sp(F) = \frac{C(F)}{\#(\mathrm{s}_i,\mathrm{t}_i) \text{ pairs seperated by F}}$.

10.1.1 LP formulation

General Sparsest Cut:

Input: Graph G, $\{(s_i,\ t_i) \text{ pairs}\}_{i=1}^{k}$
Goal: It is easy to see that $sp(F)$ can be rewritten as $\frac{\sum_{e \in E} C_e X_e}{\sum_i d(s_i,t_i)}$ where $d(s_i,t_i)$ is defined as the shortest distance between vertices s_i and t_i in the graph defined with weight X_e on the edges. The reason is $d(s_i,t_i)$ = 0, if s_i and t_i are on the same side. Still, this objective function is not linear. So, we can get rid of the denominator by assuring that $\sum_i d(s_i,t_i) = 1$, which can be ensured by scaling operation. From the previous class we know, $d_e \leq X_e$. Hence, while minimizing $sp(S)$, we should use d_e instead of X_e. So, the *LP* for *General Sparsest Cut* which is given below, returns F with objective to

$$\begin{aligned} \text{minimize} & \sum_{e \in F} C_e d_e \\ \text{s.t.} \quad & d_{uw} \leq d_{uv} + d_{vw} && \forall \{u, v, w\} \in V \\ & \sum_{i=1}^{k} d_{s_i t_i} = 1 \\ & d_e \geq 0 && \forall e = (u, v) \in E \end{aligned}$$

Uniform Sparsest Cut:

In this scenario, every (u,v) pair is a (s_i,t_i) pair. So, our *LP* becomes,

$$\begin{aligned} \text{minimize} \quad & \sum_{e \in F} C_e d_e \\ \text{s.t.} \quad & d_{uw} \leq d_{uv} + d_{vw} && \forall \{u, v, w\} \in V \\ & \sum_{(u,v) \in V \times V} d_{uv} = 1 \\ & d_{uv} \geq 0 && \forall (u, v) \in v \times v \end{aligned}$$

10.1.2 *Sweep-cut* algorithm for *Uniform Sparsest Cut*

Sweep-Cut:

1. Fix a vertex s.
2. Rename the vertices $(v_1, v_2, \ldots, v_n)$ s.t. $d_{sv_1} \le d_{sv_2} \ldots \le d_{sv_n}$. We may assume $s = v_1$, as $d_{ss} = 0$.
3. Let $A_i := \{v_1, v_2, \ldots, v_i\}\ \forall\ i \in \{1, 2, \ldots, n\}$.
4. Return the A_i s.t. sp(A_i) is minimum.

Analysis of Sweep-Cut:

Let ALG be the sparsity of the cut returned. We define the following notions.

a. $B_r(s) = B_r := \{v|\ d_{sv} \le r\}$. We may assume $r \in [0, R]$ where $R = d_{sv_n}$. Note that for any r, B_r is one of the A_is.

b. $n_r(s) = n_r := |\bar{B}_r| =$ no. of vertices s.t. $d_{sv} > r$. $\bar{n}_r(s)$ is defined similarly.

As ALG returns the set of vertices with minimum sparsity, hence we have,

$$\text{ALG} \le \text{sp}(\text{B}_r) = \frac{C(\delta B_r)}{|B_r|.|\bar{B}_r|}$$

Which implies,

$$\begin{aligned} C(\delta B_r) &\ge ALG.|B_r|.|\bar{B}_r| \\ &= ALG.\bar{n}_r.n_r \\ &\ge ALG.n_r \quad (\bar{n}_r \ge 1 \text{ as it always contains s.)} \end{aligned}$$

Integrating both sides, we get

$$\int_0^R C(\delta B_r)dr \ge ALG \int_0^R n_r dr = ALG.\sum_v d_{sv} \tag{10.1}$$

