Differential Privacy for Decentralized Optimization

CIRM 2022

Edwige Cyffers (Inria Lille)
Work done with Aurélien Bellet, Mathieu Even and Laurent Massoulié

Why Privacy?

Right to be left alone [Warren and Brandeis 1890]

The press is overstepping in every direction the obvious bounds of propriety and of decency. Gossip is no longer the resource of the idle and of the vicious, but has become a trade, which is pursued with industry as well as effrontery.

- New technologies that force us to rethink privacy: broader scope of surveillance, improvement of data mining and make the intrusion invisible [Snowden 2019]
- Technical implementation are the Missing Masses of our social changes, compared to other agents, such as legal requirement (GDPR [Conseil 2016]), public opinion, financial incentive.

Privacy 2.0: High Dimension Data destroy anonymity

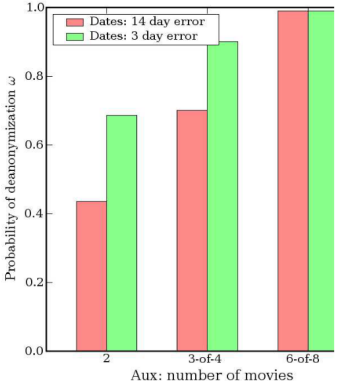
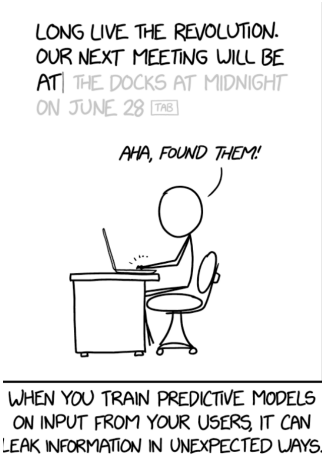
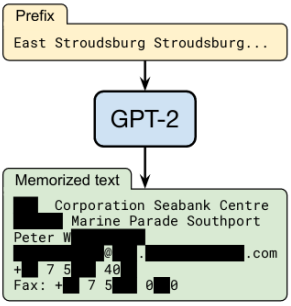


Figure 1: De-anonymization: adversary knows exact ratings and approximate dates.

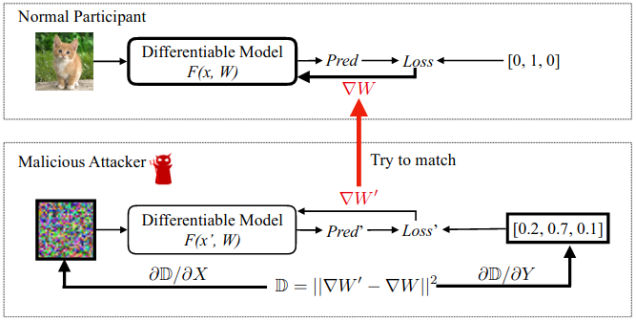
- Large anonymized database published for the Netflix Prize
- Joint with the MovieLens dataset, users who are in both datasets have high probability of re-identification
- Leaks sensitive private rankings (political movies, sexual preferences, religious belief, ...)

Machine Learning Models Privacy Leakage





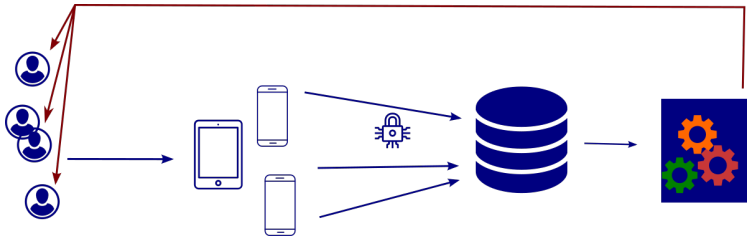
Attacks on Machine Learning models



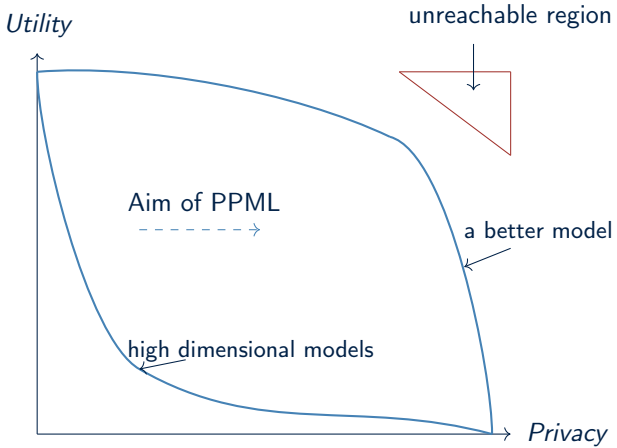
- Young and active research field, better attacks yet to come
- Various dangers: membership inference, attribute inference, full reconstruction attacks
- Various threat models: white/black box model, passive/active, computational power, auxiliary information...

