Cut Sparsification

Definition: A $(1 \pm \varepsilon)$-cut sparsifier $H=(V,E_H)$ is a weighted subgraph of a given graph $G=(V,E_G)$ such that for all subsets $\emptyset \subset S \subset V$:

$$(1-\varepsilon)\,|E_G(S,V\setminus S)| \leq |E_H(S,V\setminus S)| \leq (1+\varepsilon)\,|E_G(S,V\setminus S)|$$

$$|E_H(S,V\setminus S)| = \sum_{\substack{(u,v)\in E_H \\ u\in S,\, v\in V\setminus S}} w_H(u,v)$$

Today: G undirected, unweighted

Straightforward idea:

→ Include each edge (u,v) of G in H with probability p (and if so, set $w_H(u,v) = \frac{1}{p}$) (in general $w_H(u,v) = \frac{1}{p} \cdot w_G(u,v)$)

Analysis

- Consider for every edge $(u,v) \in E_G$ random variable $X_{(u,v)} = \begin{cases} 1 & \text{if } (u,v) \text{ is sampled} \\ 0 & \text{otherwise} \end{cases}$

- Let $(S, V\setminus S)$ be an arbitrary cut

$$E[|E_H(S, V\setminus S)|] = E\Big[\sum_{\substack{(u,v)\in E_G \\ u\in S, v\in V\setminus S}} X_{(u,v)} \cdot \frac{1}{p}\Big] = \frac{1}{p} \sum_{\substack{(u,v)\in E_G \\ u\in S, v\in V\setminus S}} \overbrace{E[X_{u,v}]}^{p} = \frac{1}{p}\cdot p \cdot |E_G(S, V\setminus S)| = |E_G(S, V\setminus S)|$$

<u>But</u>: This only gives expectation for each cut

<u>Markov Bound</u>: $Pr[|E_H(S, V\setminus S)| > (1+\varepsilon)|E_G(S, V\setminus S)|] < \frac{1}{1+\varepsilon}$

$\hookrightarrow$ only constant probability, cannot use union bound to get guarantee for <u>all</u> (possibly exponentially many) cuts

$\hookrightarrow$ only gives a bound in one direction

<u>Chernoff Bound</u>:

$$Pr\Big[\sum_{(u,v)\in E_G} X_{(u,v)} > (1+\varepsilon)\cdot \underbrace{p\cdot |E_G(S, V\setminus S)|}_{\mu} \vee \sum_{(u,v)\in E_G} X_{(u,v)} < (1-\varepsilon)\cdot p\, E_G(S, V\setminus S)\Big]$$

$$\leq 2 \cdot \frac{1}{e^{\varepsilon^2/3 \cdot \mu}} \leq \frac{1}{n^c} \quad \text{if } \mu \geq \frac{3}{\varepsilon^2}\cdot (c+1)\cdot \ln(n)$$

$$\Rightarrow |E_G(S, V\setminus S)| \geq \frac{1}{p}\cdot \frac{3}{\varepsilon^2}(c+1)\ln(n)$$

$$2 \cdot \frac{1}{e^{\varepsilon^2/3 \cdot \mu}} \leq 2 \cdot \frac{1}{e^{\cancel{\varepsilon^2}/\cancel{3} \cdot \frac{\cancel{3}}{\cancel{\varepsilon^2}}(c+1)\ln(n)}}$$

$n \geq 2$ because $n = 1$ does not need spars.

$$= 2 \cdot \frac{1}{\left(e^{\ln(n)}\right)^{c+1}} = 2 \frac{1}{n^{c+1}} \leq \frac{n}{n^{c+1}} = \frac{1}{n^c}$$

Summary: By Chernoff Bound, each cut of size $\Omega\left(\frac{1}{p} \cdot \frac{\log(n)}{\varepsilon^2}\right)$ will be preserved with quality $(1 \pm \varepsilon)$ with high probability

Size of H:

$$E[|E_H|] = p \cdot |E_G| = p \cdot m \leq p \cdot n^2$$

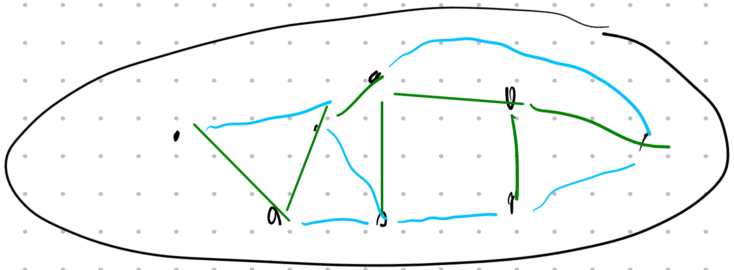
Goal: Approximately preserve all cuts while reducing the number of edges by a factor of $1/2$ (in expectation) ["half-sparsify"]

We will prove the following Theorem:

Thm: The half-sparsify algorithm computes, with probability $\geq 1 - \frac{1}{n^{c+1}}$ (for a given constant c), a $(1 \pm \varepsilon)$-cut sparsifier $H = B \cup G'$ of a given unweighted graph G of size $|E(G')| \leq \frac{3}{4} |E(G)|$ and $|B| = O\left(\frac{n \log n}{\varepsilon^3}\right)$.