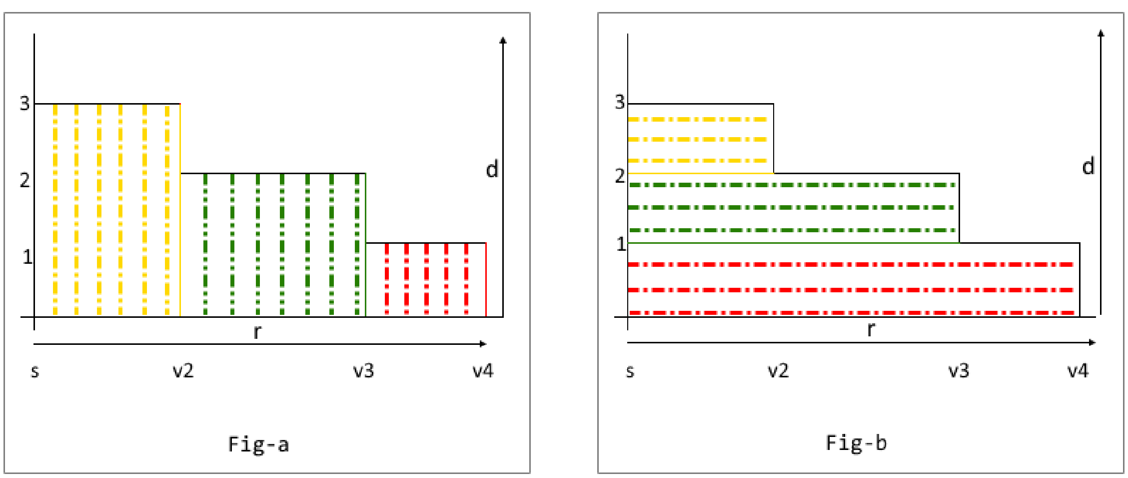

Figure 10.1: $\int_0^R n_r dr$ and $\sum_v d_{sv}$

The equality ($\int_0^R n_r\,dr = \sum_v \mathrm{d}_{sv}$) comes because *l.h.s.* represents *Fig-a* and *r.h.s.* is *Fig-b*, and both of those essentially represent the same area under the curve (*double-counting*).

Now, we have,

$$\begin{aligned}
1 = \sum_{u,v} d_{uv} &\leq \sum_{u,v} (d_{su} + d_{sv}) \\
&= \sum_u \sum_v d_{su} + \sum_u \sum_v d_{sv} \\
&\text{[Since we are summing over all vertices, } u \text{ can be replaced with } v.] \\
&= \sum_u \sum_v d_{sv} + \sum_u \sum_v d_{sv} \\
&= n. \sum_v d_{sv} + n \sum_v d_{sv} \\
&= 2n. \sum_v d_{sv} \\
\Rightarrow \sum_v d_{sv} &\geq \frac{1}{2n}
\end{aligned}$$

Using this lower bound of $\sum_v \mathrm{d}_{sv}$ in eqn *10.1*, we get

$$\int_0^R C(\delta B_r) dr \geq \frac{ALG}{2n}$$

Again, by definition, we get

$$\begin{aligned}
\int_0^R C(\delta B_r) dr &= \sum_{u,v} C(u,v).|d_{sv} - d_{su}| \quad [\text{as } d_{su} \leq r \leq d_{sv}] \\
&\leq \sum_{u,v} C(u,v).d_{uv} \quad [\text{From triangle inequality}] \\
&= LP
\end{aligned}$$

Applying the above two inequalities in eqn *10.1* we obtain

$$\begin{aligned}
LP &\geq \frac{ALG}{2n} \\
ALG &\leq O(n).LP
\end{aligned}$$

Hence, *Sweep-Cut* is an O(n) approximation algorithm for *Uniform Sparsest Cut*. ■

10.1.3 A better approximation factor for Uniform Sparsest Cut

Let us look at at a modified version of *Sweep-Cut*, where instead of taking a single vertex s, we take a set of vertices T at the beginning. Then the algorithm goes like this.

Modified Sweep-Cut:

1. Fix a vertex set T of size at-least $\frac{n}{3}$.
2. Rename the vertices (v_1, v_2,..., v_n) s.t. $\mathrm{d}_{Tv_1} \leq \mathrm{d}_{Tv_2}$ ... $\leq \mathrm{d}_{Tv_n}$, where $\mathrm{d}_{Tv_i} := \min_{t \in T} \mathrm{d}_{tv_i}$.
3. Let $A_i := \{v_1,\ v_2,...,\ v_i\}\ \forall\ i \in \{1,\ 2,...,\ n\}$.
4. Return the A_i s.t. $\mathrm{sp}(A_i)$ is minimum.