From binary to stochastic privacy

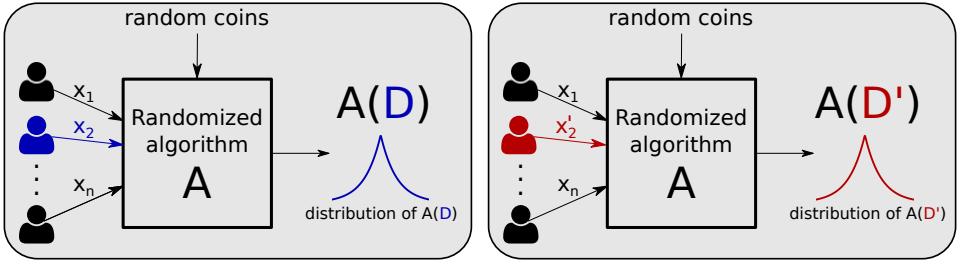
- Not private or non-private, but a probability to leak information
- Differential Privacy only tackles the risk of leakage in algorithms outputs
- It enforces constraints on how much a single instance can influence outputs through noise injection



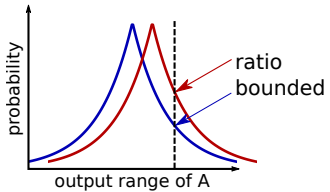
Ultimate goal: good utility *and* privacy



Differential privacy Intuition



- Neighboring datasets $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{D}' = \{x_1, x_2', x_3, \ldots, x_n\}$
- **Requirement**: $\mathcal{A}(\mathcal{D})$ and $\mathcal{A}(\mathcal{D}')$ should have "similar" distributions



Differential Privacy

Definition

Let $\varepsilon \geq 0$ and $\delta \in [0, 1]$. A mechanism, $\mathcal{M}$ is ε-differentially private with respect to $\sim$ if for every pair of databases $X \sim X'$ and every subset $\mathcal{S} \subset Z$:

$$\mathbb{P}(\mathcal{M}(X) \in \mathcal{S}) \leq \exp(\varepsilon)\mathbb{P}(\mathcal{M}(X') \in \mathcal{S}) + \delta$$

- Robust to post-processing
- Based on the protection of the participation
- Sensitive to the definition of adjacency
- Sensitive to the outputs of the mechanism

Renyi Differential Privacy

Rényi Differential Privacy [Mironov 2017]

An algorithm $\mathcal{A}$ satisfies (α, ε)-Rényi Differential Privacy (RDP) for $\alpha > 1$ and $\varepsilon > 0$ if for all pairs of neighboring datasets $\mathcal{D} \sim \mathcal{D}'$:

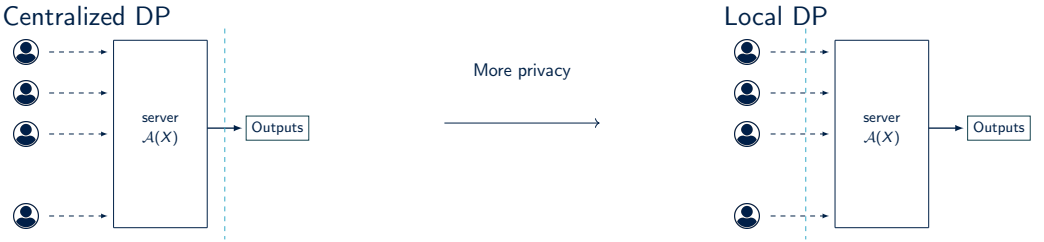
$$D_\alpha\left(\mathcal{A}(\mathcal{D}) || \mathcal{A}(\mathcal{D}')\right) \leq \varepsilon\,, \tag{1}$$

where for two r.v. X, Y with densities μ_X, μ_Y, $D_\alpha\big(X \,||\, Y\big)$ is the Rényi divergence of order α:

