E0 249: Approximation Algorithms February 7, 2022

Week-6
Lecture 1: Sparsest Cut

Instructor: Anand Louis *Scribe: V S S Aniketh*

1 Introduction

In this lecture, we will look at solving two related problems- sparsity (generalized) of a graph and expansion of a graph. First, the definitions of two problems will be introduced, followed by two Integer Programming formulations for solving the sparsity problem. Then, we will discuss the general sparsest cut problem and its reduction to the LP relaxation of the seconds formulation specified above. Also, we will briefly look at solving the general sparsest cut problem with O(log(n)) approximation. Finally, we will look into the expander problem.

2 Problem definitions

Consider a Weighted Graph $G = (V, E, w)$, where V is the set of vertices, E is the set of edges, w is the weight function mapping each edge to a positive real number. In the case of unweighted graph, we can consider each edge in the graph G being mapped to 1, under the weight function. Also let $|V| = n$
First, we define the following:

2.1 Some Definitions

Definition 1. ***(Weighted degree of vertex)****:*
The (weighted) degree of a vertex $u \in V$ is defined as $d_u = \sum_{v \in V} w_{uv}$.

Definition 2. ***(Uniform Sparsity of cut)****:*
Let a cut be $(S, V-S)$. Sparsity of this cut, $\Phi_G(S)$ is defined as:

$$\Phi_G(S) = \frac{\sum_{u \in S, v \in V-S} w_{uv}}{|S| \cdot |V-S|} \cdot |V|$$

Definition 3. ***(Generalized sparsity of a cut, Sparsest cut)****:*
Consider an undirected graph G with k commodities (s_i, t_i, d_i), $i \in \{1, 2, ...k\}$, where d_i is the positive demand associated with the i^{th} commodity and $s_i, t_i \in V$. Sparsity(generalized) of the graph is defined as:

$$\Phi_G(S) = \frac{\sum_{u \in S, v \in V-S} w_{uv}}{\sum_{i:|S \cap \{s_i, t_i\}|=1} d_i}$$

Sparsest cut in G is defined as $\operatorname{argmin}_{S \subset V, S \neq \phi} \Phi_G(S)$

Definition 4. ***(Expansion of a cut, Expansion of Graph)****:*
Let a cut be $(S, V-S)$. Expansion of this cut, $\phi_G(S)$ is defined as:

$$\phi_G(S) = \frac{\sum_{u \in S, v \in V-S} w_{uv}}{\min(\sum_{u \in S} d_u, \sum_{u \in V-S} d_u)}$$

Expansion of a graph G defined as $\phi_G = \operatorname{argmin}_{S \subset V, S \neq \phi} \phi_G(S)$

Definition 5. ***(Distance Label):***
Distance label z is defined as $z : E \to \mathbb{R}^+$. Also, we can extend the function to all vertex pairs, $x : V \cdot V \to \mathbb{R}^+$ as follows:

$$x_{uv} = z_{uv}, \forall (u,v) \in E$$
$$x_{uv} = Shortest\ distance\ betweenu, v\ under\ distance\ label\ z, for (u,v) \notin E$$

3 Uniform Sparsest Cut Problem Formulation

Formulation-1

Suppose for some $S \subset V$,

$$y_u = \begin{cases} 1, if u \in S \\ 0, if u \notin S \end{cases}$$

Then $\sum_{u,v \in V} w_{uv} \cdot |y_u - y_v| = w(S, V - S)$. Also, $\sum_{u \in V} y_u = |S|$.
Let $k = |S|$. For a particular value of k, we can write the Integer Programming Problem under this formulation as follows:

$$\begin{aligned} \min \ & \frac{1}{k \cdot (n-k)} \cdot \sum_{u,v \in V} w_{uv} \cdot |y_u - y_v| \\ st. \ & \sum_{u \in V} y_u = k \\ & y_u \in \{0,1\}, \forall u \in V \end{aligned}$$

Now, since there only $|V| = n$ values of k, we can solve the above integer programming problem for all the n possible values of k, to get the Uniform Sparsest Cut.
Now consider, the LP relaxation of the above Integer Programming problem, in which $y_u \in \{0,1\}$ is replaced by $0 \leq y_u \leq 1$. For any value of k, it is obvious to see that minimum cost function is attained by setting $y_u = \frac{k}{n}$, for all $u \in V$ and the value is 0.

Hence, this formulation is not much useful.

