1 Differential Privacy on graphs

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 σ-field treated as metric space with the D_∞ distance. We remind that for measures μ, μ', $D_\infty(\mu, \mu') = \| \log \mu - \log \mu' \|_\infty$.

We start with two observations.

Proposition 1.1. *A mapping $\mu : (\mathcal{G}_n, \delta_v) \to (\mathcal{M}, D_\infty)$ corresponds to an ε-differential private mechanism if and only if μ is ε-Lipschitz with respect to δ_v and D_∞.*

Proof. Follows from the definition. □

Proposition 1.2. *Suppose that for some $H_n \subset \mathcal{G}_n$ a function $\hat{\mu} : (H_n, \delta_v) \to (\mathcal{M}, D_\infty)$ is ε-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 2ε-Liptschitz and for every $G \in H_n$, $\mu(G) = \hat{\mu}(G)$.*

Proof. We define for every $G \in \mathcal{G}_n$ and A in the σ-field

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

Both the properties follow. The differential privacy follows like for the exponential mechanism. □

1.1 The minimax rate

Given the Proposition the rate we want to find is

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

1.2 A $n^{\frac{3}{2}}$-upper bound

Proposition 1.3. *For the minimax rate defined above it holds*

$$R \leq O\left(\frac{1}{n} + \max\{p(1-p), \sqrt{\frac{\log n}{n}}\} \frac{\sqrt{\log n}}{n^{\frac{3}{2}} \varepsilon} \right).$$

We start with a lemma.

Lemma 1.4. *Let $p \in [0,1]$. For every $S \subseteq V(G)$, $|S| = k$ set the event*

$$A_{p,S} := \{|E(S,S^c) + E(S,S) - p\left[k\,(n-k) + \binom{k}{2}\right]| \leq \max\{p(1-p), \sqrt{\frac{\log n}{n}}\} 2k\sqrt{n \log n}|\}.$$

Then it holds

$$\mathbb{P}_{G\sim G_{n,p}}\left[\bigcup_{S\subseteq V(G)} A^c_{p,S} \right] \leq \frac{1}{n^2}.$$

Proof. Set $c = \max\{p(1-p), \sqrt{\frac{\log n}{n}}\}$. By a union bound, Berstein inequality and basic algebra we have

$$\begin{aligned}
&\mathbb{P}_{G\sim G_{n,p}}\left[\bigcup_{S\subseteq V(G)} A^c_{p,S}\right] \\
&\leq \sum_{k=1}^{n}\binom{n}{k}\exp\left(-4\frac{c^2k^2 n\log n}{\left(k(n-k)+\binom{k}{2}\right)p(1-p)+2ck\sqrt{n\log n}}\right) \\
&\leq \sum_{k=1}^{n} n^k n^{-4k} \\
&\leq n\frac{1}{n^3} = \frac{1}{n^2}
\end{aligned}$$

□

Proof. We now begin the proof of Proposition (1.3). We remind the reader that the sampling error is $\frac{1}{n}$.

Given the Proposition (1.2) a strategy would be to find a subset H_n of all the graphs on n vertices so that

$$\max_{p\in[0,1]} \mathbb{P}_{G\sim G_{n,p}}\left(G \notin H_n\right) \leq \frac{1}{n} \tag{1.1}$$

and furthermore define an ε-Lip function $\hat{\mu}$ on H_n so that for all $G \in H_n$,

$$\mathbb{E}_{\hat{p}\sim\hat{\mu}_G}[|\hat{p}-e(G)|] \leq \max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\frac{100}{n^{\frac{3}{2}}\varepsilon}\sqrt{\log n} \tag{1.2}$$

Then, from Proposition (1.2) we could extend this mapping to a 2ε-Lipschitz mapping on the space of all graphs and furthermore have for all p,

$$\begin{aligned}
&\mathbb{E}_{G\sim G_{n,p},\hat{p}\sim\mu_G}[|\hat{p}-e(G)|] \leq \mathbb{P}_{G\sim G_{n,p}}\left(G\notin H_n\right) + \max_{G\in H_n}\mathbb{E}_{\hat{p}\sim\hat{\mu}_G}[|\hat{p}-e(G)|] \\
&= O\left(\frac{1}{n} + \max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\frac{\sqrt{\log n}}{n^{\frac{3}{2}}\varepsilon}\right)
\end{aligned}$$

Given lemma (1.4) we define

$$H_n = \bigcup_{p\in[0,1]}\bigcap_{S\subseteq V(G)} A_{p,S},$$

that is all the graphs on n vertices for which for some $p \in [0,1]$ all $A_{p,S}$ are satisfied. This represents for us the class of **homogeneous** graphs. Given the Lemma (1.4) we know that indeed (1.1) is satisfied.