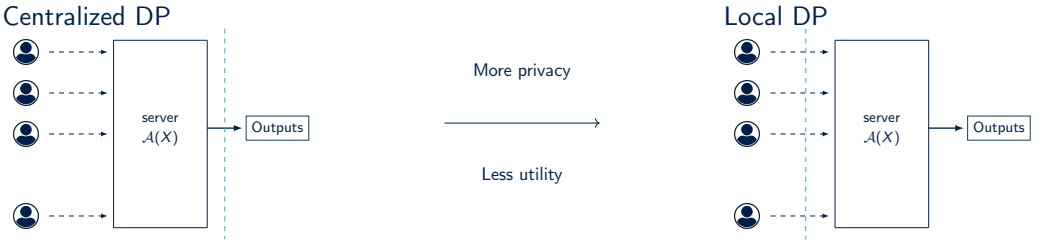
$$D_\alpha\big(X \,||\, Y\big) = \frac{1}{\alpha - 1} \ln \int \left(\frac{\mu_X(z)}{\mu_Y(z)}\right)^\alpha \mu_Y(z) dz\,.$$

- Conversion to (ε, δ)-DP: (α, ε)-RDP implies $(\varepsilon + \frac{\ln(1/\delta)}{\alpha - 1}, \delta)$-DP for any $\delta \in (0, 1)$
- Composition: if $\mathcal{A}_1$ is (α, ϵ_1)-RDP and $\mathcal{A}_2$ is (α, ϵ_2)-RDP, then $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ is $(\alpha, \epsilon_1 + \epsilon_2)$-RDP
- If $\sup_{\mathcal{D} \sim \mathcal{D}'} \|f(D) - f(D)\| \leq \Delta$; $\mathcal{A}(\cdot) = f(\cdot) + \mathcal{N}(0, \sigma^2 \Delta^2)$ is $(\alpha, \frac{\alpha}{2\sigma^2})$-RDP for any $\alpha > 1$

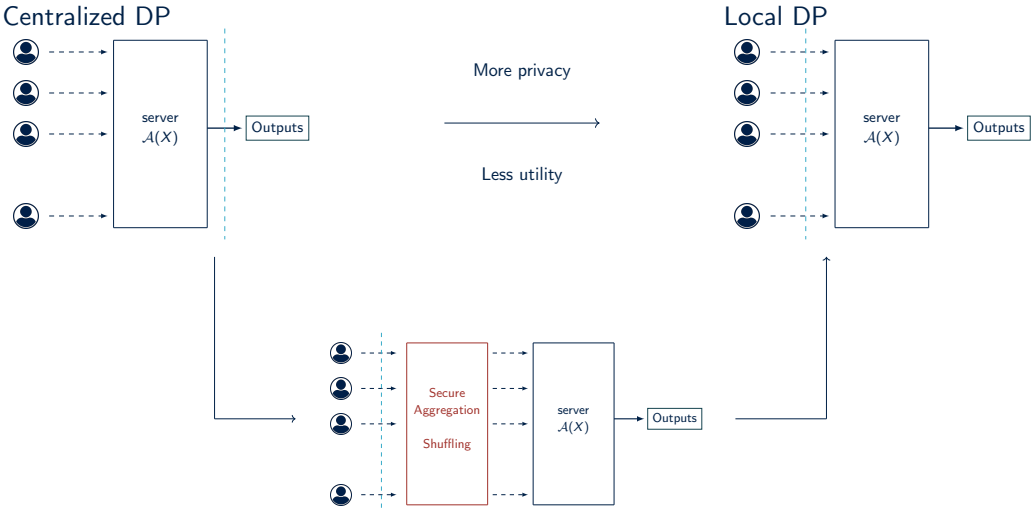
Local or Central DP



Local or Central DP



Local or Central DP



Differentially Private SGD

Algorithm 1 Differentially Private SGD (DP-SGD) [Bassily, Smith, and Thakurta 2014; Abadi et al. 2016]

