Journal of Privacy and Confidentiality
Vol. 11 (1) 2021 TPDP 2019 Submitted Apr 2020
Published Jan 2021

EFFICIENTLY ESTIMATING ERDŐS-RÉNYI GRAPHS WITH NODE DIFFERENTIAL PRIVACY *

ADAM SEALFON AND JONATHAN ULLMAN

Department of Statistics, UC Berkeley
e-mail address: asealfon@berkeley.edu

Khoury College of Computer Sciences, Northeastern University
e-mail address: jullman@ccs.neu.edu

Abstract. We give a simple, computationally efficient, and node-differentially-private algorithm for estimating the parameter of an Erdős-Rényi graph—that is, estimating p in a $G(n, p)$. Our algorithm nearly matches the information-theoretically optimal accuracy of the exponential-time algorithm due to Borgs et al. (FOCS 2018). More generally, we give an optimal, computationally efficient, private algorithm for estimating the edge-density of any graph whose degree distribution is concentrated in a small interval.

1. Introduction

Network data modeling individuals and relationships between individuals are increasingly central in data science. However, while there is a highly successful literature on differentially private statistical estimation for traditional i.i.d. data, the literature on estimating network models is far less well developed.

Early work on private network data focused on *edge-differential-privacy*, in which the algorithm is required to "hide" the presence or absence of a single edge in the graph (see, e.g. Nissim, Raskhodnikova and Smith, 2007; Hay et al., 2009; Karwa et al., 2014; Gupta, Roth and Ullman, 2012; Blocki et al., 2012; Xiao, Chen and Tan, 2014; Karwa and Slavković, 2016, and many others). A more desirable notion of privacy is *node-differential privacy (node-DP)*, which requires the algorithm to hide the presence or absence of an arbitrary set of edges incident on a single node. Although node-DP is difficult to achieve without compromising accuracy, the beautiful works of Blocki et al. (2013) and Kasiviswanathan et al. (2013) showed how to design accurate node-DP estimators for many interesting graph statistics via *Lipschitz extensions.* However, many of the known constructions of Lipschitz extensions require exponential running time, and constructions of computationally efficient Lipschitz extensions (Raskhodnikova and Smith, 2016; Cummings and Durfee, 2018; Canonne et al., 2019) lag behind. As a result, even for estimating very simple graph models, there

Key words and phrases: Differential privacy, graph privacy, node differential privacy.
* A previous version of this work was published in NeurIPS 2019.

JPrivacy Confidentiality
www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.745

are large gaps in accuracy between the best known computationally efficient algorithms and the information theoretically optimal algorithms.

In this work we focus on what is arguably the simplest model of network data, the *Erdős-Rényi graph*. In this model, denoted $G(n,p)$, we are given a number of nodes n and a parameter $p \in [0,1]$, and we sample an n-node graph G by independently including each edge (i,j) for $1 \le i < j \le n$ with probability p. The goal is to design a node-DP algorithm that takes as input a graph $G \sim G(n,p)$ and outputs an estimate $\hat{p}$ of the edge density parameter p.

Surprisingly, until an elegant recent work of Borgs et al. (2018), the optimal accuracy for estimating the parameter p in a $G(n,p)$ via node-DP algorithms was unknown. Although that work essentially resolved the optimal accuracy of node-DP algorithms,[1] their construction is again based on generic Lipschitz extensions, and thus results in an exponential-time algorithm, and, in our opinion, gives little insight for how to construct an efficient estimator with similar accuracy.

The main contribution of this work is to give a simple, polynomial-time estimator for Erdős-Rényi graphs whose error very nearly matches that of Borgs et al.'s estimator, and indeed matches it in a wide range of parameters. We achieve this by giving a more general result, showing how to optimally estimate the edge-density of any graph whose degree distribution is concentrated in a small interval.

1.1. Background: Node-Private Algorithms for Erdős-Rényi Graphs.

Without privacy, the optimal estimator is simply to output the *edge-density* $p_G = |E|/\binom{n}{2}$ of the realized graph $G \sim G(n,p)$, which guarantees

$$\mathop{\mathbb{E}}_{G}\left[(p - p_G)^2\right] = \frac{p(1-p)}{\binom{n}{2}}.$$

The simplest way to achieve ε-node-DP is to add zero-mean noise to the edge-density with standard-deviation calibrated to its *global-sensitivity*, which is the amount that changing the neighborhood of a single node in a graph can change its edge-density. The global sensitivity of p_G is $\Theta(1/n)$, and thus the resulting private algorithm $\mathcal{A}_{naïve}$ satisfies

$$\mathop{\mathbb{E}}_{G}\left[(p - \mathcal{A}_{naïve}(G))^2\right] = \Theta\left(\frac{1}{\varepsilon^2 n^2}\right)$$

Note that this error is *at least* on the same order as the non-private error, and can asymptotically dominate the non-private error.

Borgs et al. (2018) gave an improved ε-node-DP algorithm such that, when both p and ε are $\gtrsim \log(n)/n$,

$$\mathbb{E}\left[(p - \mathcal{A}_{bcsz}(G))^2\right] = \underbrace{\frac{p(1-p)}{\binom{n}{2}}}_{\text{non-private error}} + \underbrace{\tilde{O}\left(\frac{p}{\varepsilon^2 n^3}\right)}_{\text{overhead due to privacy}}$$

What is remarkable about their algorithm is that, unless ε is quite small (roughly $\varepsilon \lesssim n^{-1/2}$), the first term dominates the error, in which case privacy comes essentially *for free*. That is, the error of the private algorithm is only larger than that of the optimal non-private

[1] We note that the lower bound shown in Borgs et al. (2018) is actually for the uniform random graph model $G(n,m)$ with a fixed number of edges m, rather than for $G(n,p)$.

algorithm by a $1 + o(1)$ factor. However, as we discussed above, this algorithm is not computationally efficient.

The only computationally efficient node-DP algorithms for computing the edge-density apply to graphs with small maximum degree (Blocki et al., 2013; Kasiviswanathan et al., 2013; Raskhodnikova and Smith, 2016), and thus do not give optimal estimators for Erdős-Rényi graphs unless p is very small.

