Efficiently Estimating Erdős-Rényi Graphs with Differential Privacy

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

1 – DP FOR GRAPHS

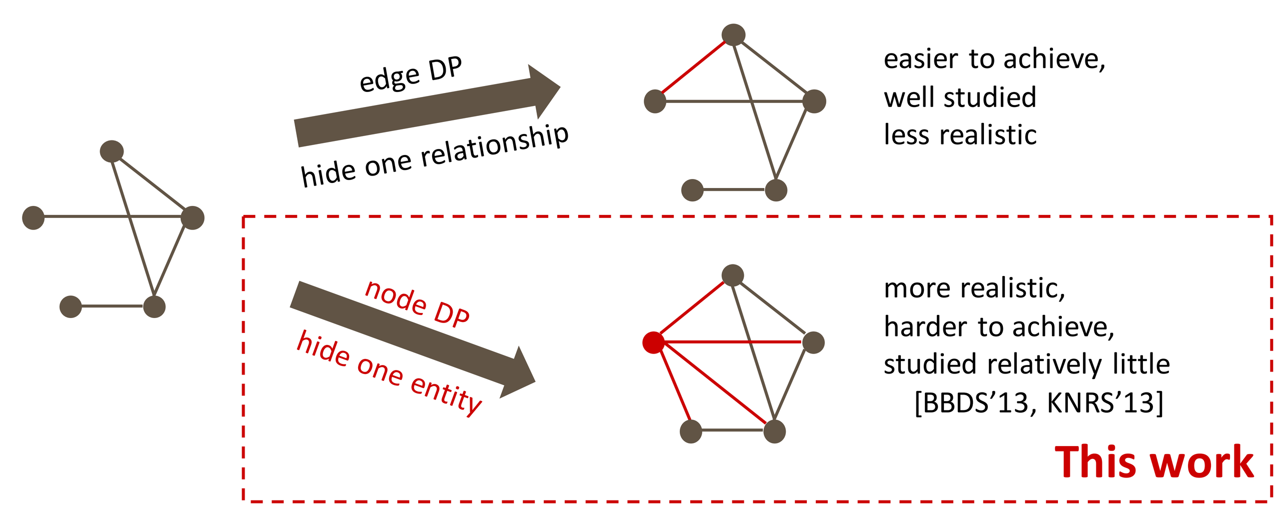
Differential privacy (DP) allows the release of aggregate statistics from a dataset while hiding individual entries

Graphs model entities and relationships

undirected, simple graph $G = (V, E)$

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

2 – COUNTING EDGES IN NICE GRAPHS

Our work: node-DP estimators for the edge density $p(G) = m/\binom{n}{2}$ in nice graphs

Baseline

Any Graph
global sensitivity is $^2/_n$
error is $\Theta(^1/_{\varepsilon n})$

Improvements for *nice graphs*

Based on Lipschitz extensions

Max Degree D [KNRS'13]
restricted sensitivity is $O(^D/_{n^2})$
error is $\Theta(^D/_{\varepsilon n^2})$

Erdős-Rényi graph model

Random Graphs $G(n,p)$ [BCSZ'18]
exponential time algorithm with
error $\Theta\left(^{\sqrt{p}}/_n + ^{\sqrt{p}}/_{\varepsilon n^{3/2}}\right)$

compare

3 – RESULTS: NEW EFFICIENT ESTIMATORS

Theorem: a poly time ε-node-DP algorithm for computing edge density in *k^*-concentrated* graphs with error $\tilde{O}\left(^{k^*}/_{\varepsilon n^2} + ^1/_{\varepsilon^2 n^2}\right)$

all degrees lie in an interval of width $2k^*$

Optimality: any ε-node-DP algorithm for computing edge density in *k^*-concentrated* graphs must have error $\Omega\left(^{k^*}/_{\varepsilon n^2} + ^1/_{\varepsilon^2 n^2}\right)$

