Efficiently Estimating Erdős-Rényi Graphs with Differential Privacy

Adam Sealfon, MIT and UC Berkeley
Jonathan Ullman, Northeastern University

1 – DP FOR GRAPHS

Differential privacy (DP) allows the release of aggregate statistics from a dataset while hiding individual entries. Graphs model entities and relationships: undirected, simple graph $G = (V, E)$ on $n$ nodes, $m$ edges. Two models of DP for graphs: edge and node.

Most graph statistics are highly sensitive to arbitrary changes of a single node.

4 – LIPSCHITZ EXTENSIONS

A graph statistic $f: G \rightarrow R$ is everywhere differentially private if there is an extension $\hat{f}: \mathcal{G} \rightarrow R$ satisfying:

- $f(G) = \hat{f}(G')$ for nice graphs in $\mathcal{G}$
- $\hat{g}_{LIP}(G) \leq \hat{g}_{LIP}(G')$

where $\hat{g}_{LIP}(G)$ is the Lipschitz extension of $g(G)$.

Our Work (Relaxed Lipschitz Extension):

A new graph statistic $f: G \rightarrow R$ is everywhere differentially private if there is an extension $\hat{f}: \mathcal{G} \rightarrow R$ satisfying:

- $f(G) = \hat{f}(G')$ for nice graphs in $\mathcal{G}$
- $\hat{g}_{SmallSensitivity} = O(\hat{g}_{LIP}(G)) [NRS'07, M'34]$

We give an explicit polynomial time algorithm for computing $\hat{f}$.

2 – COUNTING EDGES IN NICE GRAPHS

Our work: node-DP estimators for the edge density $p(G) = m/n$ in nice graphs.

Baseline

- Any Graph
  - global sensitivity is $\Omega(1/n)$
  - error is $\Theta(\sqrt{m/n})$

Max Degree $D$ [KNRS'13]

- restricted sensitivity is $O(\sqrt{m/D})$
- error is $\Theta(\sqrt{m/n})$

Improvements for nice graphs

Random Graphs $G(n, p)$ [BCSZ'18]

- exponential time algorithm with error $\Theta(\sqrt{m/n})$

Our graph model

Efficiently Estimating $G(n, p)$, for nice graphs $G$

• 3 – RESULTS: NEW EFFICIENT ESTIMATORS

Theorem: a poly time $\epsilon$-node-DP algorithm for computing edge density in $k^\epsilon$-concentrated graphs with error $O\left((\sqrt{\epsilon}/\mathbb{E}_v|d_v| + \sqrt[4]{m/n})\right)$

Optimality: any $\epsilon$-node-DP algorithm for computing edge density in $k^\epsilon$-concentrated graphs must have error $\Omega\left((\sqrt{\epsilon}/\mathbb{E}_v|d_v| + \sqrt[4]{m/n})\right)$

Application: there is a poly time $\epsilon$-node-DP algorithm for estimating $G(n, p)$ with error $\Theta\left((\sqrt{\epsilon}/n + \sqrt{m/n})\right)$

Privacy for free when $\epsilon$ is not too small!

5 – OUR ESTIMATOR

Step 1: assign a weight $w_{T_L}(v) \in [0, 1]$ to each node based on how “typical” its degree is.

- $S_{LIP} = \{v \in V : |\deg(v) - \bar{d}_G| > \epsilon\}$
- $k = \min(k, |S_{LIP}|)$
- $\bar{d}_G = \frac{1}{n} \sum_{v \in V} \deg(v)$

Step 2: For edges incident on low-weight nodes, replace each edge with the average edge density.

- $w_{T_L}(u, v) = \min(w_{T_L}(u), w_{T_L}(v))$
- $w_{T_L}(u, v) = w_{T_L}(e) \mathbb{E}_{e \in E(G')} + (1 - w_{T_L}(e))p_G$

Lemma: $f(G) = |E|$ for $k^\epsilon$-concentrated graphs

Smooth Sensitivity Algorithm:

- $\left|\frac{d}{dz}f(G)\right| < \frac{\epsilon}{\mathbb{E}_v|d_v|}$

where $Z$ is sampled from a Student’s $t$-distribution with 3 d.f. $[NRS'07]$