1.2. **Our Results.** Our main result is a computationally efficient estimator for Erdős-Rényi graphs.

Theorem 1.1 (Erdős-Rényi Graphs, Informal). *There is an $O(n^2)$-time ε-node-DP algorithm $\mathcal{A}$ such that for every n and every $p \gtrsim 1/n$ if $G \sim G(n, p)$ then*

$$\mathop{\mathbb{E}}_{G,\mathcal{A}}\left[(p - \mathcal{A}(G))^2\right] = \underbrace{\frac{p(1-p)}{\binom{n}{2}}}_{\text{non-private error}} + \underbrace{\tilde{O}\left(\frac{p}{\varepsilon^2 n^3} + \frac{1}{\varepsilon^4 n^4}\right)}_{\text{overhead due to privacy}}$$

The error of Theorem 1.1 matches that of the exponential-time estimator of Borgs et al. (2018) up to the additive $\tilde{O}(1/\varepsilon^4 n^4)$ term, which is often not the dominant term in the overall error. In particular, the error of our estimator is still within a $1 + o(1)$ factor of the optimal non-private error unless ε or p is quite small—for example, when p is a constant and $\varepsilon \gtrsim n^{-1/2}$.

Our estimator actually approximates the edge density for a significantly more general class of graphs than merely Erdős-Rényi graphs. Specifically, Theorem 1.1 follows from a more general result for the family of *concentrated-degree graphs*. For $k \in \mathbb{N}$, define $\mathcal{G}_{n,k}$ to be the set of n-node graphs such that the degree of every node is between $\bar{d} - k$ and $\bar{d} + k$, where $\bar{d} = 2|E|/n$ is the average degree of the graph.

Theorem 1.2 (Concentrated-Degree Graphs, Informal). *For every $k \in \mathbb{N}$, there is an $O(n^2)$-time ε-node-DP algorithm $\mathcal{A}$ such that for every n and every $G \in \mathcal{G}_{n,k}$,*

$$\mathop{\mathbb{E}}_{\mathcal{A}}\left[(p_G - \mathcal{A}(G))^2\right] = O\left(\frac{k^2}{\varepsilon^2 n^4} + \frac{1}{\varepsilon^4 n^4}\right)$$

where $p_G = |E|/\binom{n}{2}$ is the empirical edge density of G.

Theorem 1.1 follows from Theorem 1.2 by using the fact that for an Erdős-Rényi graph, with overwhelming probability the degree of every node lies in an interval of width $\tilde{O}(\sqrt{pn})$ around the average degree.

The main technical ingredient in Theorem 1.2 is to construct a *low sensitivity* estimator $f(G)$ for the number of edges. The first property we need is that when G satisfies the concentrated degrees property, $f(G)$ equals the number of edges in G. The second property of the estimator we construct is that its *smooth sensitivity* (Nissim, Raskhodnikova and Smith, 2007) is low on these graphs G. At a high level, the smooth sensitivity of f at a graph G is the most that changing the neighborhood of a small number of nodes in G can change the value of $f(G)$. Once we have this property, it is sufficient to add noise to $f(G)$ calibrated to its smooth sensitivity. We construct f by carefully reweighting edges that are incident on nodes that do not satisfy the concentrated-degree condition.

Finally, we are able to show that Theorem 1.2 is optimal for concentrated-degree graphs. In additional to being a natural class of graphs in its own right, this lower bound demonstrates that in order to improve Theorem 1.1 we will need techniques that are more specialized to Erdős-Rényi graphs.

Theorem 1.3 (Lower Bound, Informal)**.** *For every n and k, and every ε-node-DP algorithm $\mathcal{A}$, there is some $G \in \mathcal{G}_{n,k}$ such that*

$$\mathbb{E}_{A}\left[(p_G - \mathcal{A}(G))^2\right] = \Omega\left(\frac{k^2}{\varepsilon^2 n^4} + \frac{1}{\varepsilon^4 n^4}\right)$$

The same bound applies to (ε, δ)-node-DP algorithms with sufficiently small $\delta \lesssim \varepsilon$.

2. Preliminaries

Let $\mathcal{G}_n$ be the set of n-node graphs. We say that two graphs $G, G' \in \mathcal{G}_n$ are *node-adjacent*, denoted $G \sim G'$, if G' can be obtained by G modifying the neighborhood of a single node i. That is, there exists a single node i such that for every edge e in the symmetric difference of G and G', e is incident on i. As is standard in the literature on differential privacy, we treat n as a fixed quantity and define adjacency only for graphs with the same number of nodes. We could easily extend our definition of adjacency to include adding or deleting a single node itself.

Definition 2.1 (Differential Privacy (Dwork et al., 2006))**.** A randomized algorithm $\mathcal{A}\colon \mathcal{G}_n \to \mathcal{R}$ is (ε, δ)-node-differentially private if for every $G \sim G' \in \mathcal{G}_n$ and every $R \subseteq \mathcal{R}$,

$$\mathbb{P}\left[\mathcal{A}(G) \in R\right] \leq e^{\varepsilon} \cdot \mathbb{P}\left[\mathcal{A}(G') \in R\right] + \delta$$

If $\delta = 0$ we will simply say that $\mathcal{A}$ is ε-node-differentially private. As we only consider node differential privacy in this work, we will frequently simply say that $\mathcal{A}$ satisfies differential privacy.

The next lemma is the basic composition property of differential privacy.

Lemma 2.2 (Composition (Dwork et al., 2006))**.** *If $\mathcal{A}_1, \mathcal{A}_2\colon \mathcal{G}_n \to \mathcal{R}$ are each (ε, δ)-node-differentially private algorithms, then the mechanism $\mathcal{A}(G) = (\mathcal{A}_1(G), \mathcal{A}_2(G))$ satisfies $(2\varepsilon, 2\delta)$-node-differential privacy. The same holds if $\mathcal{A}_2$ may depend on the output of $\mathcal{A}_1$.*

We say that two graphs G, G' are at *node distance c* if there exists a sequence of graphs

$$G = G_0 \sim G_1 \ldots G_{c-1} \sim \ldots G_c = G'$$

The standard *group privacy* property of differential privacy yields the following guarantees for graphs at node distance $c > 1$.