Formulation-2

Let $k = |S|$. The Integer programming problem under the current formulation is written as follows:

$$\begin{aligned} min \ & \frac{1}{k \cdot (n-k)} \cdot \sum_{u,v \in V} w_{uv} \cdot |y_u - y_v| \\ st. \ & \sum_{u,v \in V} |y_u - y_v| = k \cdot (n-k) & (1) \\ & y_u \in \{0,1\}, \forall u \in V & (2) \end{aligned}$$

In the LP relaxation, we once again replace the condition(2) with $0 \leq y_u \leq 1$. Now, in the absence of condition(1), we could have added 2 constraints $z_{uv} \geq (y_u - y_v)$ and $z_{uv} \geq (y_v - y_u)$, for every $u, v \in V$ and at the optimal solution we would have $z_{uv} = |y_u - y_v|$. But this won't work in the presence of condition (1).

Adding metric constraints on z_{uv}, we get the following LP formulation:

$$\begin{aligned} \min \; & \frac{1}{k \cdot (n-k)} \sum_{u,v \in V} w_{uv} \cdot z_{uv} \\ st. \; & \sum_{u,v \in V} z_{uv} = k \cdot (n-k) \\ & z_{uv} + z_{vw} \geq z_{uw}, \forall u, v, w \in V \\ & z_{uv} \geq 0, \forall u, v \in V \end{aligned}$$

Now, if we set $\frac{z_{uv}}{k \cdot (n-k)}$ to be z_{uv}, in the above LP, we get:

$$\begin{aligned} min \; & \sum_{u,v \in V} w_{uv} \cdot z_{uv} \\ st. \; & \sum_{u,v \in V} z_{uv} = 1 \\ & z_{uv} + z_{vw} \geq z_{uw}, forallu, v, w \in V \\ & z_{uv} \geq 0, \forall u, v \in V \end{aligned}$$

4 Generalized Sparsest Cut

4.1 Problem Formulation

Consider a graph $G = (V, E, w)$ and the multi commodity demand flow problem [1] on this graph, with k commodities (s_i, t_i, d_i), $i \in \{1, 2, ...k\}$, where $s_i, t_i \in V$ and d_i is the demand associated with the commodity.
Let P_i be the set of paths from s_i to t_i.
Let f_p be a flow in path p.
Let c_e be a capacity of edge e, c_e is 0 if edge is not present in Graph G.
The LP for this Multi-commodity demand flow problem is given as:

$$\begin{aligned} \max \; & f \\ st. \; & \sum_{P : P \in P_i} f_P \geq f \cdot d_i, \forall i \in \{1, 2, ...k\} \\ & \sum_{P : P \ni e} f_P \leq c_e, \forall e \in E \end{aligned}$$

The Dual for the above problem is:

$$\begin{aligned} \min. \; & \sum_{e \in E} c_e \cdot x_e \\ st. \; & \sum_{i=1}^{k} d_i \cdot y_i \geq 1 \\ & \sum_{e : e \in P} x_e \geq y_i, \forall P \in P_i, 1 \leq i \leq k \\ & y_i \geq 0, \forall 1 \leq i \leq k \\ & x_e \geq 0, \forall e \in E \end{aligned} \tag{LP-1}$$

This Dual is the LP for Generalized Sparsest Cut problem. It can be shown [1] that the above LP has an optimal solution which satisfies the following:

- The distance label x_e assigned to each edge $e \in E$ forms a metric, i.e., $\forall u, v, w \in V (x_{uw} \le x_{uv} + x_{vw})$.
- $dist(s_i, t_i) = x_{s_i t_i} = y_i$, for all $i \in \{1, 2, ...k\}$
- $\sum_{i=1}^{k} y_i \cdot d_i = 1$

Therefore, the distance labels x_e for the Optimal solution satisfy the following equations:

$$\sum_{i=1}^{k} x_{s_i t_i} \cdot d_i = 1$$
$$x_{uv} + x_{vw} \ge x_{uw}, \forall u, v, w \in V$$

Hence, the LP for Generalized sparsest cut has the same optimal value as that of the LP below:

$$\begin{aligned} min. & \sum_{e \in E} c_e \cdot x_e \\ st. & \sum_{i=1}^{k} d_i \cdot x_{s_i t_i} = 1 \\ & x_{uv} + x_{vw} \ge x_{uw}, \forall u, v, w \in V \end{aligned} \tag{LP-2}$$

Let f^* be the optimal value for the Multi-commodity demand flow problem.
For any set $S \subset V$, define the following:

- $\delta(S)$ = Set of edges crossing the cut $(S, V - S)$
- $c(S) = \sum_{e \in \delta(S)} c_e$
- $d(S) = \sum_{i:|\{s_i, t_i\} \cap S| = 1} d_i$