Initialize $\theta^{(0)} \in \mathbb{R}^p$ (must be independent of $\mathcal{D}$)
for $t = 0, \ldots, T-1$ **do**
  Pick $i_t \in \{1, \ldots, n\}$ uniformly at random
  $\eta^{(t)} \leftarrow (\eta_1^{(t)}, \ldots, \eta_p^{(t)}) \in \mathbb{R}^p$ where each $\eta_j^{(t)} \sim \mathcal{N}(0, \sigma^2\Delta^2)$
  $\theta^{(t+1)} \leftarrow \theta^{(t)} - \gamma^{(t)}\big(\nabla\ell(\theta^{(t)}; x_{i_t}, y_{i_t}) + \eta^{(t)}\big)$
Return $\theta^{(T)}$

- The sensitivity $\Delta = \sup_\theta \sup_{x,y,x',y'} \|\nabla\ell(\theta^{(t)}; x, y) - \nabla\ell(\theta^{(t)}; x', y')\|$ can be controlled by assuming $\ell(\cdot; x, y)$ Lipschitz for all x, y, or using gradient clipping [Abadi et al. 2016] straightforward
- More data (larger n) $\rightarrow$ less noise added to each gradient
- More iterations (larger T) $\rightarrow$ more noise added to each gradient

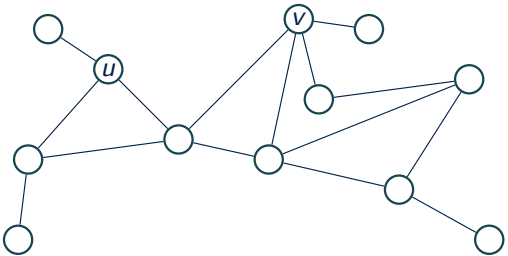
Privacy-utility trade-off of DP-SGD

- Utility analysis: same as non-private SGD (with additional noise due to privacy)
- Privacy analysis: DP-SGD is $(\alpha, \frac{\alpha}{2n^2\sigma^2})$ by subsampled Gaussian mechanism + composition of RDP
  - Note: if one only releases the last iterate, recent refined analysis (e.g., based on Langevin diffusion) show convergence in T of the privacy loss
- In any case, setting σ^2 to satisfy (ε, δ)-DP and choosing T to balance optimization and privacy errors, we get the following suboptimality gap:

Convex, Lipschitz, smooth loss	$\tilde{O}\left(\frac{\sqrt{p}\ln(1/\delta)}{n\varepsilon}\right)$
Convex, Lipschitz, smooth loss, strongly convex F	$\tilde{O}\left(\frac{p\ln(1/\delta)}{n^2\varepsilon^2}\right)$

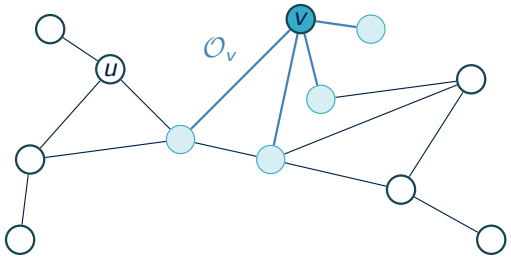
- This is optimal [Bassily, Smith, and Thakurta 2014]: cannot do better without additional assumptions

Decentralized algorithms: good for privacy?



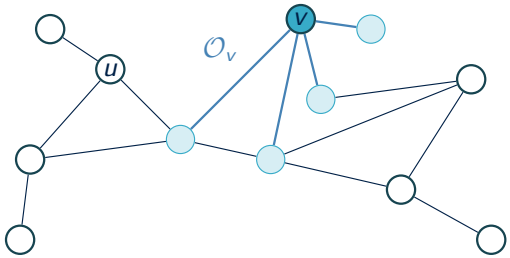
- Decentralized optimization algorithms like decentralized SGD [Lian et al. 2017] [Koloskova et al. 2020] are increasingly popular in ML due to their scalability

Decentralized algorithms: good for privacy?



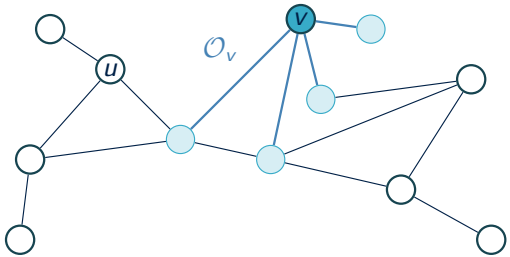
- Decentralized optimization algorithms like decentralized SGD [Lian et al. 2017] [Koloskova et al. 2020] are increasingly popular in ML due to their scalability
- **Folklore: "Decentralized algorithms are good for privacy"**

Decentralized algorithms: good for privacy?