Lemma 2.3 (Group Privacy (Dwork et al., 2006))**.** *If $\mathcal{A}\colon \mathcal{G}_n \to \mathcal{R}$ is (ε, δ)-node-differentially-private and G, G' are at node-distance c then for every $R \subseteq \mathcal{R}$,*

$$\mathbb{P}\left[\mathcal{A}(G) \in R\right] \leq e^{c\varepsilon} \mathbb{P}\left[\mathcal{A}(G') \in R\right] + c e^{c\varepsilon} \delta$$

Sensitivity and Basic DP Mechanisms. The main differentially private primitive we will use is *smooth sensitivity* (Nissim, Raskhodnikova and Smith, 2007). Let $f\colon \mathcal{G}_n \to \mathbb{R}$ be a real-valued function. For a graph $G \in \mathcal{G}_n$, we can define the *local sensitivity* of f at G to be

$$LS_f(G) = \max_{G' : G' \sim G} |f(G) - f(G')|$$

and the *global sensitivity* of f to be

$$GS_f = \max_G LS_f(G) = \max_{G,G':G'\sim G} |f(G) - f(G')|$$

A basic result in differential privacy says that we can achieve privacy for any real-valued function f by adding noise calibrated to the global sensitivity of f.

Theorem 2.4 (DP via Global Sensitivity (Dwork et al., 2006))**.** *Let $f : \mathcal{G}_n \to \mathbb{R}$ be any function. Then the algorithm*

$$\mathcal{A}(G) = f(G) + \frac{GS_f}{\varepsilon} \cdot Z,$$

where Z is sampled from a standard Laplace distribution, satisfies $(\varepsilon, 0)$-differential privacy.[2] *Moreover, this mechanism satisfies $\mathbb{E}_{\mathcal{A}}\left[(\mathcal{A}(G) - f(G))^2\right] = O(GS_f/\varepsilon)$, and for all $t > 0$ we have that*

$$\mathbb{P}_{\mathcal{A}}\left[|\mathcal{A}(G) - f(G)| \geq t \cdot GS_f/\varepsilon\right] \leq \exp(-t).$$

In many cases the global sensitivity of f is too high, and we want to use a more refined mechanism that adds instance-dependent noise that is more comparable to the local sensitivity. This can be achieved via the *smooth sensitivity* framework of Nissim, Raskhodnikova and Smith (2007).

Definition 2.5 (Smooth Upper Bound (Nissim, Raskhodnikova and Smith, 2007))**.** Let $f : \mathcal{G}_n \to \mathbb{R}$ be a real-valued function and $\beta > 0$ be a parameter. A function $S : \mathcal{G}_n \to \mathbb{R}$ is a *β-smooth upper bound on LS_f* if

(1) for all $G \in \mathcal{G}_n$, $S(G) \geq LS_f(G)$, and
(2) for all neighboring $G \sim G' \in \mathcal{G}_n$, $S(G) \leq e^{\beta} \cdot S(G')$.

The key result in smooth sensitivity is that we can achieve differential privacy by adding noise to $f(G)$ proportional to any smooth upper bound $S(G)$.

Theorem 2.6 (DP via Smooth Sensitivity (Nissim, Raskhodnikova and Smith, 2007; Bun and Steinke, 2019))**.** *Let $f : \mathcal{G}_n \to \mathbb{R}$ be any function and S be a β-smooth upper bound on the local sensitivity of f for any $\beta \leq \varepsilon$. Then the algorithm*

$$\mathcal{A}(G) = f(G) + \frac{S(G)}{\varepsilon} \cdot Z,$$

where Z is sampled from a Student's t-distribution with 3 degrees of freedom, satisfies $(O(\varepsilon), 0)$-differential privacy.[3] *Moreover, for any $G \in \mathcal{G}_n$, this algorithm satisfies*

$$\mathbb{E}_{\mathcal{A}}\left[(\mathcal{A}(G) - f(G))^2\right] = O(S(G)^2/\varepsilon^2).$$

3. An Estimator for Concentrated-Degree Graphs

In this section we describe and analyze a node-differentially-private estimator for the edge density of a concentrated-degree graph.

[2]The standard Laplace distribution Z has $\mathbb{E}[Z] = 0, \mathbb{E}\left[Z^2\right] = 2$, and density $\mu(z) \propto e^{-|z|}$.

[3]The Student's t-distribution with 3 degrees of freedom can be efficiently sampled by choosing $X, Y_1, Y_2, Y_3 \sim \mathcal{N}(0,1)$ independently from a standard normal and returning $Z = X/\sqrt{Y_1^2 + Y_2^2 + Y_3^2}$. This distribution has $\mathbb{E}[Z] = 0$ and $\mathbb{E}\left[Z^2\right] = 3$, and its density is $\mu(z) \propto 1/(1+z^2)^2$.

3.1. **The Estimator.** In order to describe the estimator we introduce some key notation. The input to the estimator is a graph $G = (V, E)$ and a parameter k^*. Intuitively, k^* should be an upper bound on the concentration parameter of the graph, although we obtain more general results when k^* is not an upper bound, in case the user does not have an *a priori* upper bound on this quantity.