For the next condition we define for every graph $G \in H_n$ the distribution over $[0,1]$ to come from the addition of "truncated" Laplacian noise given by

$$\hat{\mu}_G(q) \propto 2^{-\varepsilon c \min\{\frac{n^{\frac{3}{2}}}{\max\{p(1-p),\sqrt{\frac{\log n}{n}}\}\sqrt{\log n}}|e(G)-q|, n\}}$$

for $q \in [0,1]$. The constant $c > 0$ will be satisfied later on.

It is easy to prove that (1.2) is satisfied but we need to prove that our mapping is ε-Lip.

To do this it is easy to establish first by triangle inequality that for any graphs on n vertices G_1, G_2

$$D_\infty\left(\hat{\mu}_{G_1}, \hat{\mu}_{G_2}\right) \leq 2\varepsilon c \min\{\frac{n^{\frac{3}{2}}}{\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\sqrt{\log n}}|e(G_1) - e(G_2)|, n\}$$

Hence we only need to prove that for some $c > 0$ small enough and for any $G_1, G_2 \in H$ it holds

$$c \min\{\frac{n^{\frac{3}{2}}}{\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\sqrt{\log n}}|e(G) - e(G')|, n\} \leq \delta_V(G, G').$$

This is what we prove in the next claim we completes the proof.

Claim 1.5. *There exists a universal constant $c > 0$ such that for any $G, G' \in H$, it holds*

$$c \min\{\frac{n^{\frac{3}{2}}}{\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\sqrt{\log n}}|e(G) - e(G')|, n\} \leq \delta_V(G, G').$$

Proof. Let $G, G' \in H$. By assuming $c < \frac{1}{4}$ we may assume that $\delta_V(G, G') \leq \frac{n}{4}$. In that case we will prove that for some universal $c > 0$,

$$c\frac{n^{\frac{3}{2}}}{\sqrt{\log n}}|e(G) - e(G')| \leq \max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\delta_V(G, G').$$

Let p, q such that $G \in \bigcap_{S \subseteq V(G)} A_{p,S}$ and $G' \in \bigcap_{S \subseteq V(G)} A_{q,S}$. Consider $S_0 \subseteq V(G)$ the vertices that need to be rewired to change G to G'. In particular it holds $\delta_V(G, G') = |S_0| =: k$. Now we have

$$\begin{aligned} &|E(G) - E(G')| \\ &= |E_G(S_0, S_0) + E_G(S_0, S_0^c) - E_{G'}(S_0, S_0) - E_{G'}(S_0, S_0^c)| \\ &\leq |p - q|\left(k(n-k) + \binom{k}{2}\right) + 4\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}k\sqrt{n\log n} \text{ ,using } G \in A_{p,S_0}, G' \in A_{q,S_0} \end{aligned}$$

Now observe that since $G \in A_{p,V(G)}, G' \in A_{q,V(G')}$ it holds $|E(G) - p\binom{n}{2}| \leq 2\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}n\sqrt{n\log n}$, $|E(G') - q\binom{n}{2}| \leq 2\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}n\sqrt{n\log n}$. Hence,

$$|p - q| \leq \frac{1}{\binom{n}{2}}|E(G) - E(G')| + \frac{4\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}n\sqrt{n\log n}}{\binom{n}{2}}$$

Plugging this into the previous inequality we have,

$$|E(G) - E(G')| \leq$$
$$\left[\frac{1}{\binom{n}{2}}|E(G) - E(G')| + \frac{4\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}n\sqrt{n\log n}}{\binom{n}{2}}\right]\left(k(n-k) + \binom{k}{2}\right)$$
$$+ 4\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}k\sqrt{n\log n},$$

Since $k(n-k) + \binom{k}{2} \leq kn$ we can equivalently write the inequality as

$$\left(\binom{n}{2} - kn\right)|e(G) - e(G')| \leq 8\frac{n^2}{\binom{n}{2}}\max\{p(1-p), \sqrt{\frac{\log n}{n}}\}k\sqrt{n\log n}.$$

But now as we have assumed $k \leq \frac{n}{4}$ we have $\binom{n}{2} - kn \geq \frac{n^2}{8}$ (large n) and since $\frac{n^2}{\binom{n}{2}} \leq 4$ (large n) the inequality gives for some universal $c > 0$

$$c\frac{n^{\frac{3}{2}}}{\sqrt{\log n}}|e(G) - e(G')| \leq \max\{p(1-p), \sqrt{\frac{\log n}{n}}\}k = \max\{p(1-p), \sqrt{\frac{\log n}{n}}\}\delta_V(G, G'),$$

as we wanted.

