DP cuts:

"Global" Min-Cut:

Input: $G=(V,E)$

Output: $S \subset V$, $S \neq \emptyset$

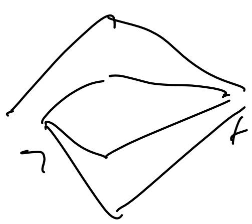
Objective: min $|E(S,\bar{S})| = |\{\{u,v\}\in E : u\in S, v\in \bar{S}\}|$

Non-private:

- Can compute min s-t cut $\forall s,t$, take min, but slow
- "Contraction"-based algorithm:
  - Karger: $\tilde{O}(n^2 m)$, Karger-Stein: $\tilde{O}(n^2)$
- Some super useful structural facts about cuts in graphs.
  - In particular:

<u>Thm</u> (Karger): If G has min-cut C, there are at most $n^{2\alpha}$ cuts of size at most αC.

- Very not true for min s-t cut!



$C = n-2$

min-cuts $= 2^C$

Privacy:

- Edge DP.
- If want #, not structure: Lap(1/ε) noise,

$$E[ALG] \le OPT + O(1/\varepsilon)$$

- Structure harder!

Paper: "Differentially Private Combinatorial Optimization":

Gupta, Ligett, McSherry, Roth, Talwar SODA '10

Idea: exponential mechanism!

- Almost works, but not quite

Thm (exp. mechanism): If we use the exp. mechanism to select $r \in R$ using score fn $q(D,r)$, then

$$\Pr\left[q(EM(D,R,q)) < \max_{r \in R} q(D,r) - \frac{1}{\varepsilon}\ln|R| - \frac{t}{\varepsilon}\right] \le \exp(-t)$$

So apply EM to cuts, w/ -size as score fn.

Say OPT = 1. # cuts of size $t \approx n^{2t}$: too large!

EM only downweights by e^{-t}.

What if $OPT(G) \ge 2\ln n$?

$\Rightarrow$ # cuts of size $OPT + t$ = # cuts of size $(1+\frac{t}{OPT})OPT$

$\leq n^{2(1+\frac{t}{OPT})} \leq n^2 \cdot n^{2t/OPT} = n^{2} \cdot n^{\frac{2t}{2\log n}}$

$= n^2 \cdot e^t$

So right ballpark for EM!

How do we ensure OPT large?

- Add some edges using EM!

__Alg:__

- Let $H_0 \subset H_1 \subset \dots \subset H_{\binom{n}{2}}$ arbitrary strictly increasing sets of edges
- Choose $i \in [0, \binom{n}{2}]$ w/ prob. prop. to $\exp(-\varepsilon\,|OPT(G \cup H_i) - \frac{8\ln n}{\varepsilon}|)$ (EM w/ score fn $-|OPT - \frac{8\ln n}{\varepsilon}|$)
- Choose cut $S \subset V$, $S \neq \emptyset$, w/ prob. prop. to $\exp(-\varepsilon \cdot |E_{G \cup H_i}(S, \bar{S})|)$ (EM w/ score fn = $-$size).
- Output S

__Thm:__ 2ε-DP

__Pf:__ two calls to EM, basic composition.

Thm: $E[c(A(G))] \le OPT + O(\frac{1}{\varepsilon}\ln n)$

Pf:

Claim: $\frac{4}{\varepsilon}\ln n \le OPT(G \cup H_i) \le OPT + O(\frac{1}{\varepsilon}\ln n)$
w/ prob. $\ge 1 - \frac{1}{n^2}$

Pf: Sps $OPT \ge \frac{8}{\varepsilon}\ln n$.
$\Rightarrow$ first inequality true, optimal
choice is H_0, score $-(OPT - \frac{8\ln n}{\varepsilon})$

$|R| = \Theta(n^2)$, set $t = 2\ln n$

$\Rightarrow$ by EM, $Pr[OPT(G \cup H_i) \ge OPT + \frac{2}{\varepsilon}\ln n + \frac{2}{\varepsilon}\ln n] \le \frac{1}{n^2}$ ✓

Sps $OPT < \frac{8}{\varepsilon}\ln n$
$\Rightarrow$ there is some i^* s.t. $|OPT(G \cup H_{i^*})| = \frac{8}{\varepsilon}\ln n$

$\Rightarrow Pr[|OPT(G \cup H_i) - \frac{8\ln n}{\varepsilon}| > 0 + \underbrace{\frac{2}{\varepsilon}\ln n + \frac{2}{\varepsilon}\ln n}_{= \frac{4}{\varepsilon}\ln n}) \le \frac{1}{n^2}$

$\Rightarrow \frac{4}{\varepsilon}\ln n \le OPT(G \cup H_i) \le OPT + \frac{4}{\varepsilon}\ln n$ ✓

Assume claim holds.