- Decentralized optimization algorithms like decentralized SGD [Lian et al. 2017] [Koloskova et al. 2020] are increasingly popular in ML due to their scalability
- **Folklore: "Decentralized algorithms are good for privacy"**
- **Question:** can we formalize and quantify this gain?

Decentralized algorithms: good for privacy?



- Decentralized optimization algorithms like decentralized SGD [Lian et al. 2017] [Koloskova et al. 2020] are increasingly popular in ML due to their scalability
- **Folklore: "Decentralized algorithms are good for privacy"**
- **Question:** can we formalize and quantify this gain? **Yes!**

Definition of Network Differential Privacy

- Graph of n nodes with edges for communication
- Private database D_u (several contributions for a fixed u)
- Define the observations $\mathcal{O}_v$ of a node u

Network Diffferential Privacy

An algorithm $\mathcal{A}$ is (ε, δ)-network DP if for all pairs of distinct users $u, v \in V$ and all pairs of datasets $D \sim_u D'$, we have

$$\mathbb{P}(\mathcal{O}_v(\mathcal{A}(D))) \leq e^{\varepsilon}\mathbb{P}(\mathcal{O}_v(\mathcal{A}(D'))) + \delta.$$

Definition of Network Differential Privacy

- Graph of n nodes with edges for communication
- Private database D_u (several contributions for a fixed u)
- Define the observations $\mathcal{O}_v$ of a node u

Network Diffferential Privacy

An algorithm $\mathcal{A}$ is (ε, δ)-network DP if for all pairs of distinct users $u, v \in V$ and all pairs of datasets $D \sim_u D'$, we have

$$\mathbb{P}(\mathcal{O}_v(\mathcal{A}(D))) \leq e^{\varepsilon}\mathbb{P}(\mathcal{O}_v(\mathcal{A}(D'))) + \delta.$$

Network Pairwise Differential Privacy

- We will also consider privacy guarantees that are specific to each pair of nodes, rather than uniform over all pairs

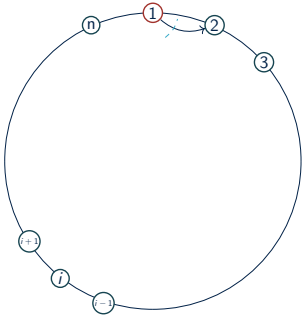
Definition (Pairwise Network DP [Cyffers et al. 2022])

For $f : \mathcal{V} \times \mathcal{V} \to \mathbb{R}^+$, an algorithm $\mathcal{A}$ satisfies (α, f)-Pairwise Network DP (PNDP) if for all pairs of distinct users $u, v \in \mathcal{V}$ and neighboring datasets $\mathcal{D} \sim_u \mathcal{D}'$:

$$D_\alpha\big(\mathcal{O}_v(\mathcal{A}(\mathcal{D})) \,||\, \mathcal{O}_v(\mathcal{A}(\mathcal{D}'))\big) \leq f(u, v)\,. \tag{2}$$

- For comparison with central and local DP baselines, we will report the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u \in \mathcal{V}\setminus\{v\}} f(u, v)$ under the constraint $\overline{\varepsilon} = \max_{v \in \mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$
- Note: $\overline{\varepsilon}_v$ is not a proper privacy guarantee (we simply use it to summarize our gains)

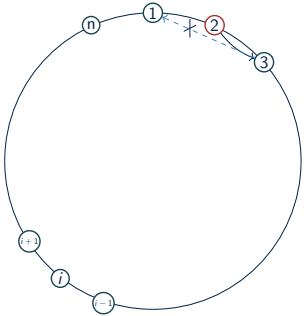
Toy example: Case of a Ring Network



Real Aggregation

- Same privacy-utility trade-off as a *trusted aggregator*
- Gain of $O(1/\sqrt{n})$ compared to LDP