Lemma 6. *For every set $S \subset V$, $f^* \le \frac{c(S)}{d(S)}$*

Proof. Let $A = \{i : |S \cap \{s_i, t_i\}| = 1\}$. Hence $d(S) = \sum_{i \in A} d_i$
$f^* \cdot d_i \le \sum_{P \in P_i} f_P$, for all $i \in \{1, 2, ...k\}$

$$f^* \cdot d(S) = \sum_{i \in A} f^* \cdot d_i \le \sum_{i \in A} \sum_{P: P \in P_i} f_P. \tag{3}$$

Also, for every $i \in A$, for every path P from s_i to t_i, there is at least one edge e in $\delta(S)$ which is contained in the path P.
Hence

$$\sum_{i \in A} \sum_{P: P \in P_i} f_P \le \sum_{e \in \delta(S)} \sum_{P: P \ni e} f_P \le \sum_{e \in \delta(S)} c_e = c(S) \tag{4}$$

From (3) and (4), we get:

$$f^* \cdot d(S) \le c(S)$$
$$f^* \le \frac{c(S)}{d(S)}$$

Hence proved. □

4.2 Reduction to Sparsest Cut Problem

Introduce a commodity for each pair of points($^nC_2$ in all) $u, v \in V, u \neq v$, with demand being 1, i.e.$(u, v, 1)$. Substituting this into (LP-2) for Generalized Sparsest Cut problem , we get:

$$\begin{aligned} min. \sum_{e \in E} c_e * x_e \\ st. \sum_{u,v \in V} x_{uv} = 1 \\ x_{uv} + x_{vw} \geq x_u w, \forall u, v, w \in V \end{aligned}$$

which is the same as the second formulation given for Uniform Sparsest Cut Problem.

5 Solving Generalized Sparsest Cut Problem

Definition 7. ***(Metric Embeddings)*** *A metric (V, x), x is distance label assigned to edges, embeds into l_1 with distortion α, if there exits an $f : V \to \mathbb{R}^m$ for some m such that:*

$$x(u, v) \leq ||f(u) - f(v)|| \leq \alpha \cdot x(u, v), \forall u, v \in V$$

Definition 8. ***(Cut packing)*** *Let $G = (V, E)$ be a graph, with a distance labels x assigned to the edges, which are in turn satisfying metric property. Let y be function $y : 2^V \to \mathbb{R}^+$. y is said to be α approximate cut packing (iff):*

$$x_e \leq \sum_{S: e \in \delta(S)} y_S \leq \alpha \cdot x_e, \forall e \in E$$

where $\delta(S)$ is the set of edges crossing the cut $(S, V - S)$.

Lemma 9. *Let $\sigma : V \to \mathbb{R}^m$ be a mapping. There is a cut packing $y : 2^V \to \mathbb{R}^+$, st. for all edges (u, v), we have:*

$$\|\sigma(u) - \sigma(v)\|_1 = \sum_{S:(u,v) \in \delta(S)} y_S$$

Moreover, the number of non-zero y_S, where $S \in 2^V$, is at the most $m \cdot (n - 1)$, where $|V| = n$

Proof. Refer to Lemma 21.10 in [1] □

Lemma 10. *Let $y : 2^V \to \mathbb{R}^+$ be a cut packing with m non-zero y_S. Then, there is a mapping $\sigma : V \to \mathbb{R}^m$, such that for each edge (u, v):*

$$\sum_{S:(u,v) \in \delta(S)} y_S = \|\sigma(u) - \sigma(v)\|_1$$

Proof. Refer to Lemma 21.11 in [1] □

Lemma 11. *Given a graph $G = (V, E)$, with distance labels d assigned to edges, satisfying metric property: There is an α-approximate cut-packing y (iff) There is an embedding σ with α-distortion.*

Proof. Follows from Lemma-9 and Lemma-10. □

Lemma 12. *Let $y : 2^V \to \mathbb{R}^+$ be an α-approximate cut packing for (V, x), where x is the distance labels, as obtained on solving (LP-2) of the Generalized Sparsest Cut problem. Further, let $S^* = \text{argmin}(\{\frac{c(S)}{d(S)}$: where S is st. $y_S > 0\})$. Then*

$\frac{c(S^*)}{d(S^*)} \leq \alpha * |OPT|$

where $c(S)$, $d(S)$ for a set $S \subset V$ are as defined in Section 4.1 and OPT is optimal solution for the generalized sparsest cut problem