Analysis of Modified Sweep-Cut:

Let ALG_2 be the sparsity of the cut returned. We define the following notions.

1. $B_r(T) = B_r := \{v |\ d_{Tv} \leq r\}$.
2. $n_r(T) = n_r := |\bar{B}_r|$ = no. of vertices s.t. $d_{Tv} > r$. $\bar{n}_r(T)$ is defined similarly.

By the same logic, we have,

$$\begin{aligned} C(\delta B_r) &\geq ALG_2.|B_r|.|\bar{B}_r| \\ &= ALG_2.\bar{n}_r.n_r \\ &\geq ALG_2.\frac{n}{3}.n_r \quad (\bar{n}_r \geq \frac{n}{3} \text{ as it always contains T.)} \end{aligned}$$

Integrating both sides, we get

$$\int_0^R C(\delta B_r)dr \geq \frac{n}{3}.ALG_2 \int_0^R n_r dr = \frac{n}{3}.ALG_2.\sum_v d_{Tv} \tag{10.2}$$

Now, we have,

$$\begin{aligned} 1 = \sum_{u,v} d_{uv} &\leq \sum_{u,v} (d_{Tu} + d_{Tv} + diam(T)) \\ &\text{[Since, most likely the nearest vertices to } u \text{ and } v \text{ in T are different and can be furthest apart.]} \\ &= \sum_u \sum_v d_{Tu} + \sum_u \sum_v d_{Tv} + n^2.diam(T) \\ &\text{[Since we are summing over all vertices, } u \text{ can be replaced with } v.] \\ &= \sum_u \sum_v d_{Tv} + \sum_u \sum_v d_{Tv} + n^2.diam(T) \\ &= n.\sum_v d_{Tv} + n \sum_v d_{Tv} + n^2.diam(T) \\ &= 2n.\sum_v d_{Tv} + n^2.diam(T) \end{aligned}$$

Now suppose T had small diameter – that is, $diam(T) \leq 1/2n^2$. Then, we would get $\sum_v d_{Tv} \geq 1/2n$, and using this lower bound of $\sum_v \mathrm{d}_{sv}$ in eqn *10.2*, we get

$$\int_0^R C(\delta B_r)dr \geq c.ALG_2$$

The analysis for the upper bound still remains the same, hence we get LP as the upper bound. Applying the above two inequalities in eqn *10.2* we obtain

$$LP \geq c.ALG_2$$
$$ALG_2 \leq O(1).LP$$

This implies the following theorem

Theorem 10.1 *If there is a set T with $|T| \geq n/3$* **and** *$diam(T) \leq 1/2n^2$, then Modified Sweep-Cut from T is an $O(1)$-approximation algorithm for the Uniform Sparsest Cut problem.*

Of course such a special set T may not exist. Next, we see a different algorithm which implies a $O(\log n)$ approximation if no such 'teeny-diameter-with-many-many-points' set exist. To do so we need the following general purpose lemma.

Theorem 10.2 (Low Diameter Decomposition Lemma) *Given an undirected graph $G = (V,E)$ with cost C_e on each each e, and a distance d between all pairs of vertices, let $L = \sum_{e \in E} C_e d_e$. Given any $R > 0$, we can partition V into $\{V_1, V_2, \ldots, V_T\}$ in polynomial time such that*

1. $diam(V_i) \leq 2R$, $\forall$ $i \in \{1, 2, \ldots, T\}$
2. $\sum_{e \in E(V_1, V_2, \ldots, V_T)} C_e \leq O\left(\frac{\log n}{R}\right) L$ *where* $E(V_1, \ldots, V_T) := \{(u,v) \in E : u \in V_i, v \in V_j, i \neq j\}$.

We now describe the $O(\log n)$-approximation for uniform sparsest cut. Run the low diameter decomposition algorithm with $R = 1/4n^2$. Two cases arise.

