DP Densest Subgraph (DSG):

DSG: Input: - $G=(V,E)$

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

Objective: $\max \rho(S) = \frac{|E(S)|}{|S|} = \frac{|\{\{u,v\} \in E : u,v \in S\}|}{|S|}$

$\rho = \frac{1}{2}$ avg degree in subgraph induced by S.

Privacy: Edge-DP: $G \sim G'$ if they differ in one edge.

Local edge-DP: untrusted curator, nodes know only their own incident edges.

Requirement: entire transcript of communication is ε-DP.

Charikar's Algorithm:

- $S_1 = V$
- for $i = 2$ to n
  - Let v_i be min-deg. node in $G[S_{i-1}]$
  - $S_i = S_{i-1} \setminus \{v_i\}$
- Return $\underset{i \in [n]}{\operatorname{argmax}} \rho(S_i)$

<u>Thm</u>: 2-approximation! : $\forall G,\ \rho(ALG(G)) \geq \frac{1}{2}\rho(OPT(G))$

Want to make private. But lots of iterations!

Can we get fewer iterations?

Parallel version (Bahmani, Kumar, Vassilvitskii):

- $S_1 = V$
- $i = 1$
- while $S_i \neq \emptyset$
  - Let $L_i = \{v \in S_i : d_{S_i}(v) \leq 2(1+\eta)\rho(S_i)\}$
  - $S_{i+1} = S_i \setminus L_i$
  - $i = i+1$
- return $\arg\max_i \rho(S_i)$

<u>Thm</u>: $2(1+\eta)$-approx

<u>Thm</u>: # iterations $\leq O(\frac{1}{\eta}\log n)$

<u>Pf</u>: Consider iteration i.

avg $d_{S_i}(v)$ for nodes in S_i is $\frac{\sum_{v \in S_i} d_{S_i}(v)}{|S_i|} = \frac{2|E(S_i)|}{|S_i|}$

$= \mathbb{E}_{v \sim S_i}[d_{S_i}(v)] \qquad = 2\rho(S_i)$

$\Rightarrow |L_i| = |\{v \in S_i : d_{S_i}(v) \leq 2(1+\eta)\rho(S_i)\}|$

$= |S_i| - |\{v \in S_i : d_{S_i}(v) > (1+\eta)\cdot \mathbb{E}_{v \sim S_i}[d_{S_i}(v)]\}|$

$\geq |S_i| - \frac{1}{1+\eta}|S_i|$ (Markov's inequality)

$\geq \frac{\eta}{1+\eta} |S_i|$

$\Rightarrow |S_{i+1}| \leq |S_i| - \frac{\eta}{1+\eta} |S_i| \leq \frac{1}{1+\eta} |S_i|$

$\Rightarrow$ total iterations $\leq O(\log_{1+\eta} n) = O(\frac{1}{\eta} \log n)$

$\Rightarrow$ if we can make each iteration private, use basic composition $\Rightarrow$ only extra $O(\log n)$ loss.

So want to make private (and local)

- $S_1 = V$, $i = 1$
- while $S_i \neq \emptyset$:
  - every node $v \in S_i$:
    - draws $N_i(v) \sim Lap(\frac{2}{\varepsilon})$
    - computes $\hat{d}_i(v) = d_i(v) + N_i(v)$, sends to curator
  - Curator:
    - computes $\hat{\rho}(S_i) = \frac{1}{2|S_i|} \sum_{v \in S_i} \hat{d}_i(v)$
    - computes $T_i = 2(1+\eta)\hat{\rho}(S_i)$
    - computes $L_i = \{v \in S_i : \hat{d}_i(v) \leq T_i\}$
    - Computes and makes public $S_{i+1} = S_i \setminus L_i$
    - $i = i+1$
- Curator selects $\arg\max_i \hat{\rho}(S_i)$

Lemma: # iterations $\leq O(\frac{1}{\eta} \log n)$

Pf: Same as before!

T_i is $(1+\eta) \cdot \underset{v \sim S_i}{E} [\hat{d}_i(v)]$

after $\hat{d}_i(v)$ is all calculated!

$\Rightarrow \leq \frac{1}{1+\eta} |S_i|$ vertices above threshold

$\Rightarrow |S_{i+1}| \leq \frac{1}{1+\eta} |S_i|$

$\Rightarrow$ # iterations $\leq \log_{1+\eta} n \leq O(\frac{1}{\eta} \log n)$

Thm: $O(\frac{\varepsilon}{\eta} \log n)$-LEDP

Pf: In each iteration i, $\Delta d_i \leq 2$

$\Rightarrow$ Computing $\hat{d}_i$ is ε-DP

Everything else in iteration i is post-processing

$\Rightarrow$ each iteration is ε-DP

$\Rightarrow$ by basic composition, $O(\frac{\varepsilon}{\eta} \log n)$-DP.