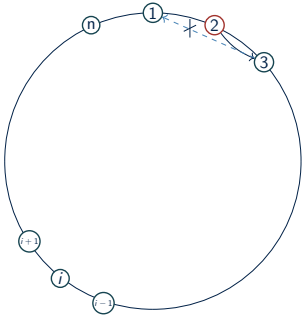
Toy example: Case of a Ring Network



Real Aggregation

- Same privacy-utility trade-off as a *trusted aggregator*
- Gain of $O(1/\sqrt{n})$ compared to LDP

Toy example: Case of a Ring Network



Real Aggregation

- Same privacy-utility trade-off as a *trusted aggregator*
- Gain of $O(1/\sqrt{n})$ compared to LDP

Histogram computation

- keeps trace of the multiset of answers
- amplification by shuffling
- Gain of $\mathcal{O}_\delta(1/\sqrt{n})$ compared to LDP

Private Random walk-based Decentralized SGD

- Consider the standard objective $F(\theta; \mathcal{D}) = \frac{1}{n} \sum_{v=1}^{n} F_v(\theta; \mathcal{D}_v)$
- We consider a decentralized SGD algorithm where the model is updated sequentially by following a random walk, aka incremental gradient [B. Johansson, Rabi, and M. Johansson 2009]

Algorithm 2 Private random walk-based SGD [Cyffers2022a]

Initialize $\theta \in \mathbb{R}^p$
for $t = 1$ to T **do**
  Draw random user $v \sim \mathcal{U}(1, \dots, n)$
  $\eta = [\eta_1, \dots, \eta_p]$, where each $\eta_j \sim \mathcal{N}(0, \sigma^2 \Delta^2)$
  $\theta \leftarrow \theta - \gamma[\nabla_\theta F_v(\theta; \mathcal{D}_v) + \eta]$
return θ

Privacy amplification for walk-based SGD

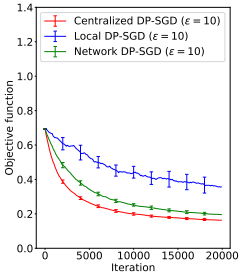
Theorem ([Cyffers2022a], informal)

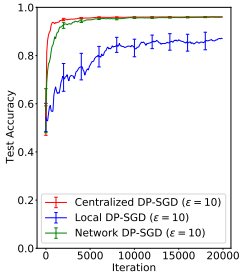
Let $F_1(\cdot; \mathcal{D}_1), \dots, F_n(\cdot; \mathcal{D}_n)$ be convex and smooth. Given $\alpha > 1$, $\varepsilon > 0$, let $T = \tilde{\Omega}(n^2)$ and σ^2 be such that private random walk-based decentralized SGD on the complete graph satisfies (α, ε)-local RDP. Then the algorithm also satisfies $(\alpha, \frac{\ln^2 n}{n}\varepsilon)$-network DP.

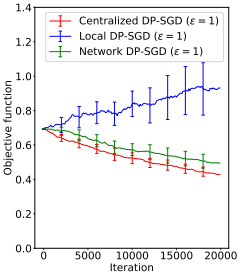
- With NDP, we recover the privacy-utility trade-off of DP-SGD under central DP up to a log factor
- For $T = o(n^2)$, the amplification effect is still strong and can be computed numerically
- Utility analysis: same as DP-SGD
- Privacy analysis: leverages privacy amplification by iteration **[Feldman18]** and exploits the randomness of the walk through "weak convexity" of Rényi divergence
- We show some robustness to collusion (albeit with smaller privacy amplification)

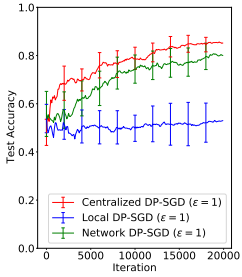
Case of a Complete graph - SGD

- The amplification by decentralization lead to significant utility gains.
- Experiments with σ set to ensure $\varepsilon = 10$ (left) or $\varepsilon = 1$ (right) and $\delta = 10^{-6}$.









Ongoing work: arbitrary graphs

- Let W be a bistochastic matrix encoding the transitions of the random walk and consider Pairwise Network DP
- The privacy is composed of a global term corresponding to the complete graph and a constant graph-dependent factor.