Our estimator will weight vertices according to the typicality of their degrees, where vertices close to the average degree will have weight 1 and vertices far from the average degree will have weight 0. For a graph $G = (V, E)$, let $p_G = |E|/\binom{n}{2}$ be the empirical edge density of G, and let $\bar{d}_G = (n-1)p_G$ be the empirical average degree of G. Let k_G be the smallest positive integer value such that at most k_G vertices of G have degrees differing from $\bar{d}_G$ by more than $k' := k^* + 3k_G$. Define $I_G = [\bar{d}_G - k', \bar{d}_G + k']$. For each vertex $v \in V$, let $t_v = \min\{|t| : [\deg_G(v) \pm t] \cap I_G \neq \emptyset\}$ be the distance between $\deg_G(v)$ and the interval I_G, and define the *weight* $\mathsf{wt}_G(v)$ of v as follows. For a parameter $\beta > 0$ to be specified later, let

$$\mathsf{wt}_G(v) = \begin{cases} 1 & \text{if } t_v = 0 \\ 1 - \beta t_v & \text{if } t_v \in (0, 1/\beta] \\ 0 & \text{otherwise.} \end{cases}$$

That is, $\mathsf{wt}_G(v) = \max(0, 1 - \beta t_v)$. For each pair of vertices $e = \{u, v\}$, define the *weight* $\mathsf{wt}_G(e)$ and *value* $\mathsf{val}_G(e)$ as follows. Let

$$\mathsf{wt}_G(e) = \min(\mathsf{wt}_G(u), \mathsf{wt}_G(v))$$

and let

$$\mathsf{val}_G(e) = \mathsf{wt}_G(e) \cdot x_e + (1 - \mathsf{wt}_G(e)) \cdot p_G$$

where x_e denotes the indicator variable on whether $e \in E$. As above, define the function f to be the total value of all pairs of vertices in the graph,

$$f(G) = \sum_{u,v \in V} \mathsf{val}_G(\{u, v\}),$$

where the sum is over unordered pairs of distinct vertices.

Once we construct this function f, we add noise to f proportional to a β-smooth upper bound on the sensitivity of f, which we derive in this section. Pseudocode for our estimator is given in Algorithm 1.

3.2. **Analysis using Smooth Sensitivity.** We begin by bounding the local sensitivity $LS_f(G)$ of the function f defined above.

Lemma 3.1 . $LS_f(G) = O((k_G + k^*)(1 + \beta k_G) + \frac{1}{\beta})$.

Proof. Consider any pair of graphs G, G' differing in only a single vertex v^*, and note that the empirical edge densities p_G and $p_{G'}$ can differ by at most $\frac{2}{n} < \frac{2}{n-1}$, so $\bar{d}_G$ and $\bar{d}_{G'}$ can differ by at most 2. Moreover, for any vertex $v \neq v^*$, the degree of v can differ by at most 1 between G and G'. Consequently, by the Triangle Inequality, for any $v \neq v^*$, $|\bar{d}_G - \deg_G(v)|$ can differ from $|\bar{d}_{G'} - \deg_{G'}(v)|$ by at most 3 and $\mathsf{wt}_G(v)$ can differ from $\mathsf{wt}_{G'}(v)$ by at most 3β. It follows from the former statement that k_G and $k_{G'}$ differ by at most 1.

Let Far_G denote the set of at most k_G vertices whose degree differs from $\bar{d}_G$ by more than $k' = k^* + 3k_G$. For any vertices $u, v \notin \mathsf{Far}_G \cup \mathsf{Far}_{G'} \cup \{v^*\}$, we have that $\mathsf{wt}_G(\{u, v\}) = \mathsf{wt}_{G'}(\{u, v\}) = 1$, and so $\mathsf{val}_G(\{u, v\}) = \mathsf{val}_{G'}(\{u, v\})$, since the edge $\{u, v\}$ is present in G if and only if it is present in G'.

Algorithm 1: Estimating the edge density of a concentrated-degree graph.

Input: A graph $G \in \mathcal{G}_n$ and parameters $\varepsilon > 0$ and $k^* \geq 0$.
Output: A parameter $0 \leq \hat{p} \leq 1$.

1 Let $p_G = \frac{1}{\binom{n}{2}} \sum_e x_e$ and $\bar{d}_G = (n-1)p_G$.

2 Let $\beta = \min(\varepsilon, 1/\sqrt{k^*})$.

3 Let $k_G > 0$ be the smallest positive integer such that at most k_G vertices have degree outside $[\bar{d}_G - k^* - 3k_G, \bar{d}_G + k^* + 3k_G]$.

4 For $v \in V$, let $t_v = \min\{|t| : \deg_G(v) \pm t \in [\bar{d}_G - k^* - 3k_G, \bar{d}_G + k^* + 3k_G]\}$ and let $\mathsf{wt}_G(v) = \max(0, 1 - \beta t_v)$.

5 For each $u, v \in V$, let $\mathsf{wt}_G(\{u,v\}) = \min(\mathsf{wt}_G(u), \mathsf{wt}_G(v))$ and let $\mathsf{val}_G(e) = \mathsf{wt}_G(e) \cdot x_e + (1 - \mathsf{wt}_G(e))p_G$.

6 Let $f(G) = \sum_{u \neq v} \mathsf{val}_G(\{u,v\})$, where the sum is over unordered pairs of vertices.

7 Let $s = \max_{\ell \geq 0} Ce^{-\beta\ell} \cdot (k_G + \ell + k^* + \beta(k_G + \ell)(k_G + \ell + k^*) + 1/\beta)$, where C is the constant implied by Lemma 3.1.

8 Return $\frac{1}{\binom{n}{2}} \cdot (f(G) + (s/\varepsilon) \cdot Z)$, where Z is sampled from a Student's t-distribution with three degrees of freedom.

Now consider edges $\{u,v\}$ such that $u, v \neq v^*$ but $u \in \mathsf{Far}_G \cup \mathsf{Far}_{G'}$ (and v may or may not be as well). If $\deg_G(u) \notin [\bar{d}_G - k'', \bar{d}_G + k'']$ for $k'' = k' + 1/\beta + 3$, then $\mathsf{wt}_G(u) = \mathsf{wt}_{G'}(u) = 0$ and so $|\mathsf{val}_G(\{u,v\}) - \mathsf{val}_{G'}(\{u,v\})| = |p_G - p_{G'}| \leq 2/n$. Otherwise, $\deg_G(u) \in [\bar{d}_G - k'', \bar{d}_G + k'']$. We can break up the sum

$$f_u(G) := \sum_{v \neq u} \mathsf{val}_G(\{u,v\}) = \sum_{v \neq u} \mathsf{wt}_G(\{u,v\}) \cdot x_{\{u,v\}} + \sum_{v \neq u} (1 - \mathsf{wt}_G(\{u,v\}))p_G.$$