Let C_t = # cuts of size $\le OPT(G \cup H_i) + t$ in $G \cup H_i$

$\Rightarrow$ by Karger:

$$C_t \le n^{2\left(1+\frac{t}{OPT(G\cup H_i)}\right)} \le n^{2\left(1+\frac{t}{\frac{4}{\varepsilon}\ln n}\right)}$$

$$= n^2 \cdot e^{\frac{\varepsilon t}{2}}$$

what prob. does EM assign to each cut in C_t?

"proportional" to $\exp(-|E_{G\cup H_i}(S,\bar{S})|)$

Since $\exists$ cut in $G\cup H_i$ of size $OPT(G\cup H_i)$, any cut in C_t has prob. $\le e^{-\varepsilon t} \cdot$ prob. of optimal cut

$$\le e^{-\varepsilon t}$$

$\Rightarrow$ $Pr[$output cut of size $> OPT(G\cup H_i)+b] \le$

$$\le \sum_{t>b} e^{-\varepsilon t}(C_t - C_{t-1})$$

$$\le \sum_{t>b} \left(e^{-\varepsilon t} C_t - e^{-\varepsilon t} C_{t-1}\right)$$

$$\le \sum_{t>b} \left(e^{-\varepsilon t} - e^{-\varepsilon(t+1)}\right) C_t$$

$$= \sum_{t>b} e^{-\varepsilon t}\left(1-e^{-\varepsilon}\right) C_t$$

$$\le \left(e^{\varepsilon}-1\right) \sum_{t>b} e^{-\varepsilon t} C_t$$

$$\le \left(e^{\varepsilon}-1\right) \sum_{t>b} e^{-\varepsilon t} n^2 e^{\frac{\varepsilon t}{2}}$$

$$= (e^{\varepsilon}-1) \sum_{t > b} n^2 e^{-\frac{\varepsilon t}{2}}$$

$$\le (e^{\varepsilon}-1) \cdot n^2 \cdot \frac{\exp(-\frac{\varepsilon b}{2})}{\exp(\frac{\varepsilon}{2}-1)}$$

$\varepsilon < 1$

$$\le O(1) \cdot n^2 \cdot \exp(-\frac{\varepsilon b}{2})$$

Let $b = \frac{8 \ln n}{\varepsilon}$

$$\le O(1) \cdot n^2 \cdot \exp(-4 \ln n) = O(1) \cdot n^2 \cdot n^{-4}$$

$$= O(n^{-2})$$

So also within $O(\frac{1}{\varepsilon}\ln n)$ of $\mathrm{OPT}(G \cup H_i)$ except w/ prob $O(\frac{1}{n^2})$

$$\Rightarrow E[c(ALG)] \le \underbrace{O(n^2 \cdot \tfrac{1}{n^2})}_{\substack{\text{either of the} \\ \text{two events fail}}} + \mathrm{OPT}(G \cup H_i) + O(\tfrac{1}{\varepsilon} \ln n)$$

$$\le \mathrm{OPT}(G) + O(\tfrac{1}{\varepsilon} \ln n) \quad \checkmark$$

Ok, but what if want poly time?

Useful fact from Karger's generation:

$n^{2\alpha+1}$ runs of Karger will output <u>all</u> cuts of size $\le \alpha \cdot \mathrm{OPT}$ except w/ exponentially small prob.

So after generating H_i, run Karger n^2 times to generate n^2 cuts

$\Rightarrow$ w.h.p., all cuts of size $\leq 3 \cdot OPT(G \cup H_i)$ in set.

run EM just on those cuts.

<u>Thm</u>: $(2\varepsilon, O(\frac{1}{n^2}))$-DP

<u>pf</u>: Know from prev analysis that w/ prob. $\geq 1 - O(n^{-2})$,

alg returns cut of size $\leq OPT(G \cup H_i) + \frac{8 \ln n}{\varepsilon}$

$\leq 3 \cdot OPT(G \cup H_i)$ (since $OPT(G \cup H_i) \geq \frac{4 \ln n}{\varepsilon}$)

So except w/ prob. $O(\frac{1}{n^2})$, original alg. outputs something from our set.

Now "couple" 2 distributions: they can assign diff. probs to same cut, but just a rescaling!

$\rightarrow (2\varepsilon, O(\frac{1}{n^2}))$-DP

<u>Thm</u>: Same utility guarantee.

<u>pf</u>: Same dist. except w/ $O(\frac{1}{n^2})$ mass.

Extensions:

- min st cut: very different!

  DMN NeurIPS '23: ε-DP alg w/ error $O(\frac{n}{\varepsilon})$

  - Any (ε, δ)-DP alg has error $\Omega(n)$

    for $\varepsilon \leq 1$, $\delta \leq 0.1$

  (But also works for weights, abrig like in APSD)

- min k-cut: (CDFZ) $O(\frac{k}{\varepsilon} \log)$ error

- Multiway cut:

  DMN: $(2, \frac{n \log k}{\varepsilon})$

  CDFZ: $(1.2965, \frac{nk}{\varepsilon})$