$G(n,p)$ graphs are $\sqrt{pn}$ concentrated whp

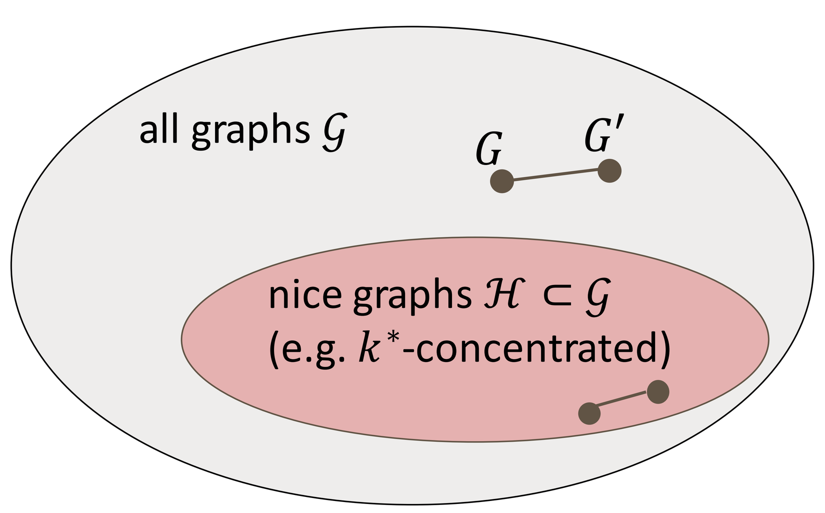
Application: there is a poly time ε-node-DP algorithm for estimating $G(n,p)$ graphs with error $\Theta\left(^{\sqrt{p}}/_n + ^{\sqrt{p}}/_{\varepsilon n^{3/2}} + ^1/_{\varepsilon^2 n^2}\right)$

Sampling error

Privacy overhead

Privacy for free when ε is not too small!

4 – LIPSCHITZ EXTENSIONS



graph statistic:
$f: \mathcal{G} \to \mathbb{R}$

global sensitivity:
$$GS_f = \max_{G \sim G' \in \mathcal{G}} f(G) - f(G')$$

restricted sensitivity:
$$RS_{f,\mathcal{H}} = \max_{G \sim G' \in \mathcal{H}} f(G) - f(G')$$

Often, $RS_{f,\mathcal{H}} \ll GS_f$

Want privacy for all graphs but noise calibrated to $RS_{f,\mathcal{H}}$ for nice graphs in $\mathcal{H}$

Lipschitz extension [BBDS'13,KNRS'13]:

A new graph statistic $\tilde{f}$ s.t.

- $\tilde{f}(G) = f(G)$ for nice graphs in $\mathcal{H}$
- $GS_{\tilde{f}} = RS_{f,\mathcal{H}}$

$\tilde{f}$ always exists [M'34], but can require exponential time to compute

Our Work (Relaxed Lipschitz Extension):

A new graph statistic $\tilde{f}$ s.t.

- $\tilde{f}(G) \approx f(G)$ for nice graphs in $\mathcal{H}$
- $\text{SmoothSens}_{\tilde{f}} = O(RS_{f,\mathcal{H}})$ [NRS'07]