□

□

2 The lower bound

Let $n, k \in \mathbb{N}$ and $N = \binom{n}{2}, M = (N-k)/2$.

2.1 The 1-distance case

We consider two models. The first is $\mathbb{P}_1 = G(n, M)$, that is sample a uniform graph on n vertices and M edges. The second is $\mathbb{P}_2 = G(n, M, k)$: sample first uniformly a graph on n vertices and $M + k$ edges, choose a uniformly chosen maximum-degree vertex and then delete $\min\{d_{\max}, k\}$ edges which are adjacent to the vertex uniformly at random.

Theorem 2.1. *Suppose $k = \frac{1}{\sqrt{2}}\sqrt{n\log n}$. Then*

$$\lim_{n\to+\infty} \mathrm{TV}(\mathbb{P}_1, \mathbb{P}_2) = 0.$$

Proof. By Pinsker's inequality it suffices to deal with the KL-divergence and show

$$\lim_{n\to+\infty} \mathbb{E}_{G_0\sim\mathbb{P}_1}\left[\log\frac{\mathbb{P}_2[G=G_0]}{\mathbb{P}_1[G=G_0]}\right]=0.$$

By Jensen inequality we have for all n,

$$\mathbb{E}_{G_0\sim\mathbb{P}_1}\left[\log\frac{\mathbb{P}_2[G=G_0]}{\mathbb{P}_1[G=G_0]}\right]\le\log\mathbb{E}_{G_0\sim\mathbb{P}_1}\left[\frac{\mathbb{P}_2[G=G_0]}{\mathbb{P}_1[G=G_0]}\right]=\log 1=0.$$

So it suffices to show

$$\liminf_{n\to+\infty} \mathbb{E}_{G_0\sim\mathbb{P}_1}\left[\log\frac{\mathbb{P}_2[G=G_0]}{\mathbb{P}_1[G=G_0]}\right]=0.$$

Now for any G_0 on n vertices with M edges we lower bound $\mathbb{P}_2[G=G_0]$ as follows,

$$\begin{aligned}
\mathbb{P}_2[G=G_0] &= \sum_{G' \text{ with M+k edges}} \mathbb{P}(G' \text{ is chosen in the first step})\mathbb{P}(G_0|G') \\
&= \sum_{G' \text{ with M+k edges and } \mathbb{P}(G_0|G')>0} \frac{1}{\binom{N}{M+k}}\mathbb{P}(G_0|G')\ (G' \text{ is chosen u.a.r.}) \\
&= \frac{1}{\binom{N}{M+k}} \sum_{v\in V(G_0):d^{G_0}(v)\ge d^{G_0}_{max}-k}\ \sum_{G' \text{ is plausible by } G_0 \text{ via } v} \mathbb{P}(G_0|G') \\
&= \frac{1}{\binom{N}{M+k}} \sum_{v\in V(G_0):d^{G_0}(v)\ge d^{G_0}_{max}-k}\ \sum_{G' \text{ is plausible by } G_0 \text{ via } v} \frac{1}{|\text{max. degree vertices in } G'|\binom{d^{G_0}(v)+k}{k}} \\
&= \frac{1}{\binom{N}{M+k}} \sum_{v\in V(G_0):d^{G_0}(v)\ge d^{G_0}_{max}-k} \frac{\binom{n-d^{G_0}(v)-1}{k}}{|\text{max. degree vertices in } G'|\binom{d^{G_0}(v)+k}{k}} \\
&\ge \frac{1}{\binom{N}{M+k}} \sum_{v\in V(G_0):d^{G_0}(v)\ge d^{G_0}_{max}-k+2} \frac{\binom{n-d^{G_0}(v)-1}{k}}{\binom{d^{G_0}(v)+k}{k}}\ (\text{in these cases unique max degree vertex}) \\
&\ge \frac{1}{\binom{N}{M+k}} \sum_{v\in V(G_0):d^{G_0}(v)\ge d^{G_0}_{max}-k+2} \left(\frac{n-d^{G_0}(v)-1}{d^{G_0}(v)+k}\right)^k (1+O(\frac{k}{d^{G_0}(v)})) \\
&\ge \frac{1}{\binom{N}{M+k}} \sum_{v\in V(G_0):d^{G_0}_{max}-k+k/\log\log n\ge d^{G_0}(v)\ge d^{G_0}_{max}-k+2} \left(\frac{n-d^{G_0}(v)-1}{d^{G_0}(v)+k}\right)^k (1+O(\frac{k}{d^{G_0}(v)})) \\
&\ge \frac{1}{\binom{N}{M+k}} \mathcal{Z}\left(\frac{n-d^{G_0}_{max}+k+o(k)-3}{d^{G_0}_{max}+o(k)}\right)^k (1+O(\frac{k}{d^{G_0}_{max}-k})
\end{aligned}$$

