# 1 Differential Privacy on graphs

## 1.1 Edge density

**Proposition 1.1.** *The sufficient and necessary level of the noise to $\varepsilon$-differentiably private estimate the edge density of a graph (where the set of graphs have a "known" degree distribution, scales like $\frac{\max_i d_i}{n^2\varepsilon}$. In paricular for $\rho$-sparse graphon like $\frac{\rho_n}{n\varepsilon}$.*

## 1.2 SBMs

**Definition 1.2.** *We say that two (undirected) graphs $G_1$ and $G_2$ have distance 1 if $G_1$ is generated from $G_2$ by deleting a node from $G_1$ and all its adjacent edges. Let $d$ be the natural metric induced by this operation.*

We start with k-Stochastic Block Models where the clustering sizes can vary arbitrarily.

**Proposition 1.3.** *Let $M, \Delta \in \mathbb{N}$ and $\varepsilon, \eta > 0$ such that $M > 2^{\Delta\varepsilon+1}$.*

*Suppose that $Q_1, \dots, Q_M \in [0,1]^{k\times k}$ induce k-SBM's with the property that $\min_{i,j} \|Q_i - Q_j\|_2 \geq \eta$ and any two $Q_i, Q_j$ induce almost surely graphs that are at most $\Delta$-close. Then any $\varepsilon$-differential private mechanism $M$ should have rate at least $\eta/2$.*

*Proof.* Same idea as in the survey. Also $k$-close $Q's$ imply $k$-close graphs (check this with a coupling possible argument). □

**Proposition 1.4.** *Let $k > 2$ (possibly not necessary). The rate for $\varepsilon$-diff approximating the matrix $Q$ from the SBM is at least of the order $\frac{k}{n\varepsilon} = \frac{1}{\varepsilon}\sqrt{\frac{k^2}{n^2}}$. That is for any $\varepsilon$-differential private estimator $\hat{Q}$ it holds $\|\hat{Q} - Q\|_2^2 = \Omega(\frac{1}{\varepsilon}\sqrt{\frac{k^2}{n^2}})$.*

*Proof.* Choose a family $F$ of $2^{k/2}$ sets $S \subset \{0,1\}^k$ with $|S \triangle T| \geq k/4$ for any two $S \neq T$ both elements of $F$. (check constants).

Then choose $Q = 0$ with the block structure that the first $k/2$ vertices correspond to a block of size $\lambda > 0$ and after these we have blocks of size $2(n - \lambda k)/k$.The parameter $\lambda > 0$ will be tuned later. We enumerate the blocks so that the first $k/2$ blocks correspond to the sizes $\lambda > 0$.

Now for every $S \in F$, we define $Q_S$ we define the $k$-SBM probability matrix such that for the block corresponding to $i \in S$, we adjust the inwards and outwards probabilities to be equal to 1, that is $Q_S = 1_{b(i)\in S, j\in[n]}$, where by $b(i)$ we refer to the block that vertex $i$ belongs.

Now clearly every pair $Q_S, Q_T$ is at most $\lambda k$-away and $\|Q_S - Q_T\|_2^2 \geq c'|S\triangle T|n = c\lambda kn$. Hence by setting $\lambda = \frac{1}{7\varepsilon}$ we have that $M := |F| > 2^{\Delta\varepsilon+1}$ since $k > 2$ we are done.

□

Remark: The above argument using the result from Gao et al [16] can be boosted for the equipartition case to $\lambda\rho$ rate. In particular, no reasonable estimation seems possible in this case with differentiable private estimation. **Differential private algorithms can not give the clustering partition!**

**Proposition 1.5.** *There exists a constant $c > 0$ small enough such that if $\varepsilon < c\frac{k^2}{n}$ then the trivial estimator is the optimal $\varepsilon$-differentiable private estimator for estimating the probability matrix of a $k$-SBM.*

*Proof.* Using the Vashanorv-Gilber bound (probabilistic method) we can find $M = 2^{\Omega(k^2)} > 2^{\varepsilon n+1}$ matrices $B_i \in \{0,1\}^{k\times k}$ such that $\inf_{i\neq j} \|B_i - B_j\|_2 = \Omega(1)$. Since each two $k$-SBMs induces graphs that are $n$-close we get the result using the Proposition 1.3. $\square$

**Proposition 1.6.** *There exists $c > 0$ such that if $\varepsilon < c\log k$, then the trivial estimator is the optimal $\varepsilon$-differentiable private estimator for estimating the probability matrix of the induced graph of a $k$-SBM.*

*Proof.* Indeed in pages 26-27 of Gao et al, we see the existence of $\exp(cn\log k)$ probability matrices of graphs induced by balances $k$-SBMs with pairwise $\ell_2$ distance constant. The result follows from the obvious variant of Proposition 1.3. $\square$

**Proposition 1.7.** *Suppose for some graphon $W$ we see the $A = G_n(W)$ and then we want to $\varepsilon$-differential private estimate the graphon $W$ given $A$, by producing say the estimator $\hat{W}$. Then if $1 > \varepsilon > \frac{1}{n}$, it necessarily holds for some $W$ that $\mathbb{E}\left[\delta_2(\hat{W}, W)\right] \geq \Omega(\frac{1}{n\varepsilon})$.*

*Proof.* Assume not. Then for all $W$ using Cauchy-Shwartz we get that

$$\mathbb{E}\left[\mathbb{P}\left(|\|\hat{W}\|_1 - \|W\|_1| < \frac{1}{100n\varepsilon}|A\right)\right] \geq 1-\delta$$