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

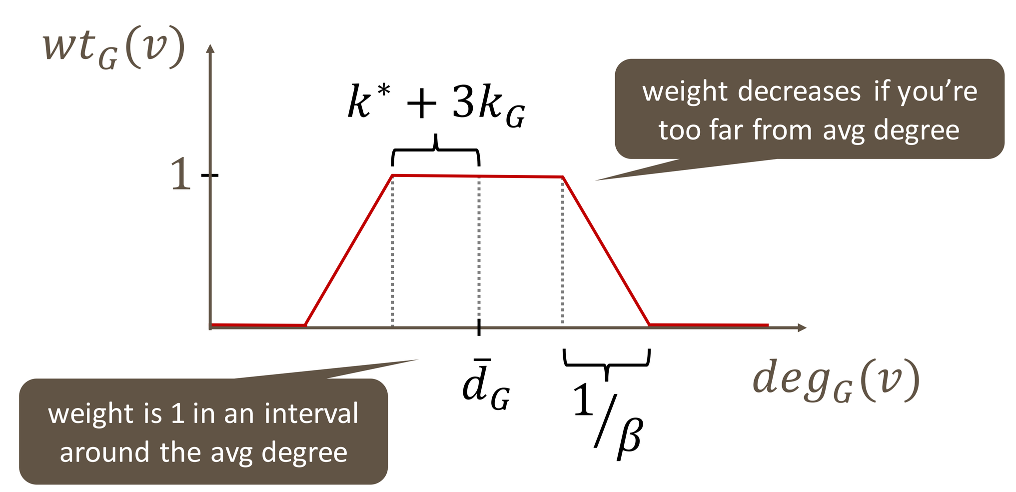
5 – OUR ESTIMATOR

Step 1: assign a weight $wt_G(v) \in [0,1]$ to each node based on how "typical" its degree is

$$S_{G,t} = \{v \in V : |\deg(v) - \bar{d}_G| > t\}$$

$G \in \mathcal{H}: |S_{G,k^*}| = k_G = 0$

$$k_G = \min\{k : |S_{G,3k+k^*}| \le k\}$$



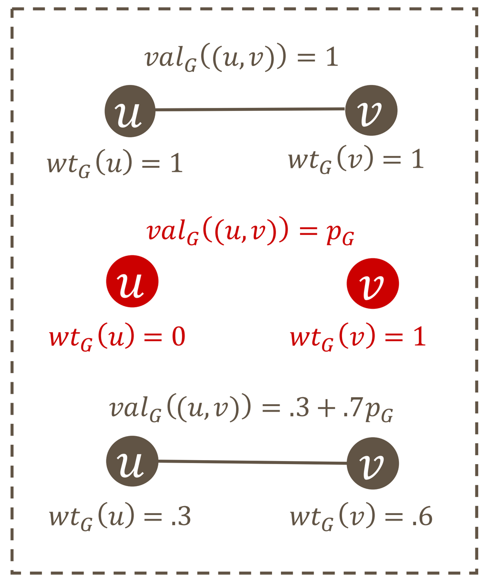
Step 2: For edges incident on low-weight nodes, replace each edge with the average edge density.

$$wt_G((u,v)) = \min(wt_G(u), wt_G(v))$$

$$val_G(e) = wt_G(e)\mathbb{I}_{e \in E} + (1 - wt_G(e))p_G$$

$$f(G) = \frac{1}{2}\sum_{u \ne v} val_G((u,v))$$

examples



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

Lemma: there is a poly-time computable, β-smooth upper bound on the local sensitivity of f satisfying
$$S(G) = O((k_G + k^*)(1 + \beta k_G) + 1/\beta)$$

Smooth Sensitivity Algorithm:
$\binom{n}{2}^{-1}\left(f(G) + \frac{S(G)}{\varepsilon} \cdot Z\right)$
where Z is sampled from a Student's t-distribution with 3 d.f. [NRS'07]

[BBDS'13] Blocki, Blum, Datta, Sheffet. *Differentially private data analysis of social networks via restricted sensitivity.* ITCS'13

[BCSZ'18] Borgs, Chayes, Smith, Zadik. *Revealing Network Structure, Confidentially: Improved Rates for Node-Private Graphon Estimation.* FOCS'18

[KNRS'13] Kasiviswanathan, Nissim, Raskhodnikova, Smith. *Analyzing Graphs with Node Differential Privacy.* TCC'13

[M'34] Edward McShare. *Extension of Range Functions.* Bulletin of the AMS 1934

[NRS'07] Nissim, Raskhodnikova, Smith. *Smooth Sensitivity and Sampling in Private Data Analysis.* STOC'07