for $\mathcal{Z}$ is the number of vertices in G_0 with degree between $d^{G_0}_{max}-k$ and $d^{G_0}_{max}-k+k/\log\log n$. As by definition $\mathbb{P}_1[G=G_0]=\frac{1}{\binom{N}{M}}=\frac{1}{\binom{N}{M+k}}$ ($M+k=N-M$) we conclude

$$\mathbb{E}_{G_0\sim\mathbb{P}_1}\left[\log\frac{\mathbb{P}_2[G=G_0]}{\mathbb{P}_1[G=G_0]}\right]$$

is at least

$$\mathbb{E}_{G_0\sim\mathbb{P}_1}\log\mathcal{Z}+k\mathbb{E}_{G_0\sim\mathbb{P}_1}\log\left(\frac{n-d^{G_0}_{max}+k+o(k)-3}{d^{G_0}_{max}+o(k)}\right)+\mathbb{E}_{G_0\sim\mathbb{P}_1}\log(1+O(\frac{k}{d^{G_0}_{max}-k})).$$

Lemma 2.2. *With probability* $1-\exp(-c(\log n)^{1/4})$, $|d^{G_0}_{max}-(n-1)/2-k|\le\sqrt{n}/(\log n)^{\frac{1}{4}}$ *and* $\mathcal{Z}\ge n-10\sqrt{n}\log n$.

Proof. To be added. □

Using Lemma 2.2 we have (using also $\log(1+x)=x+o(x)$ twice)

$$\begin{aligned}&\mathbb{E}_{G_0\sim\mathbb{P}_1}\log\mathcal{Z}+k\mathbb{E}_{G_0\sim\mathbb{P}_1}\log\left(\frac{n-d^{G_0}_{max}+k+o(k)-3}{d^{G_0}_{max}+o(k)}\right)+\mathbb{E}_{G_0\sim\mathbb{P}_1}\log(1+O(\frac{k}{d^{G_0}_{max}-k})).\\&\ge\log n+O(\log n/\sqrt{n})+k\log\left(\frac{n/2+o(k)-3}{n/2+k+o(k)}\right)+O(k/n)\\&=\log n+k\left(\frac{n/2+o(k)-3}{n/2+k+o(k)}-1\right)+o(1)\\&=\log n-2k^2/n+O(k/n)\\&=o(1),\text{ since }k^2=2n\log n.\end{aligned}$$

The proof is complete. □

2.2 The general case

Let $C\in\mathbb{N}$ and $k_1,\dots,k_C>0$ with $k=\sum_{i=1}^C k_i$. We consider two random generating models. The first is again $\mathbb{P}_1=G(n,M)$, that is sample a uniform graph on n vertices and M edges. The second is $\mathbb{P}_2=G(n,M,k_1,k_2,\dots,k_C)$: sample first uniformly a graph on n vertices and $M+k$ edges, list (solving ties randomly) the vertices into a decreasing order and then for $i=1,2,\dots,C$ delete from the i-th vertex $\min\{d_i,k_i\}$ adjacent edges uniformly at random.

Theorem 2.3. *Suppose* $C=o(\sqrt{n})$ *and for all* $i=1,2,\dots,C$, $k=\frac{1}{\sqrt{2}}\sqrt{n\log(n/i)}$. *Then*

$$\lim_{n\to+\infty}\mathrm{TV}\left(\mathbb{P}_1,\mathbb{P}_2\right)\to 0.$$

Proof. Similarly with the previous proof we lower bound $\mathbb{P}_2[G=G_0]$ as follows

$$\begin{aligned}\mathbb{P}_2[G=G_0]&=\sum_{G'\text{ with M+k edges}}\mathbb{P}(G'\text{ is chosen in the first step})\mathbb{P}(G_0|G')\\&=\sum_{G'\text{ with M+k edges and }\mathbb{P}(G_0|G')>0}\frac{1}{\binom{N}{M+k}}\mathbb{P}(G_0|G')\ (G'\text{ is chosen u.a.r.})\\&=\frac{1}{\binom{N}{M+k}}\sum_{v_1,\dots,v_C\in V(G_0):d^{G_0}(v_i)\ge d_i^{G_0}-k_i}\ \sum_{G'\text{ is plausible by }G_0\text{ via }v_1,\dots,v_C}\mathbb{P}(G_0|G')\end{aligned}$$