Case 1: Among the T partitions of V, if $\exists$ i, s.t. $|V_i| \geq \frac{n}{3}$ and diam$(V_i) \leq \frac{1}{2n^2}$ we are done. Here we'll get a constant factor approximation from Theorem 10.1

Case 2: If there is no such partition, then initialize $S = \emptyset$. Order the parts $V_1, \ldots, V_T$ arbitrarily and go on inserting parts into S until $|S| > n/3$. As the initial parts are of relatively small size (i.e. all of them have size $< \frac{n}{3}$), $|S| \leq 2n/3$ implying $|\bar{S}| \geq n/3$. Also note that $\delta S \subseteq E(V_1, \ldots, V_T)$. This gives us

$$\begin{aligned}
sp(S) &= \frac{C(\delta(S))}{|S| \cdot |\bar{S}|} \\
&\leq \frac{9}{n^2} \cdot C(\delta(S)) \\
&\leq \frac{9}{n^2} \cdot C[E(V_1, V_2, \ldots, V_T)] \\
&\leq \frac{9}{n^2} \cdot O\left(n^2 \cdot \log n \cdot LP\right) \quad \text{[From the lemma]} \\
&= O(\log n) \cdot LP
\end{aligned}$$

■

10.1.4 Proof of Low Diameter decomposition lemma

We start with some definitions. Recall $B_r = B_r(s) := \{v : d(s,v) \leq r\}$, $\delta B_r := \{(u,v) : u \in B_r, v \notin B_r\}$ and $E[B_r] := \{(u,v) : u, v \in B_r\}$.

We define the "total LP volume" in a ball of radius r around a vertex s, as :

$$Vol(B_r) := \frac{L}{n} + \sum_{(u,v)\in E[B_r]} c_{uv}d_{uv} + \sum_{(u,v)\in \delta B_r} c_{uv}(r - d_{su}) \tag{10.3}$$

where $L = \sum_{e\in E} c_e d_e$ is the value of relaxed LP solution and n is the total number of vertices of the graph $G = (V, E)$. The first component in r.h.s is the "initial LP mass" which is same for all the balls "grown", the second component accounts for the mass due to internal edges in B_r, while the third is for the cross-over edges.

Fix a vertex s. Our goal is to find an $r \in [0,R)$ such that $c(\delta B_r) \le 4\frac{\log n}{R}.Vol(B_r)$. We claim that such a ball can be found. To see this, look at the rate at which the volume of the ball wrt r. To do so, we differentiate the equation on both sides w.r.t r, and get,

$$\frac{d}{dr}(Vol(B_r)) = \sum_{(u,v)\in\delta B_r} c_{uv} = c(\delta B_r)$$

Since the ball is still growing, we can assume $c(\delta B_r) > 4\frac{\log n}{R}.Vol(B_r)$ and arrive at a contradiction if possible. It is easy to check that at $r = R$, Vol(B_r) = $(L + \frac{L}{n}) = L(1 + \frac{1}{n})$. Thus we have

$$\begin{aligned}
\frac{d}{dr}(Vol(B_r)) &> 4\frac{\log n}{R}.Vol(B_r) \\
\implies \frac{d(Vol(B_r))}{Vol(B_r)} &> 4\frac{\log n}{R}.dr \qquad \textit{So,} \\
\int_{Vol(B_r)=\frac{L}{n}}^{L(1+\frac{1}{n})} \frac{d(Vol(B_r))}{Vol(B_r)} &> \int_{r=0}^{R} 4\frac{\log n}{R}.dr \\
\implies [\log(Vol(B_r))]_{\frac{L}{n}}^{L(1+\frac{1}{n})} &> 4\frac{\log n}{R}.[r]_0^R \\
\implies \log(n+1) &> 4.\log(n) \qquad \textit{which is a contradiction}
\end{aligned}$$