Proof. Consider $\sum_{e \in E} c_e \cdot x_e$:

$$\begin{aligned}\sum_{e \in E} c_e \cdot x_e &\geq \sum_{e \in E} c_e \cdot (\sum_{S: e \in \delta(S)} y_S) \cdot \frac{1}{\alpha} \\ &= \frac{1}{\alpha} \cdot \sum_S y_S \sum_{e: e \in \delta(S)} c_e \\ &= \frac{1}{\alpha} \cdot \sum_S c(S) \cdot y_S\end{aligned}$$

Now let $e_i = (s_i, t_i)$, where (s_i, t_i, d_i) is the i^{th} commodity. Also define $I_{e_i}^{(S)}$ as follows:

$$I_{e_i}^{(S)} = \begin{cases} 1, if |\{s_i, t_i\} \cap S| = 1 \\ 0, otherwise \end{cases}$$

Then, we have:

$$\begin{aligned}1 &= \sum_{i=1}^{k} x_{e_i} \cdot d_i \\ &\leq \sum_{i=1}^{k} (\sum_{S: e_i \in \delta(S)} y_S) \cdot d_i \\ &= \sum_S y_S \cdot (\sum_{i=1}^{k} I_{e_i}^{(S)} \cdot d(i)) \\ &= \sum_S y_S \cdot d(S)\end{aligned}$$

Therefore, we have:

$$\begin{aligned}\frac{\sum_{e \in E} c_e \cdot x_e}{1} &\geq \frac{1}{\alpha} \cdot \frac{\sum_S c(S) \cdot y_S}{\sum_S y_S \cdot d(S)} \\ \frac{\sum_S c(S) \cdot y_S}{\sum_S y_S \cdot d(S)} &\geq \frac{c(S^*)}{d(S^*)} \\ \alpha * \sum_{e \in E} c_e \cdot x_e &\geq \frac{c(S^*)}{d(S^*)}\end{aligned}$$

Let f^* be as defined in Sec-4.1. From Lemma-1, we have $\frac{c(OPT)}{d(OPT)} \geq f^*$, where OPT is an optimum for generalized sparsest cut problem. Also, from Strong Duality of LP, it follows that $f^* = \sum_{e \in E} c_e * x_e$.

Therefore, $\frac{c(OPT)}{d(OPT)} \geq \sum_{e \in E} c_e * x_e$ and hence:

$$\frac{c(S^*)}{d(S^*)} \leq \alpha \cdot \frac{c(OPT)}{d(OPT)}$$

□

Theorem 13. ***(Bourgain's Theorem)***
Given an n point metric space (X,d) there exists a randomized polynomial time algorithm to compute an embedding $f : X \rightarrow \mathbb{R}^{poly(n)}$ with $O(\log(k))$-distortion, with high probability, where k is the number of commodities

Algorithm

Combining Bourgain's Theorem with Lemma-2 and Lemma-5, we get a polynomial time randomized algorithm that is of $O(log(k))$-approximation for the generalized sparsest cut problem, with high probability.

In Uniform sparsest cut problem, since the number of commodities is ${}^nC_2$, we get $O(\log({}^nC_2)) = O(\log(n))$-approximate algorithm for the Uniform Sparsest cut problem.

6 Solving Expansion of a Graph

Procedure

- Introduce commodities for each pair of vertices u,v : $(u, v, d_u * d_v)$, where d_u is weighted degree of vertex u, d_v is the weighted degree of vertex v.
- With the above commodities run the algorithm for generalized sparsest cut problem and let the outputted set be S^*
- Return S^*

Lemma 14. *Set S^* returned by the above Procedure is $O(\log(n))$-approximation of the expansion of G*

Proof. Observe that with demands for commodities as defined in the procedure, we get:

$$d(S) = \sum_{u \in S} d_u \cdot \sum_{v \in V-S} d_v$$

where $S \subset V$ and d_u is weighted degree of vertex $u \in V$.
Let OPT be the optimal for Generalized Sparsest Cut problem. Because the Algorithm suggested in Section-5 is an $O(log(n))$-approximation algorithm, we have:

$$\frac{c(S^*)}{d(S^*)} \leq O(log(n)) \cdot \frac{c(OPT)}{d(OPT)} \leq O(log(n)) \cdot \frac{c(S_1)}{d(S_1)}. \tag{5}$$