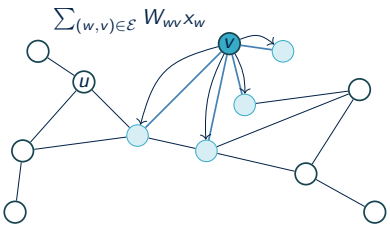
Theorem (Privacy loss for RW in arbitrary graph (informal))

Let $\alpha > 1$, $\varepsilon > 0$ and σ^2 be such that the RW DP-SGD on the graph of adjacency matrix W is (α, ε)-local RDP for a step. After T (large enough) steps, the algorithm is (α, f)-PNDP for

$$f(u,v) \leq \frac{CT}{n\varepsilon}\left(\frac{\ln(T)}{n} - \ln(I - W + \frac{1}{n}11^\top)_{uv}\right)$$

- Interpretation as leakage within communities? Relative importance of the two terms $\rightarrow$ requires tight bounds on RW mixing times.

Gossip protocols



Algorithm 3 Gossip_Averaging$\left(\{x_v\}_{v\in\mathcal{V}}, W, K\right)$

for all nodes v in parallel **do**
  $x_v^0 \leftarrow x_v$
for $k = 0$ to $K-1$ **do**
  for all nodes v in parallel **do**
    $x_v^{k+1} \leftarrow \sum_{w\in\mathcal{N}_v} W_{v,w}x_w^k$, where $\mathcal{N}_v = \{w : W_{v,w} > 0\}$
return $x_1^K, \ldots, x_n^K$

- Random walk-based SGD is sequential (no parallel computation)
- A popular parallel alternative is gossip-based decentralized SGD [Lian et al. 2017] [Koloskova et al. 2020], which builds upon gossip averaging [Boyd et al. 2006]

Gossip-based Decentralized SGD

- Consider again $F(\theta; \mathcal{D}) = \frac{1}{n}\sum_{v=1}^{n} F_v(\theta; \mathcal{D}_v)$ with $F_v(\theta; \mathcal{D}_v) = \frac{1}{|\mathcal{D}_V|}\sum_{(x_v, y_v)\in\mathcal{D}_v} \ell(\theta; x_v, y_v)$

Algorithm 4 Gossip-based decentralized SGD

Initialize $\theta_1^{(0)}, \dots, \theta_n^{(0)} \in \mathbb{R}^p$
for $t = 0$ to $T-1$ **do**
  for all nodes v in parallel **do**
    $\hat{\theta}_v^t \leftarrow \theta_v^t - \gamma \nabla_\theta \ell(\theta_v^t; x_v^t, y_v^t)$ where $(x_v^t, y_v^t) \sim \mathcal{D}_v$
  $\theta_v^{t+1} \leftarrow \text{Gossip\_Averaging}\big(\{\hat{\theta}_v^t\}_{v\in\mathcal{V}}, W, K\big)$
return $\theta_1^T, \dots, \theta_n^T$

- To improve the dependence on the topology in the convergence rate we actually use accelerated gossip [Berthier, Bach, and Gaillard 2020]

Private Gossip-based Decentralized SGD

- To make the algorithm private, we add Gaussian noise before gossiping

Algorithm 5 Muffliato$\big(\{x_v\}_{v\in\mathcal{V}}, W, K, \sigma^2\big)$

for all nodes v in parallel **do**
  $\tilde{x}_v^0 \leftarrow x_v + \eta_v$ where $\eta_v \sim \mathcal{N}(0, \sigma^2)$
$x_1^K, \dots, x_n^K \leftarrow$ Gossip_Averaging$\big(\{\tilde{x}_v^0\}_{v\in\mathcal{V}}, W, K\big)$
return $x_1^K, \dots, x_n^K$

Algorithm 6 Private gossip-based decentralized SGD

Initialize $\theta_1^{(0)}, \dots, \theta_n^{(0)} \in \mathbb{R}^p$
for $t = 0$ to $T-1$ **do**
  for all nodes v in parallel **do**
    $\hat{\theta}_v^t \leftarrow \theta_v^t - \gamma \nabla_\theta \ell(\theta_v^t; x_v^t, y_v^t)$ where $(x_v^t, y_v^t) \sim \mathcal{D}_v$
  $\theta_v^{t+1} \leftarrow$ Muffliato$\big(\{\hat{\theta}_v^t\}_{v\in\mathcal{V}}, W, K, \nu^2\sigma^2\big)$
return $\theta_1^T, \dots, \theta_n^T$