Since at most k_G other vertices can have weight less than the weight of u, we can bound the first term by

$$\left| \deg_G(u)\mathsf{wt}_G(u) - \sum_{v \neq u} \mathsf{wt}_G(\{u,v\}) \cdot x_{\{u,v\}} \right| \leq k_G \mathsf{wt}_G(u)$$

and the second term by

$$\left| \bar{d}_G - \bar{d}_G \mathsf{wt}_G(u) - \sum_{v \neq u} (1 - \mathsf{wt}_G(\{u,v\}))p_G \right| \leq p_G k_G \mathsf{wt}_G(u).$$

so the total sum is bounded by

$$\left| \bar{d}_G + (\deg_G(u) - \bar{d}_G)\mathsf{wt}_G(u) - f_u(G) \right| \leq 2k_G \mathsf{wt}_G(u).$$

Since $|\mathsf{wt}_G(u) - \mathsf{wt}_{G'}(u)| \le 3\beta$, it follows that

$$|f_u(G) - f_u(G')| \le 7 + 3\beta(k'' + 3) + 9\beta + 6\beta k_G = O(1 + \beta(k_G + k^*)).$$

Since there are at most $k_G + k'_G \le 2k_G + 1$ vertices in $u \in \mathsf{Far}_G \cup \mathsf{Far}_{G'} \setminus \{v^*\}$, the total difference in the terms of $f(G)$ and $f(G')$ corresponding to such vertices is at most $O(k_G + \beta k_G(k_G + k^*))$. However, we are double-counting any edges between two vertices in $u \in \mathsf{Far}_G \cup \mathsf{Far}_{G'}$; the number of such edges is $O(k_G^2)$, and for any such edge e, $|\mathsf{val}_G(e) - \mathsf{val}_{G'}(e)| \pm O(\beta)$. Consequently the error induced by this double-counting is at most $O(\beta k_G^2)$, so the total difference between the terms of $f(G)$ and $f(G')$ corresponding to such vertices is still $O(k_G + \beta k_G(k_G + k^*))$.

Finally, consider the edges $\{u, v^*\}$ involving vertex v^*. If $\mathsf{wt}_G(v^*) = 0$ then

$$f_{v^*}(G) = \sum_{u \ne v^*} \mathsf{val}_G(\{u, v^*\}) = (n-1)p_G = \bar{d}_G.$$

If $\mathsf{wt}_G(v^*) = 1$ then $\deg_G(v^*) \in [\bar{d}_G - k', \bar{d}_G + k']$, so $|f_{v^*}(G) - \deg_G(v^*)| \le k_G$ and

$$|f_{v^*}(G) - \bar{d}_G| \le k' + k_G.$$

Otherwise, $\deg_G(v^*) \in [\bar{d}_G - k' - 1/\beta, \bar{d}_G + k' + 1/\beta]$. Then we have that

$$\left| f_{v^*}(G) - \left[\bar{d}_G + (\deg_G(v^*) - \bar{d}_G)\mathsf{wt}_G(v^*)\right]\right| \le 2k_G \mathsf{wt}_G(v^*),$$

so

$$\left| f_{v^*}(G) - \bar{d}_G \right| \le \left| \deg_G(v^*) - \bar{d}_G \right| + 2k_G.$$

Therefore, in either case we have that $f_{v^*}(G) \in [\bar{d}_G - O(k_G + k^* + 1/\beta), \bar{d}_G + O(k_G + k^* + 1/\beta)]$, and so $|f_{v^*}(G) - f_{v^*}(G')| \le O(k_G + k^* + 1/\beta)$.

Putting everything together, we have that $LS_f(G) = O((k_G + k^*)(1 + \beta k_G) + 1/\beta)$. □

We now compute a smooth upper bound on $LS_f(G)$. From the proof of Lemma 3.1, we have that there exists some constant $C > 0$ such that $LS_f(G) \le C((k_G + k^*)(1 + \beta k_G) + \frac{1}{\beta})$. Let

$$g(k_G, k^*, \beta) = C((k_G + k^*)(1 + \beta k_G) + \tfrac{1}{\beta})$$

be this upper bound on $LS_f(G)$, and let

$$S(G) = \max_{\ell \ge 0} e^{-\ell\beta} g(k_G + \ell, k^*, \beta).$$

Lemma 3.2 . *$S(G)$ is a β-smooth upper bound on the local sensitivity of f. Moreover,*

$$S(G) = O((k_G + k^*)(1 + \beta k_G) + \tfrac{1}{\beta}).$$

Proof. For neighboring graphs G, G', we have that

$$\begin{aligned} S(G') &= \max_{\ell \ge 0} e^{-\ell\beta} g(k_{G'} + \ell, k^*, \beta) \\ &\le \max_{\ell \ge 0} e^{-\ell\beta} g(k_G + \ell + 1, k^*, \beta) \\ &= e^{\beta} \max_{\ell \ge 1} e^{-\ell\beta} g(k_G + \ell, k^*, \beta) \\ &\le e^{\beta} \max_{\ell \ge 0} e^{-\ell\beta} g(k_G + \ell, k^*, \beta) \\ &= e^{\beta} S(G). \end{aligned}$$