Let $m_S = min\{\sum_{u \in S} d_u, \sum_{v \in V-S} d_v\}$ and $M_S = max\{\sum_{u \in S} d_u, \sum_{v \in V-S} d_v\}$. Note that for any Set $S \subset V$:

$$m_S + M_S = \sum_{u \in V} d_u$$
$$m_S \leq M_S$$

Hence,

$$0 \leq m_S \leq \frac{\sum_{u \in V} d_u}{2} \leq M_S \leq \sum_{u \in V} d_u \tag{6}$$

Combining (5), (6), we get:

$$\frac{c(S^*)}{m_{S^*}} \leq 2 \cdot O(log(n)) \cdot \frac{c(S_1)}{m_{S_1}}$$
$$\frac{c(S^*)}{m_{S^*}} \leq O(log(n)) \cdot \frac{c(S_1)}{m_{S_1}}$$

Hence, S^* is an $O(\log(n))$-approximation to S_1, which is expansion of G. ☐

7 Expanders

Let $G = (V, E)$ be d-regular expander graph $|V| = n$, $n > N_0$ where N_0 is sufficiently large. Then, $h(G) > \varepsilon$ for some constant $\varepsilon > 0$, where $h(G)$ is the edge expansion defined as:

$$h(G) = min_{S:0<|S|\leq \frac{n}{2}} \frac{e(S, V-S)}{|S|}$$

where $e(S, V-S)$ is the number of edges crossing the $(S, V-S)$. From the definition of $h(G)$, and that $h(G) > \varepsilon$, we have:

$$\frac{e(S, V-S)}{|S|} > \varepsilon$$
$$\frac{e(S, V-S)}{|S|} \cdot \frac{n}{|V-S|} > \varepsilon$$

for all $0 < |S| \leq \frac{n}{2}$. Hence $\min_{0<|S|\leq \frac{n}{2}} \frac{e(S,V-S)}{|S|} \cdot \frac{n}{|V-S|} = \Phi_G > \varepsilon$ too.
Hence, $\Phi_G = \Omega(1)$

Define capacities st. $c_e = 1$, for all $e \in E$ and $c_{uv} = 0$, for all $(u, v) \notin E$
Now, consider the metric z defined on the graph as follows: Set $z_{uv} = 1$, for all $(u, v) \in E$

$$z_{uv} = \text{shortest distance between } (u, v) \ \forall (u, v) \notin E$$

We have:

$$\sum_{u,v\in V} c_{uv} z_{uv} = \frac{n \cdot d}{2}$$

Fix a vertex $u \in V$. Number of vertices at a distance of at most k is $\leq d^1 + d^2 + \ldots . d^k \leq d^{(k+1)}$.

Hence, for $k = \lfloor (\log_d(\frac{n}{2})) \rfloor - 1$, at least $\frac{n}{2}$ vertices are at a distance of at least k from u.
Therefore,

$$\sum_{u,v\in V} z_{uv} \geq \sum_{u\in V} \frac{n}{2} \cdot k = \Omega(n^2 \log_d n)$$

Now setting $r_{uv} = \frac{z_{uv}}{\sum_{u,v\in V} z_{uv}}$, we see that r is a feasible solution for the following LP of the Uniform Sparsest cut problem:

$$\begin{aligned}
\min & \sum_{u,v\in V} w_{uv} \cdot z_{uv} \\
st. & \sum_{u,v\in V} z_{uv} = 1 \\
& z_{uv} + z_{vw} \geq z_{uw}, \ \forall u, v, w \in V \\
& z_{uv} \geq 0, \ \forall u, v \in V
\end{aligned}$$

Hence,

$$n \cdot \sum_{u,v\in V} c_{uv} \cdot r_{uv} = n \cdot \frac{\sum_{u,v\in V} c_{uv} \cdot z_{uv}}{\sum_{u,v\in V} z_{uv}} = n \cdot O(\frac{n * d}{n^2 \log_d n}) = O(\frac{1}{\log(n)})$$

Now let x be the optimal solution for the above LP. Let OPT be the uniform sparsest cut for graph G

$$\Phi_G(OPT) = \Omega(1)$$
$$n \cdot \sum_{u,v\in V} c_{uv} x_{uv} \leq n \cdot \sum_{u,v\in V} c_{uv} r_{uv} = O(\frac{1}{log(n)})$$

Integrality Factor= $\frac{\Omega(1)}{O(\frac{1}{\log(n)})} = \Omega(log(n))$
Also, Bourgain's Theorem is tight up to constant factors.

References

[1] V.V. Vazirani, *Approximation Algorithms*. Springer Berlin Heidelberg, 2013

[2] David P. Williamson and David B. Shmoys, *The Design of Approximation Algorithms*. Cambridge University Press, 2011