But then we can choose $W_1 = 0$ and $W_2 = 1(\min\{x,y\} < \frac{1}{n\varepsilon})$ and let $A_1, A_2$ the graphs they induce. Hence it holds

$$\mathbb{E}\left[\mathbb{P}\left(|\|\hat{W}\|_1| < \frac{1}{100n\varepsilon}|A_1\right)\right] \geq 1-\delta$$

$$\mathbb{E}\left[\mathbb{P}\left(|\|\hat{W}\|_1 - \frac{1}{n\varepsilon}| < \frac{1}{100n\varepsilon}|A_2\right)\right] \geq 1-\delta$$

So for $X(A) = \mathbb{P}\left(|\|\hat{W}\|_1| < \frac{1}{100n\varepsilon}|A\right)$ it holds $\mathbb{E}[X(A_1)] \geq 1-\delta$ and $\mathbb{E}[X(A_2)] \leq \delta$. But from $\varepsilon$-differentiable privacy we have a.s. with the obvious coupling that $\frac{X(A_1)}{X(A_2)} \leq 2^{\varepsilon U}$ a.s. where $U$ follows Binomial$(n, \frac{1}{n\varepsilon})$. Hence,

$$1-\delta \leq \mathbb{E}[X(A_1)]^2 \leq \mathbb{E}\left[[\frac{X(A_1)}{X(A_2)}]^2\right]\mathbb{E}\left[X(A_2)^2\right] \leq \mathbb{E}\left[2^{2\varepsilon U}\right]\delta \leq O(1)\delta,$$

a contradiction.

$\square$

Remark: Assuming our sparsity $\rho > \frac{1}{\varepsilon n}$ this min-max bound transfers in this case as well. Important Observation for the Borgs et al question:

We need to to control the $k \times k$ matrices with distances $\inf_\pi$ introduced in Borgs et al [14].

**Question 1.8.** *Given $\varepsilon$ find the maximum $\eta > 0$ such that for some $\Delta > 0$ and $M > 2^{\varepsilon \frac{n}{k} \Delta}$ there exist $B_1, \ldots, B_M \in [0,1]^{k \times k}$ such that for all $i \neq j$, $\inf_\pi \|B_{i,\pi} - B_j\|_2 \geq \eta$ and $B_i, B_j$ are $\Delta$-close.*

# 2 Differential Privacy and the Lipschitz Property

Consider $(\mathcal{G}_n, \delta_v)$ the set of all graphs with $n$ vertices treated as a metric space with the vertex-distance and $(\mathcal{M}, D_\infty)$ the set of probability measures in $[0,1]$ with the Borel $\sigma$-field treated as metric space with the $D_\infty$ distance. We remind that for measures $\mu, \mu'$, $D_\infty(\mu, \mu') = \|\log \mu - \log \mu'\|_\infty$.

**Proposition 2.1.** *A mapping $\mu : (\mathcal{G}_n, \delta_v) \to (\mathcal{M}, D_\infty)$ corresponds to an $\varepsilon$-differential private mechanism if and only if $\mu$ is $\varepsilon$-Lipschitz with respect to $\delta_v$ and $D_\infty$.*

*Proof.* Doable. □

**Proposition 2.2.** *Suppose that for some $H_n \subset \mathcal{G}_n$ a function $\hat{\mu} : (H_n, \delta_v) \to (\mathcal{M}, D_\infty)$ is $\varepsilon$-Lipschitz for some $\varepsilon > 0$. Then we can extend the function to a $\mu : (\mathcal{G}_n, \delta_v) \to (\mathcal{M}, D_\infty)$ such that it is $2\varepsilon$-Liptschitz and for every $G \in H_n$, $\mu(G) = \hat{\mu}(G)$.*

*Proof.* We define for every $G \in \mathcal{G}_n$,

$$d\mu(G) \propto \inf_{G' \in H_n} 2^{\varepsilon \delta_v(G,G')} d\hat{\mu}.$$

Both the properties follow. □

## 2.1 The minimax rate

Given the Proposition the rate we want to find is

$$R = \min_{\mu:(\mathcal{G}_n,\delta_v)\to(\mathcal{M},D_\infty)\varepsilon-Lipschitz} \max_{p\in[0,1]} \mathbb{E}_{G\sim G_{n,p},\hat{p}\sim\mu_G}[|\hat{p} - p|]$$

## 2.2 An upper bound

First we define for every $p \in [0,1]$ we define the distribution over [0,1] given by

$$\hat{\mu}_p(q) \propto 2^{\varepsilon \min\{\frac{n^2}{\log n}|p-q|, n\}}$$

for $q \in [0,1]$. Now we define $H_n \subseteq \mathcal{G}_n$ to be a subset of all the graphs such that

$$\forall p \in [0,1], \frac{1}{n^{10}} \leq \mathbb{P}_{G\sim G_{n,p}} (G \notin H_n) \leq \frac{1}{n}.$$

Then for any $G \in \mathcal{G}_n$ we set $\mu : (\mathcal{G}_n, \delta_v) \to (\mathcal{M}, D_\infty)$ the mapping given by

$$\mu_G(q) \propto \inf_{G' \in H} \left[ 2^{\varepsilon \delta_v(G,G')} \hat{\mu}_{e(G')}(q) \right],$$

for $q \in [0,1]$.

**Claim 2.3.** *The map $\mu_G$ is $2\varepsilon$-D.P.*

*Proof.* Follows from the definition. □

For typical $G_{n,p}, G_{n,q}$ it holds $\delta_V(G,G') \geq \min\{\frac{n^2}{\log n}|p-q|, n\}$.