Moreover, for fixed k_G, k^*, β, consider the function $h(\ell) = e^{-\ell\beta} g(k_G + \ell, k^*\beta)$, and consider the derivative $h'(\ell)$. We have that

$$h'(\ell) = C\beta e^{-\ell\beta}(k_G + \ell)(1 - \beta(k_G + \ell + k^*)).$$

Consequently the only possible local maximum for $\ell > 0$ would occur for $\ell = 1/\beta - k_G - k^*$; note that the function h decreases as $\ell \to \infty$. Consequently the maximum value of h occurs for some $\ell \leq 1/\beta$, and so

$$\begin{aligned} S(G) &= \max_{\ell \geq 0} h(\ell) \\ &= \max_{\ell \geq 0} Ce^{-\ell\beta}(k_G + \ell + k^* + (k_G + \ell)(k_G + \ell + k^*)\beta + 1/\beta) \\ &\leq C \cdot (k_G + 1/\beta + k^* + (k_G + 1/\beta)(k_G + 1/\beta + k^*)\beta + 1/\beta) \\ &= C \cdot (3k_G + 2k^* + \beta k_G(k_G + k^*) + 3/\beta) \\ &= O((k_G + k^*)(1 + \beta k_G) + 1/\beta) \end{aligned}$$

as desired. □

Theorem 3.3 . *Algorithm 1 is $(O(\varepsilon), 0)$-differentially private. Moreover, for any k-concentrated n-vertex graph $G = (V, E) \in \mathcal{G}_{n,k}$ with $k \geq 1$, we have that*

$$\mathop{\mathbb{E}}_{\mathcal{A}}\left[\left(\frac{|E|}{\binom{n}{2}} - \mathcal{A}_{\varepsilon,k}(G)\right)^2\right] = O\left(\frac{k^2}{\varepsilon^2 n^4} + \frac{1}{\varepsilon^4 n^4}\right),$$

where $\mathcal{A}_{\varepsilon,k}(G)$ denotes Algorithm 1 on input (G, ε, k).

Proof. Algorithm 1 computes function f and releases it with noise proportional to a β-smooth upper bound on the local sensitivity for $\beta \leq \varepsilon$. Consequently $(O(\varepsilon), 0)$-differential privacy follows immediately from Theorem 2.6.

We now analyze its accuracy on k-concentrated graphs G. If G is k-concentrated and $k^* \geq k$, then $\mathsf{wt}_G(v) = 1$ for all vertices $v \in V$ and $\mathsf{val}_G(\{u, v\}) = x_{\{u,v\}}$ for all $u, v \in V$, and so $f(G) = |E|$. Consequently Algorithm 1 computes the edge density of a k-concentrated graph with noise distributed according to the Student's t-distribution scaled by a factor of $S(G)/(\varepsilon\binom{n}{2})$.

Since G is k-concentrated, we also have that $k_G = 1$, and so $S(G) = O(k + \beta(k+1) + 1/\beta) \leq O(k + 1/\varepsilon)$ by Lemma 3.2. The variance of the Student's t-distribution with three degrees of freedom is $O(1)$, so the expected squared error of the algorithm is

$$O\left(\frac{(k + 1/\varepsilon)^2}{\varepsilon^2 n^4}\right) = O\left(\frac{k^2}{\varepsilon^2 n^4} + \frac{1}{\varepsilon^4 n^4}\right)$$

as desired. □

4. Application to Erdős-Rényi Graphs

In this section we show how to apply Algorithm 1 to estimate the parameter of an Erdős-Rényi graph. Pseudocode is given in Algorithm 2.

It is straightforward to prove that this mechanism satisfies differential privacy.

Theorem 4.1. *Algorithm 2 satisfies $(O(\varepsilon), 0)$-node-differential privacy.*

Algorithm 2: Estimating the parameter of an Erdős-Rényi graph.

Input: A graph $G \in \mathcal{G}_n$ and parameters $\varepsilon, \alpha > 0$.
Output: A parameter $0 \le \hat{p} \le 1$.

1 Let $\tilde{p}' \leftarrow \frac{1}{\binom{n}{2}} \sum_e x_e + (2/\varepsilon n) \cdot Z$ where Z is a standard Laplace random variable.

2 Let $\tilde{p} \leftarrow \tilde{p}' + 4\log(1/\alpha)/\varepsilon n$ and $\tilde{k} \leftarrow \sqrt{\tilde{p} n \log(n/\alpha)}$.

3 Return $\hat{p} \leftarrow \mathcal{A}_{\varepsilon,\tilde{k}}(G)$ where $\mathcal{A}_{\varepsilon,\tilde{k}}$ is Algorithm 1 with parameters ε and $\tilde{k}$.

Proof. The first line computes the empirical edge density of the graph G, which is a function with global sensitivity $(n-1)/\binom{n}{2} = 2/n$. Therefore by Theorem 2.4 this step satisfies $(\varepsilon, 0)$-differential privacy. The third line runs an algorithm that satisfies $(O(\varepsilon), 0)$-differential privacy for every fixed parameter $\tilde{k}$. By Lemma 2.2, the composition satisfies $(O(\varepsilon), 0)$-differential privacy. □

Next, we argue that this algorithm satisfies the desired accuracy guarantee.

Theorem 4.2. *For every $n \in \mathbb{N}$ and $\frac{1}{2} \ge p \ge 0$, and an appropriate parameter $\alpha > 0$, Algorithm 2 satisfies*

$$\mathop{\mathbb{E}}_{G \sim G(n,p),\mathcal{A}} \left[(p - \mathcal{A}(G))^2\right] = \frac{p(1-p)}{\binom{n}{2}} + \tilde{O}\left(\frac{\max\{p, \frac{1}{n}\}}{\varepsilon^2 n^3} + \frac{1}{\varepsilon^4 n^4}\right)$$

Proof. We will prove the result in the case where $p \ge \frac{\log n}{n}$. The case where p is smaller will follow immediately by using $\frac{\log n}{n}$ as an upper bound on p. The first term in the bound is simply the variance of the empirical edge-density $\bar{p}$. For the remainder of the proof we will focus on bounding $\mathbb{E}\left[(\bar{p} - \hat{p})^2\right]$.

A basic fact about $G(n,p)$ for $p \ge \frac{\log n}{n}$ is that with probability at least $1 - 2\alpha$: (1) $|\bar{p} - p| \le \sqrt{\frac{\log(1/\alpha)}{n(n-1)}}$, and (2) the degree of every node i lies in the interval $[\bar{d} \pm O(\sqrt{pn\log(n/\alpha)})]$ where $\bar{d}$ is the average degree of G. This follows from Hoeffding's and Bernstein's inequalities, and we condition on these events for the remainder of the argument.

Using Theorem 2.4, we also have that with probability at least $1-\alpha$, the estimate $\tilde{p}'$ satisfies $|\bar{p} - \tilde{p}'| \le 4\log(1/\alpha)/\varepsilon n$. We will also condition on this event for the remainder. Therefore, we have $p \le \tilde{p}$ and $p \ge \tilde{p} - 8\log(1/\alpha)/\varepsilon n$.