Utility:

Lemma: $\Pr[|\hat{\rho}(S_i) - \rho(S_i)| \geq \Omega(\frac{1}{\varepsilon} \log n)] \leq \frac{1}{n^4}$

Pf: Bernstein's inequality for concentration of sum of subexponential random vars.

<u>Lemma</u>: fix iteration i. Let $v \in S_i$. Then if

$$d_{S_i}(v) \geq (1+\eta)\, \overset{=2\rho(S_i)}{d_{avg}(S_i)} + c\,\frac{1+\eta}{\varepsilon}\log n,$$

then $\Pr[v \in L_i] \leq \frac{1}{n^3}$

<u>Pf</u>: By previous lemma, $\overset{\text{w.p. } \geq}{1-\frac{1}{n^4}}$ $T_i \leq d_{avg}(S_i) + c'\frac{1}{\varepsilon}\log n$. for some c'

Let v such a node.

$\Rightarrow v \in L_i$ only if

$$d_{S_i}(v) + N_i(v) \leq T_i$$

$$\Leftrightarrow N_i(v) \leq T_i - d_{S_i}(v)$$

$$\leq 2(1+\eta)\hat{\rho}(S_i) \quad -(1+\eta)d_{avg}(S_i) - c\frac{1+\eta}{\varepsilon}\log n$$

$$\leq 2(1+\eta)\left(\rho(S_i) + c'\frac{1}{\varepsilon}\log n\right) - (1+\eta)d_{avg}(S_i) - c\frac{1+\eta}{\varepsilon}\log n$$

$$= 2(1+\eta)c'\frac{1}{\varepsilon}\log n - c\frac{1+\eta}{\varepsilon}\log n$$

Set $c = 8c'$

$$\Rightarrow N_i(v) \leq -6\left(\frac{1+\eta}{\varepsilon}\right)\log n \leq -4\frac{1}{\varepsilon}\log n$$

$$\Pr[N_i(v) \leq -4\tfrac{1}{\varepsilon}\log n] \leq e^{-4\log n} \leq \frac{1}{n^4}$$

$$\Rightarrow \text{total } \Pr[v \in L_i] \leq \frac{1}{n^4} + \frac{1}{n^4} \leq \frac{1}{n^3}$$

Thm: With prob. $\geq 1 - \frac{2}{n}$,

$$\rho(ALG(G)) \geq \frac{\rho(OPT(G))}{2(1+\eta)} - O\left(\frac{1}{\varepsilon}\log n\right)$$

Pf: Know from previous lemma that when each node removed, has deg. $\leq (1+\eta)\, d_{avg}(S_i) + \Lambda$ ($\Lambda = O(\frac{1}{\varepsilon}\log n)$)

(w.p. $\geq 1 - O\left(\frac{\log n}{n^3}\right) \geq 1 - \frac{1}{n}$)

Orient each edge towards whichever endpoint removed earlier.

$\Rightarrow$ each node has $\leq (1+\eta)\, d_{avg}(S_i) + \Lambda$ edges pointing to it

i = iteration node removed.

Lemma: $\rho(OPT(G)) \leq \min_{\text{orientations}} \max_v (\text{indegree}(v))$

$$\Rightarrow \rho(OPT(G)) \leq \max_i \left((1+\eta)\, d_{avg}(S_i) + \Lambda\right)$$
$$\leq 2(1+\eta) \max_i \rho(S_i) + \Lambda$$

So one of the S_i's good.

But $|\hat{\rho}(S_i) - \rho(S_i)| \leq \Lambda$ w.p. $\geq 1 - \frac{1}{n}$ $(\forall i)$!

$$\Rightarrow \rho(OPT(G)) \leq 2(1+\eta) \max_i \hat{\rho}(S_i) + O\left(\frac{1}{\varepsilon}\log n\right) \checkmark$$

Thm: Let $\varepsilon, \eta > 0$. There is an ε-LEDP alg. which returns $S \subseteq V$ s.t.

$$\rho(OPT(G)) \leq 2(1+\eta)\rho(S) + \frac{1}{\varepsilon\eta}\log^2 n$$

Pf: Use $\varepsilon' = \frac{\varepsilon\eta}{\log n}$

State of the art:

ε-DP: $(1+\eta, O(\frac{1}{\varepsilon}\log^4 n))$

ε-LEDP: $(2, O(\frac{1}{\varepsilon}\log n))$

(ε,δ)-DP: $(1, O(\frac{1}{\varepsilon}\sqrt{\log n \log \frac{n}{\delta}}))$

(ε,δ)-LEDP: $(1, O(\frac{1}{\varepsilon}\log n \sqrt{\log \frac{1}{\delta}}))$