DP APSD:

Graph DP, but can't do node- or edge-DP!

- One edge removed: distance jumps from finite to ∞
  $\Rightarrow$ infinite sensitivity, can't do anything.

Instead: graph public, weights private.

- Graph $G=(V,E)$
- $w: E \to \mathbb{R}_{\geq 0}$
- w, w' neighboring databases if $\|w-w'\|_1 \leq 1$
- APSD: output $d(u,v)$ $\forall u,v \in V$ $\hat{d}$
  - want to do privately!

  error of $\hat{d} := \max_{u,v \in V} |\hat{d}(u,v) - d(u,v)| = \|\hat{d} - d\|_\infty$

Intuition: models situations where graph known, but weights influenced by private behavior.

E.g., $w(u,v)$ a function of traffic on edge, so weights function of where people are driving.

Approach 1: add noise to all distances after computing them!

$\Delta_1 d = \Theta(n^2)$

$\Rightarrow$ Laplace: add $\mathrm{Lap}\left(\frac{n^2}{\varepsilon}\right)$-noise

$\Rightarrow$ ε-DP, but error $\approx \frac{n^2}{\varepsilon}$!

$\Delta_2 d = \Theta(n)$ $\left(\sqrt{\sum_n \sum_n 1^2}\right)$

$\Rightarrow$ Gaussian: add $N(0, \sigma^2)$ noise for $\sigma = \Theta\left(\frac{n\sqrt{\log 1/\delta}}{\varepsilon}\right)$

$\Rightarrow$ error $\approx \tilde{O}\left(\frac{n}{\varepsilon}\right)$, (ε,δ)-DP

Q1: Can we get $\tilde{O}(n)$ error for ε-DP?

Idea: instead of computing distances then adding noise, add noise then compute distances!

$\hat{w} = w + \text{Noise} \Rightarrow \Delta_1 w = 1$ (def)

$\Rightarrow \hat{w}(e) = w(e) + \overset{\text{New}}{\mathrm{Lap}}(1/\varepsilon)$ $\forall e \in E$ is ε-DP.

(ε,δ)-DP: $\Delta_2 w = 1 \Rightarrow \hat{w}(e) = w(e) + N\left(0, \frac{\log 1/\delta}{\varepsilon^2}\right)$

is (ε,δ)-DP $\Rightarrow$ no gain over ε-DP.

$\Rightarrow \hat{d}$ = shortest paths in $\hat{w}$: post-processing!

Issue: what if negative cycle!

Solution: If $\hat{w}(e) < 0 \Rightarrow \hat{w}(e) = 0$ (post-processing).

Thm: w.h.p., $err(\hat{d}) \leq \tilde{O}(n/\varepsilon)$

Incorrect proof: for gaussian noise, noise on path is sum of gaussians w/ variance $\frac{\log 1/\delta}{\varepsilon^2}$ $\Rightarrow$ same as gaussian w/ variance $\frac{n \log 1/\delta}{\varepsilon^2}$ $\Rightarrow$ $\frac{\sqrt{n \log 1/\delta}}{\varepsilon}$ error w.h.p.

True, but lots of paths! Too many for union bound!

Pf: Laplace tail bound: $\Pr[|Lap(1/\varepsilon)| \geq t \cdot \frac{1}{\varepsilon}] = e^{-t}$

$\Rightarrow \Pr[N_e > \frac{1}{\varepsilon} c \ln n] \leq n^{-c}$

$\Rightarrow \Pr[\max_{e \in E} N_e > \frac{1}{\varepsilon} c \ln n] \leq n^{-c+2}$

$c \geq 3$: w.h.p. max edge error $\leq O(\frac{1}{\varepsilon} \ln n)$

Fix $u, v \in V$, consider arbitrary u-v path P.

$$\Rightarrow |\hat{w}(P) - w(P)| \overset{\text{clipping}}{\leq} \left| \sum_{e \in P} (w(e) + N_e) - \sum_{e \in P} w(e) \right|$$

$$= \left| \sum_{e \in P} N_e \right| \leq O(|P| \cdot \frac{1}{\varepsilon} \ln n) \leq O(n \cdot \frac{1}{\varepsilon} \ln n)$$

Let $\hat{P} = \operatorname{argmin}_{u\text{-}v \text{ paths } P} \hat{w}(P)$

$P^* = \operatorname{argmin}_{u\text{-}v \text{ path } P} w(P)$

$\hat{d}(u,v) = \hat{w}(\hat{P}) \leq \hat{w}(P^*) \leq w(P^*) + O(\frac{1}{\varepsilon} n \ln n) = d(u,v) + O(\frac{1}{\varepsilon} n \ln n)$

$\hat{d}(u,v) = \hat{w}(\hat{P}) \geq w(\hat{P}) - O(\frac{1}{\varepsilon} n \ln n) \geq w(P^*) - O(\frac{1}{\varepsilon} n \ln n) = d(u,v) - O(\frac{1}{\varepsilon} n \ln n)$

$$\Rightarrow |\hat{d}(u,v) - d(u,v)| \le O\left(\tfrac{1}{\epsilon}\, n \ln n\right)$$

Can we do better?

Idea: - want to only consider <u>few-hop paths</u>.