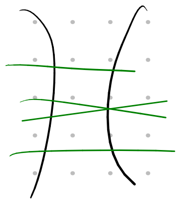
Def.: t-bundle

In a t-bundle $B = (V, F_1 \cup \dots \cup F_t)$ we have that for all $1 \leq i \leq t$: (V, F_i) is a spanning forest of $(V, E \setminus \{F_1 \cup \dots \cup F_{i-1}\})$



Straightforward algorithm
Repeat t times:

- Compute a spanning forest add it to B, remove it from G



<u>Observation</u>: Consider some cut $(S, V\setminus S)$

- if $|E_G(S, V\setminus S)| \le t$, then all edges crossing the cut are contained in t-bundle B
- if $|E_G(S, V\setminus S)| > t$, then t edges crossing the cut are contained in t-bundle B

<u>Algorithm Half-Sparsify</u> for $G=(V,E)$ (unweighted, undirected)

Set $p = \frac{1}{2}$

Set $t = 3\cdot(c+3)\cdot\frac{1}{p}\cdot\frac{1}{\varepsilon^3}\ln n$

Compute a t-bundle B

$G' = \emptyset$

Add every edge (u,v) in $G\setminus B$ to G' with probability p and set $w_{G'}(u,v) = \frac{1}{p}$

Return $H := B \cup G'$

Analysis

Size: (almost) clear

every spanning forest has size $\leq n$

- $|B| \leq n \cdot t = O\left(n \frac{\log n}{\varepsilon^3}\right)$
- $Ex[|E(G')|] = p \cdot |E(G \setminus B)| \leq p \cdot |E(G)| = \frac{1}{2} |E(G)|$

We can easily modify the algorithm such that $Pr\left[|E(G')| \leq \frac{3}{4} |E(G)|\right] \geq 1 - \frac{n}{c+1}$ by a Chernoff bound

Approximation Quality

Consider an arbitrary cut $(S, V \setminus S)$

Case 1: If $|E_G(S, V \setminus S)| \leq t$, then all edges crossing the cut are containd in $B \subseteq H$

$$|E_G(S, V \setminus S)| = w_H(E_H(S, V \setminus S))$$

$1 - \varepsilon \approx \frac{1}{1+\varepsilon}$ for small ε

Case 2: If $t < |E_G(S, V \setminus S)| \leq (1+\varepsilon) t$

- $w_H(E_H(S, V \setminus S)) \geq t \geq \frac{1}{1+\varepsilon} |E_G(S, V \setminus S)| \geq (1-\varepsilon) |E_G(S, V \setminus S)|$
- $w_H(E_H(S, V \setminus S)) \leq t + \varepsilon t \cdot \frac{1}{p} \leq (1 + 2\varepsilon) \cdot t \leq (1+2\varepsilon) \cdot |E_G(S, V \setminus S)|$

Case 3: $|E_G(S, V\setminus S)| > (1+\varepsilon)t$
$= t + \varepsilon t \cdot \alpha$ for some $\alpha \geq 1$

Let $F_S = E_G(S, V\setminus S) \setminus B$ be the set of edges crossing the cut not contained in B, and let F_S' be those edges of F that have been sample

It suffices to show that

$$w_H(F_S') = (1 \pm \varepsilon) \cdot |F_S| \quad \text{with high prob.}$$

$$(1-\varepsilon) \cdot |F_S| \leq w_H(F_S') \leq (1+\varepsilon)|F_S|$$

Let $X_{u,v}$ be random variable $= \begin{cases} 1 & \text{if } (u,v) \text{ was sampled by alg.} \\ 0 & \text{otherwis} \end{cases}$ for every edge (u,v)

$\mu := p \cdot |F_S|$

$$\Pr[w_H(F_S') \neq (1 \pm \varepsilon)|F_S|] = \Pr\Big[\sum_{(u,v) \in F_S} X_{u,v} \neq (1 \pm \varepsilon)\mu\Big]$$

Chernoff bound

$$\leq \frac{2}{e^{\varepsilon^2/3 \cdot \mu}} = \frac{2}{e^{\varepsilon^2/3 \cdot p|F_S|}} = \frac{2}{e^{\varepsilon^2/3 \, p \cdot \varepsilon t \cdot \alpha}}$$

$$= \frac{2}{e^{(c+3) \cdot \alpha \cdot \ln n}} = \frac{2}{n^{(c+3)\alpha}}$$

probability reduces as α increases

The number of cuts of size $\alpha\varepsilon t$ is at most $O(n^{2\cdot\alpha\varepsilon t})$

We can now apply
- a union bound over all cuts of size $\alpha\varepsilon t$
- a union bound over all n^2 possible values of α

→ This gives the claimed error probability