Assuming this condition holds, the graph will have $\tilde{k}$-concentrated degrees for $\tilde{k}$ as specified on line 2 of the algorithm. Since this assumption holds, taking $\alpha = 1/\mathrm{poly}(n)$, we have by Theorem 3.3 that

$$\begin{aligned}\mathbb{E}\left[(\bar{p}-\mathcal{A}_{\tilde{k},\varepsilon}(G))^2\right] &= O\left(\frac{\tilde{k}^2}{\varepsilon^2 n^4}+\frac{1}{\varepsilon^4 n^4}\right)\\ &= \tilde{O}\left(\frac{\tilde{p}n}{\varepsilon^2 n^4}+\frac{1}{\varepsilon^4 n^4}\right)\\ &= \tilde{O}\left(\frac{pn+\frac{1}{\varepsilon}}{\varepsilon^2 n^4}+\frac{1}{\varepsilon^4 n^4}\right)\\ &= \tilde{O}\left(\frac{pn}{\varepsilon^2 n^4}+\frac{1}{\varepsilon^4 n^4}\right).\end{aligned}$$

To complete the proof, we can plug in a suitably small $\alpha = 1/\mathrm{poly}(n)$ so that the $O(\alpha)$ probability of failure will not affect the overall mean-squared error in a significant way. □

5. Lower Bounds for Concentrated-Degree Graphs

In this section we prove a lower bound for estimating the number of edges in concentrated-degree graphs. Theorem 1.3, which lower bounds the mean squared error, follows by applying Jensen's Inequality.

Theorem 5.1. *For every $n, k \in \mathbb{N}$, every $\varepsilon \in [\frac{2}{n}, \frac{1}{4}]$ and $\delta \leq \frac{\varepsilon}{32}$, and every (ε, δ)-node-DP algorithm $\mathcal{A}$, there exists $G \in \mathcal{G}_{n,k}$ such that $\mathop{\mathbb{E}}_{\mathcal{A}}[|p_G - \mathcal{A}(G)|] = \Omega\left(\frac{k}{\varepsilon n^2} + \frac{1}{\varepsilon^2 n^2}\right)$.*

The proof relies on the following standard fact about differentially private algorithms. Since we are not aware of a formal treatment in the literature, we include a proof for completeness.

Lemma 5.2. *Suppose there are two graphs $G_0, G_1 \in \mathcal{G}_{n,k}$ at node distance at most $\frac{1}{\varepsilon}$ from one another. Then for every $(\varepsilon, \frac{\varepsilon}{32})$-node-DP algorithm A, there exists $b \in \{0,1\}$ such that*

$$\mathop{\mathbb{E}}_{\mathcal{A}}[|p_{G_b} - \mathcal{A}(G_b)|] = \Omega\left(|p_{G_0} - p_{G_1}|\right).$$

Proof. Let A be any ε-node-DP algorithm. Since G_0, G_1 have node distance at most $\frac{1}{\varepsilon}$, by group privacy (Lemma 2.3), for every set S and every $b \in \{0,1\}$

$$\mathbb{P}[\mathcal{A}(G_b) \in S] \leq e \cdot \mathbb{P}[\mathcal{A}(G_{1-b}) \in S] + \tfrac{e}{32}.$$

Now, let $S_b = \left\{y : |y - p_{G_b}| < \frac{1}{2}|p_{G_0} - p_{G_1}|\right\}$ and note that S_0 and S_1 are disjoint by construction. Let $\rho = \min\{\mathbb{P}[\mathcal{A}(G_0) \in S_0], \mathbb{P}[\mathcal{A}(G_1) \in S_1]\}$. Then we have

$$\begin{aligned}1-\rho &\geq \mathbb{P}[\mathcal{A}(G_0) \notin S_0]\\ &\geq \mathbb{P}[\mathcal{A}(G_0) \in S_1]\\ &\geq e^{-1}\mathbb{P}[\mathcal{A}(G_1) \in S_1] - \tfrac{1}{32}\\ &\geq e^{-1}\rho - \tfrac{1}{32}\end{aligned}$$

from which we can deduce $\rho \leq \frac{4}{5}$. Therefore, for some $b \in \{0,1\}$, we have

$$\mathbb{P}\left[|p_{G_b} - \mathcal{A}(G_b)| \geq \tfrac{1}{2}|p_{G_0} - p_{G_1}|\right] \geq \tfrac{1}{5},$$

from which the lemma follows. □

We will construct two simple pairs of graphs to which we can apply Lemma 5.2.

Lemma 5.3 (Lower bound for large k)**.** *For every $n, k \in \mathbb{N}$ and $\varepsilon \geq 2/n$, there is a pair of graphs $G_0, G_1 \in \mathcal{G}_{n,k}$ at node distance $1/\varepsilon$ such that $|p_{G_0} - p_{G_1}| = \Omega(\frac{k}{\varepsilon n^2})$.*

Proof. Let G_0 be the empty graph on n nodes. Note that $p_{G_0} = 0$, $\bar{d}_{G_0} = 0$, and G_0 is in $\mathcal{G}_{n,k}$.