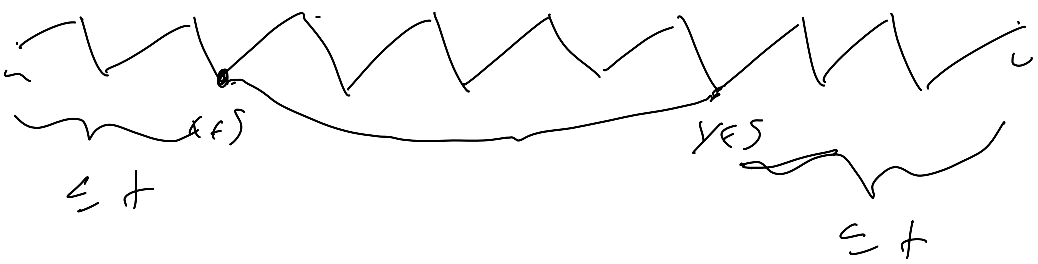
- But what if shortest paths has many hops?

- Shortcuts! (hopsets!)

Let S = subsample each node indep. w/ prob. $p = \Theta\left(\frac{\ln n}{t}\right)$

$\Rightarrow s = |S| = \Theta\left(\frac{n \ln n}{t}\right)$ w.h.p.

Idea: use S as shortcuts! If we have good estimate for distances in $S \times S$, can use that to "jump"!



For $u,v \in S$, let $\hat{d}(u,v) = d(u,v) + \text{Lap}\left(\frac{s^2}{\epsilon}\right)$ (Approach <u>1</u>)

$\forall u,v \in V$, let $\hat{d}^{t}(u,v) = \min\limits_{\substack{u\text{-}v \text{ path } P \\ \text{of length} \le t}} \hat{w}(e)$

$\forall u,v : |\{u,v\} \cap S| < 2$, let

$$\hat{d}(u,v) = \min\left(\hat{d}^{t}(u,v),\ \min_{x,y \in S} \left(\hat{d}^{t}(u,x) + \hat{d}(x,y) + \hat{d}^{t}(y,v)\right)\right)$$

Thm: ε-DP.

Pf: For distances in $S \times S$, sensitivity $\leq s^2$

$\Rightarrow$ ε-DP by Laplace mechanism

$\hat{w}(e)$ DP by previous analysis

Everything else post-processing

About t error from $\hat{d}^t$, about $s^2 \approx \frac{n^2}{t^2}$ error from $S \times S$

$\Rightarrow t = \frac{n^2}{t^2} \Rightarrow t = n^{2/3}$

So set $t = n^{2/3}$

Thm: w.h.p., error $\leq \tilde{O}\left(\frac{n^{2/3}}{\varepsilon}\right)$

Pf: Fix $u, v \in V$.

If $u, v \in S \Rightarrow$ error from $Lap\left(\frac{s^2}{\varepsilon}\right) = \tilde{O}(n^{2/3})$ w.h.p.

o/w let P^* shortest path (true).

Claim: $\hat{d}(u,v) \leq d(u,v) + O\left(\frac{n^{2/3}}{\varepsilon}\right)$

If $|P^*| \leq t \Rightarrow \hat{d}(u,v) \leq \hat{d}^t(u,v) \leq \hat{w}(P^*)$

$\leq w(P^*) + O\left(t \frac{\ln n}{\varepsilon}\right)$

$= d(u,v) + \tilde{O}\left(\frac{n^{2/3}}{\varepsilon}\right)$ ✓

else, w.h.p. $P^* \cap S \neq \emptyset$

Let x closest node in $P^* \cap S$ to u
y " " " " " v

w.h.p. x in first t hops, y in last t hops

$\Rightarrow \hat{d}(u,v) \le \hat{d}^t(u,x) + \hat{d}(x,y) + \hat{d}^t(y,v)$

w.h.p. $\rightarrow \le d(u,x) + O(t \cdot \frac{1}{\varepsilon} \ln n) + d(x,y) + \tilde{O}(n^{2/3}/\varepsilon) + d(y,v) + O(t \cdot \frac{1}{\varepsilon} \ln n)$

$\le d(u,v) + \tilde{O}\left(\frac{n^{2/3}}{\varepsilon}\right)$

Claim: $\hat{d}(u,v) \ge d(u,v) - \tilde{O}\left(\frac{n^{2/3}}{\varepsilon}\right)$

$\hat{d}^t(a,b) \ge d^t(a,b) - O\left(\frac{t \ln n}{\varepsilon}\right) \quad \forall a,b \in V$

(by earlier analysis).

$\Rightarrow$ if $\hat{d}(u,v) = \hat{d}^t(u,v)$, done ✓

else $\hat{d}(u,v) = \hat{d}^t(u,x) + \hat{d}(x,y) + \hat{d}^t(y,v)$ for some $x,y \in S$

$\ge d^t(u,x) - O\left(\frac{t \ln n}{\varepsilon}\right) + d(x,y) - \tilde{O}\left(\frac{S^2}{\varepsilon}\right) + d(y,v) - O\left(\frac{t \ln n}{\varepsilon}\right)$

$\ge d(u,x) + d(x,y) + d(y,v) - \tilde{O}\left(\frac{n^{2/3}}{\varepsilon}\right)$

$\ge d(u,v) - \tilde{O}\left(\frac{n^{2/3}}{\varepsilon}\right)$ ✓

For (ε,δ)-DP: use Gaussian mechanism for $S \times S$ distances

$\Rightarrow \tilde{O}\left(\frac{\sqrt{n}}{\varepsilon}\right)$ - error!

Further results:

- Can't do better than $\tilde{\Omega}\left(\frac{n^{1/4}}{\varepsilon}\right)$-error.
- On trees, can do polylog error! Generalizes paths, where distances $\approx$ interval queries
  $\Rightarrow$ binary tree mechanism gives polylog error
- Minor free graphs (e.g., planar graphs): $\tilde{O}((nW)^{1/3})$, where W = max edge weight allowed
  (generalized binary tree mechanism).