$\therefore$ There exists an $r \in [0, R)$ such that

$$c(\delta B_r) \le 4\frac{\log n}{R}.Vol(B_r) \tag{10.4}$$

Now coming back to the definition of $Vol(B_r)$ in (10.1), we can say

$$\begin{aligned}
r - d_{su} &\le d_{sv} - d_{su} \qquad (\because v \notin B_r, d_{sv} > r) \\
&\le d_{uv} \qquad (\textit{By triangle inequality, } d \textit{ is a metric})
\end{aligned}$$

$$\therefore Vol(B_r) \le \frac{L}{n} + \sum_{(u,v)\in E[B_r]} c_{uv}d_{uv} + \sum_{(u,v)\in\delta B_r} c_{uv}d_{uv}$$

Finally if we sum up both sides of (10.2) over all possible balls (in worst case, we could have n balls), so,

$$\begin{aligned}
R.H.S &= 4\frac{\log n}{R}.\sum_{B_r(i):i=1}^{n} Vol(B_r(i)) \\
&\le 4\frac{\log n}{R}.\Big[L + \sum_{B_r(i):i=1}^{n}\Big(\sum_{(u,v)\in E[B_r(i)]} c_{uv}d_{uv} + \sum_{(u,v)\in\delta B_r(i)} c_{uv}d_{uv}\Big)\Big] \\
&\le 4\frac{\log n}{R}.(L + 2L) = 4\frac{\log n}{R}.3L
\end{aligned}$$

and $L.H.S \ge c(E(v_1, v_2, ..., v_T))$ which conclusively proves the lemma. ■

10.2 General Sparsest Cut

The input is an undirected graph $G = (V, E)$, where each edge $e \in E$ has a non-negative capacity c_e. Also there are some k demand vertex pairs $\{(s_1, t_1), (s_2, t_2), ..., (s_k, t_k)\}$ each having some weight(demand) w_i. Given a subset of vertices $S \subseteq V$, we define "separation of S" as

$$sep(S) = \begin{cases} 1 & \textit{if } |\{s_i, t_i\} \cap S| = 1, \quad \textit{i.e } S \textit{ separates } (s_i, t_i) \textit{ pair} \\ 0 & \textit{otherwise} \end{cases}$$

Our objective is to find the cut of *minimum sparsity*, which is

$$\Psi^* = \min_{S \subseteq V} \frac{c(\delta S)}{\sum_{i=1}^{k} w_i.sep(S)}$$

It is easy to write the *general sparsest cut* in terms of d as :

$$\Psi^* = \min_{d} \frac{\sum_{e \in E} c_e d_e}{\sum_{i=1}^{k} w_i d(s_i, t_i)}$$

Like the uniform sparsest cut, the LP for this one will be :

$$LP = \min_{d} \sum_{e \in E} c_e d_e$$
$$s.t \sum_{i=1}^{k} w_i d(s_i, t_i) = 1,$$
and d is a *semi-metric* [2]

Before we go ahead with metric embedding, let us define what are so called the "nice metrics".
(1) Given a set $S \subseteq V$, we define :

$$f_S(u, v) := \begin{cases} 1 & \textit{if exactly one of u or v is in S} \\ 0 & \textit{otherwise} \end{cases}$$

Such a metric is called an *elementary cut metric* on V and it falls under the category of "nice metrics".
If the LP solution gives such a metric i.e if $d = f_S$, then trivially we can return S.
(2) However, if the cut metric f can be expressed as a linear combination of elementary cut metrics f_S i.e $f := \sum_{S \subseteq V} \alpha_S f_S$, $\alpha_S \geq 0 \ \forall S$, then also it is "nice".
If the LP solution $d = f$ with polynomially many $\alpha_S \geq 0$, then it can be shown that in the sparsity definition we could minimize over f. Since $f(u, v) = \sum\limits_{S \subseteq V} \alpha_S.f_S(u, v)$, so we can rewrite the LP as :

$$LP = \frac{\sum\limits_{e \in E} c_e \sum\limits_{S \subseteq V} \alpha_S f_S(e)}{\sum\limits_{i=1}^{k} w_i \sum\limits_{S \subseteq V} \alpha_S f_S(s_i, t_i)}$$
$$= \frac{\sum\limits_{S \subseteq V} \alpha_S \sum\limits_{e \in E} c_e f_S(e)}{\sum\limits_{S \subseteq V} \alpha_S \sum\limits_{i=1}^{k} w_i f_S(s_i, t_i)}$$
$$\geq \min_{S : \alpha_s \geq 0} \frac{\sum\limits_{e \in E} c_e f_S(e)}{\sum\limits_{i=1}^{k} w_i f_S(s_i, t_i)} = \psi^*$$