We construct G_1 as follows. Start with the empty bipartite graph with $\frac{1}{\varepsilon}$ nodes on the left and $n - \frac{1}{\varepsilon}$ nodes on the right. We connect the first node on the left to each of the first k nodes on the right, then the second node on the left to each of the next k nodes on the right and so on, wrapping around to the first node on the right when we run out of nodes. By construction, $p_{G_1} = k/\varepsilon\binom{n}{2}$, $\bar{d}_{G_1} = 2k/\varepsilon n$. Moreover, each of the first $\frac{1}{\varepsilon}$ nodes has degree exactly k and each of the nodes on the right has degree

$$\frac{k/\varepsilon}{n - 1/\varepsilon} \pm 1 = \frac{k}{\varepsilon n - 1} \pm 1$$

Thus, for n larger than some absolute constant, every degree lies in the interval $[\bar{d}_{G_1} \pm k]$ so we have $G_1 \in \mathcal{G}_{n,k}$. □

Lemma 5.4 (Lower bound for small k)**.** *For every $n \geq 8$ and $\varepsilon \in [2/n, 1/4]$, there is a pair of graphs $G_0, G_1 \in \mathcal{G}_{n,1}$ at node distance $1/\varepsilon$ such that $|p_{G_0} - p_{G_1}| = \Omega(\frac{1}{\varepsilon^2 n^2})$.*

Proof. Let $i = \lceil n\varepsilon \rceil$, and let G_0 be the graph consisting of i disjoint cliques each of size $\lfloor n/i \rfloor$ or $\lceil n/i \rceil$. Let G_1 be the graph consisting of $i + 1$ disjoint cliques each of size $\lfloor n/(i + 1) \rfloor$ or $\lceil n/(i + 1) \rceil$. We can obtain G_0 from G_1 by taking one of the cliques and redistributing its vertices among the i remaining cliques, so G_0 and G_1 have node distance $\ell := \lfloor n/(i + 1) \rfloor \leq 1/\varepsilon$. For $1/4 \geq \varepsilon \geq 2/n$ we have that $\ell \geq \lfloor 1/2\varepsilon \rfloor > 1/4\varepsilon$. Transforming G_1 into G_0 involves removing a clique of size ℓ, containing $\binom{\ell}{2}$ edges, and then inserting these ℓ vertices into cliques that already have size ℓ, adding at least ℓ^2 new edges. Consequently G_0 contains at least $\ell^2 - \ell(\ell - 1)/2 = \ell(\ell + 1)/2$ more edges than G_1, so

$$|p_{G_1} - p_{G_0}| \geq \frac{\binom{\ell+1}{2}}{\binom{n}{2}} \geq \frac{\ell^2}{n^2} \geq \Omega(1/\varepsilon^2 n^2),$$

as desired. □

Theorem 5.1 now follows by combining Lemmas 5.2, 5.3, and 5.4.

Acknowledgments

Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing, and while AS was a PhD student in EECS at MIT. AS was supported by NSF CNS-1413920, Sloan/NJIT 996698, MIT/IBM W1771646, and NSF CNS-1804794. JU was supported by NSF grants CCF-1718088, CCF-1750640, and CNS-1816028. The authors are grateful to Adam Smith for helpful discussions.

References

Blocki, Jeremiah, Avrim Blum, Anupam Datta, and Or Sheffet. 2012. "The Johnson-Lindenstrauss Transform Itself Preserves Differential Privacy." *FOCS'12*, 410–419. New Brunswick, NJ, USA. https://arxiv.org/abs/1204.2136.

Blocki, Jeremiah, Avrim Blum, Anupam Datta, and Or Sheffet. 2013. "Differentially private data analysis of social networks via restricted sensitivity." *ITCS '13*, 87–96. Berkeley, CA, USA:ACM. https://doi.org/10.1145/2422436.2422449.

Borgs, Christian, Jennifer T. Chayes, Adam D. Smith, and Ilias Zadik. 2018. "Revealing Network Structure, Confidentially: Improved Rates for Node-Private Graphon Estimation." *FOCS '18*, 533–543. Paris, France. https://doi.org/10.1109/FOCS.2018.00057.

Bun, Mark, and Thomas Steinke. 2019. "Smooth Sensitivity, Revisited." *Manuscript.*

Canonne, Clément L., Gautam Kamath, Audra McMillan, Jonathan Ullman, and Lydia Zakynthinou. 2019. "Private Identity Testing for High Dimensional Distributions." *arXiv preprint arXiv:1905.11947*. https://arxiv.org/abs/1905.11947.

Cummings, Rachel, and David Durfee. 2018. "Individual Sensitivity Preprocessing for Data Privacy." *arXiv preprint arXiv:1804.08645*. https://arxiv.org/abs/1804.08645.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." *TCC '06*, 265–284. Berlin, Heidelberg:Springer. https://doi.org/10.1007/11681878_14.

Gupta, Anupam, Aaron Roth, and Jonathan Ullman. 2012. "Iterative Constructions and Private Data Release." *TCC '12*, 339–356. Taormina, Italy:Springer. https://arxiv.org/abs/1107.3731.

Hay, Michael, Chao Li, Gerome Mikalu, and David D. Jensen. 2009. "Accurate Estimation of the Degree Distribution of Private Networks." *ICDM'09*, 169–178. Miami, FL, USA. https://doi.org/10.1109/ICDM.2009.11.

Karwa, Vishesh, and Aleksandra Slavković. 2016. "Inference using noisy degrees: Differentially private β-model and synthetic graphs." *Annals of Statistics*, 44(1): 87–112. https://doi.org/10.1214/15-AOS1358.

Karwa, Vishesh, Sofya Raskhodnikova, Adam D. Smith, and Grigory Yaroslavtsev. 2014. "Private Analysis of Graph Structure." *ACM Transactions on Database Systems*, 39(3): 22:1–22:33. https://doi.org/10.1145/2611523.

Kasiviswanathan, Shiva P., Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. 2013. "Analyzing Graphs with Node Differential Privacy." *TCC '13*, 457–476. Tokyo, Japan:Springer. https://doi.org/10.1007/978-3-642-36594-2_26.

Nissim, Kobbi, Sofya Raskhodnikova, and Adam Smith. 2007. "Smooth Sensitivity and Sampling in Private Data Analysis." *STOC '07*, 75–84. New York, NY, USA:Association for Computing Machinery. https://doi.org/10.1145/1250790.1250803.

Raskhodnikova, Sofya, and Adam D. Smith. 2016. "Lipschitz Extensions for Node-Private Graph Statistics and the Generalized Exponential Mechanism." *FOCS '16*, 495–504. New Brunswick, NJ, USA. https://doi.org/10.1109/FOCS.2016.60.

Xiao, Qian, Rui Chen, and Kian-Lee Tan. 2014. "Differentially private network data release via structural inference." *KDD'14*, 911–920. https://doi.org/10.1145/2623330.2623642.