Now as before notice that if we choose any unordered list of vertices $v_1, \dots, v_C$ that satisfy for all j,

$$l := \max_{i \in [C]} (d_i - k_i) \le d^{G_0}(v_j) \le \min_{i \in [C]} (d_i - k_i + k_i / \log\log n) =: L,$$

then after ordering them so that their degrees are decreasing, we can add any k_i edjes to v_i (among the non-adjacent edges and not edges connecting with other v_i's) to G_0 and create a plausible G'. In particular that calculation implies,

$$\begin{aligned}
\mathbb{P}_2[G = G_0] &\ge \frac{1}{\binom{N}{M+k}} \sum_{v_1,\dots,v_C \in V(G_0): l \le d^{G_0}(v_i) \le L} \prod_{i=1}^{C} \frac{\binom{n - d^{G_0}(v_i) - C - 1}{k_i}}{\binom{d^{G_0}(v_i) + k_i}{k_i}} \\
&\ge \frac{1}{\binom{N}{M+k}} \sum_{v \in V(G_0): v_1,\dots,v_C \in V(G_0): l \le d^{G_0}(v_i) \le L} \prod_{i=1}^{C} \left(\frac{n - d^{G_0}(v_i) - C}{d^{G_0}(v_i) + k_i} \right)^{k_i} (1 + O(\frac{k_i}{d^{G_0}(v_i)})) \\
&\ge \frac{1}{\binom{N}{M+k}} \mathcal{Z}_{\mathcal{C}} \prod_{i=1}^{C} \left(\frac{n - d_i^{G_0} + k_i + o(k_i) - 3}{d_i^{G_0} + o(k_i)} \right)^{k_i} (1 + O(\frac{k_i}{d_i^{G_0} - k_i}))
\end{aligned}$$

for $\mathcal{Z}$ is the number of vertices in G_0 with degree between l and L.

As by definition $\mathbb{P}_1[G = G_0] = \frac{1}{\binom{N}{M}} = \frac{1}{\binom{N}{M+k}}$, since $M + k = N - k$, we conclude

$$\mathbb{E}_{G_0 \sim \mathbb{P}_1} \left[\log \frac{\mathbb{P}_2[G = G_0]}{\mathbb{P}_1[G = G_0]} \right]$$

is at least

$$\mathbb{E}_{G_0 \sim \mathbb{P}_1} \log \mathcal{Z}_C + \sum_{i=1}^{C} k_i \mathbb{E}_{G_0 \sim \mathbb{P}_1} \log \left(\frac{n - d_i^{G_0} + k_i + o(k_i) - 3}{d_i^{G_0} + o(k_i)} \right) + \sum_{i=1}^{C} \mathbb{E}_{G_0 \sim \mathbb{P}_1} \log(1 + O(\frac{k_i}{d_i^{G_0} - k_i}).$$

Lemma 2.4. *With probability* $1 - \exp(-c(\log n)^{1/4})$, $|d_i^{G_0} - (n-1)/2 - k_i| \le \sqrt{n}/(\log n)^{\frac{1}{4}}$ *and* $\mathcal{Z} \ge \binom{n}{C}(1 + o(1))$.

Using Lemma 2.4 we have

$$\begin{aligned}
&\mathbb{E}_{G_0 \sim \mathbb{P}_1} \log \mathcal{Z}_C + \sum_{i=1}^{C} k_i \mathbb{E}_{G_0 \sim \mathbb{P}_1} \log \left(\frac{n - d_i^{G_0} + k_i + o(k_i) - 3}{d_i^{G_0} + o(k_i)} \right) + \sum_{i=1}^{C} \mathbb{E}_{G_0 \sim \mathbb{P}_1} \log(1 + O(\frac{k_i}{d_i^{G_0} - k_i}). \\
&\ge C \log n + o(1) + \sum_{i=1}^{C} k_i \log \left(\frac{n/2 + o(k_i) - C}{n/2 + k_i + o(k_i)} \right) + \sum_{i=1}^{C} O(k_i/n) \\
&= C \log n - 2 \sum_{i=1}^{C} k_i^2/n + O(C \sum k_i/n) + o(1), \text{ using } \log(1+x) = x + o(x) \\
&= 0 + O(C/\sqrt{n}) \\
&= o(1).
\end{aligned}$$

□