where the last but one inequality used the elementary fact that for positive reals $a_1, a_2, ..., a_k$ and $b_1, b_2, ..., b_k$, we have $\frac{\sum_{i=1}^k a_i}{\sum_{i=1}^k b_i} \geq \min_{i \in [k]} \frac{a_i}{b_i}$. ■

10.3 Sparsest cut from Metric embeddings

A metric f defined on V is called an $\mathcal{L}_1$ metric if there is a mapping $\phi : V \longrightarrow \mathbb{R}^k$ for some $k \geq 1$ such that $f(u, v)$ is the ℓ_1 distance between $\phi(u)$ and $\phi(v)$ i.e

$$f(u, v) = \|\phi(u) - \phi(v)\|_1 = \sum_{i=1}^{k} |\phi(u)_i - \phi(v)_i|$$

Claim : Given any metric space (V, f), where f is an $\mathcal{L}_1$ metric and can be written as the linear combination of some elementary cut metric, i.e $f = \sum_{S \subseteq V} \alpha_S f_S$, $\alpha_S \geq 0\ \forall S$, then f is also "nice".

Proof : Let's do it for $k = 1$ first and then we can generalize for any k. We plot the vertices on the real line and order them as (say) $\phi(v_1) \leq \phi(v_2) \leq ... \leq \phi(v_k)$. Let us denote $r_i = \phi(v_i)$. We define $\alpha_{S_i} := r_{i+1} - r_i$ where $S_i = \{v_1, v_2, ..., v_i\}$ and the i^{th} cut is between v_{i+1} and v_i. We put $f_{S_i}(v_{i+1}, v_i) = 1$ and 0 otherwise. It is easy to see that given any 2 vertices v_i and v_j $(j > i)$, we can write f as :

$$\begin{aligned} f(v_i, v_j) &= \sum_{c=i}^{j} \alpha_{S_c} . f_{S_c}(v_{c+1}, v_c) \\ &= \sum_{c=i}^{j} (r_{c+1} - r_c) . f_{S_c}(v_{c+1}, v_c) \\ &= r_j - r_i = \ \phi(v_j) - \phi(v_i) \end{aligned}$$

We can do the same trick now for any k. Fix a co-ordinate i and do the above procedure and check that $f(u, v) = \|\phi(u) - \phi(v)\|_1$.
Thus if the distance metric d given by LP is an $\mathcal{L}_1$ metric, we can get a constant factor approximation i.e

$$\Psi^* = \min_{d \in \mathcal{L}_1} \ \frac{\sum_{e \in E} c_e d_e}{\sum_{i=1}^{k} w_i d(s_i, t_i)}$$

However Linial, London and Rabinovich (1995) showed that given any metric space (V, d), if there exists a metric embedding $\phi : V \longrightarrow \mathbb{R}^k$ for some $k = O(poly(n))$ such that for any pair of vertices u and v, $\|\phi(u) - \phi(v)\|_1 \leq d(u, v)$, and $\|\phi(s_i) - \phi(t_i)\|_1 \geq \frac{d(s_i, t_i)}{\alpha}$, then there is an α factor approximation for the general sparsest cut i.e

$$\begin{aligned} ALG &\leq \frac{\sum\limits_{e=(u,v) \in E} c_e . \|\phi(u) - \phi(v)\|_1}{\sum\limits_{i=1}^{k} w_i . \|\phi(s_i) - \phi(t_i)\|_1} \\ &\leq \alpha . \frac{\sum\limits_{e=(u,v) \in E} c_e . d_e}{\sum\limits_{i=1}^{k} w_i . d(s_i, t_i)} = \alpha . LP \end{aligned}$$

The metric embedding result itself is due to Bourgain which will hopefully be covered in the next class. ■