Privacy-utility analysis of Private Gossip Averaging

Theorem ([Cyffers et al. 2022])

After K iterations, Private Gossip Averaging is (α, f)-PNDP with

$$f(u,v) = \frac{\alpha\Delta^2}{2\sigma^2} \sum_{k=0}^{K-1} \sum_{w:\{v,w\}\in\mathcal{E}} \frac{(W^k)_{u,w}^2}{\|(W^k)_{w,:}\|^2}$$

$$\leq \frac{\alpha\Delta^2 n}{2\sigma^2} \max_{\{v,w\}\in\mathcal{E}} W_{v,w}^{-2} \sum_{k=1}^{K} \mathbb{P}(X^k = v | X^0 = u)^2,$$

where $(X^k)_k$ is the random walk on graph G, with transitions W.

- As desired, this exhibits the fact that, for two nodes u and v, privacy guarantees improve with their "distance" in the graph

Privacy-utility analysis of Private Gossip Averaging

Theorem ([Cyffers et al. 2022])

Let λ_W be the spectral gap of W. Private Gossip Averaging verifies, for any $t \geq T^{\text{stop}}$:

$$\frac{1}{2n} \sum_{v \in \mathcal{V}} \mathbb{E}[\|x_v^t - \bar{x}\|^2] \leq \frac{3\sigma^2}{n}$$

with:

$$T^{\text{stop}} = \frac{1}{\sqrt{\lambda_W}} \ln \left(\frac{n}{\sigma^2} \max \left(\sigma^2, \frac{1}{n} \sum_{v \in \mathcal{V}} \|x_v - \bar{x}\|^2 \right) \right).$$

- There is an incompressible error due to noise so it is useless to run gossip averaging for too long

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha \Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha \Delta^2}{2n\varepsilon}$

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha\Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha\Delta^2}{2n\varepsilon}$
- Combining the two previous results and setting the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v)$ to satisfy $\overline{\varepsilon} = \max_{v\in\mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$, we get (ignoring log terms):

Graph	Arbitrary
Utility	$\frac{\alpha\Delta^2 d}{n^2\varepsilon\sqrt{\lambda_W}}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor (d: max degree)

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha\Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha\Delta^2}{2n\varepsilon}$
- Combining the two previous results and setting the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v)$ to satisfy $\overline{\varepsilon} = \max_{v\in\mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$, we get (ignoring log terms):

Graph	Arbitrary	Complete
Utility	$\frac{\alpha\Delta^2 d}{n^2\varepsilon\sqrt{\lambda_W}}$	$\frac{\alpha\Delta^2}{n\varepsilon}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor (d: max degree)

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha\Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha\Delta^2}{2n\varepsilon}$
- Combining the two previous results and setting the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v)$ to satisfy $\overline{\varepsilon} = \max_{v\in\mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$, we get (ignoring log terms):

Graph	Arbitrary	Complete	Ring
Utility	$\frac{\alpha\Delta^2 d}{n^2\varepsilon\sqrt{\lambda_W}}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2}{n\varepsilon}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor (d: max degree)

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha\Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha\Delta^2}{2n\varepsilon}$
- Combining the two previous results and setting the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v)$ to satisfy $\overline{\varepsilon} = \max_{v\in\mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$, we get (ignoring log terms):

Graph	Arbitrary	Complete	Ring	D-Torus
Utility	$\frac{\alpha\Delta^2 d}{n^2\varepsilon\sqrt{\lambda_W}}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2 D}{n^{2-1/D}\varepsilon}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor (d: max degree)

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha\Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha\Delta^2}{2n\varepsilon}$
- Combining the two previous results and setting the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v)$ to satisfy $\overline{\varepsilon} = \max_{v\in\mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$, we get (ignoring log terms):

Graph	Arbitrary	Complete	Ring	D-Torus	Expander
Utility	$\frac{\alpha\Delta^2 d}{n^2\varepsilon\sqrt{\lambda_W}}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2 D}{n^{2-1/D}\varepsilon}$	$\frac{\alpha\Delta^2}{n^2\varepsilon}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor (d: max degree)
- Some graphs (e.g., expanders) make this constant: we get privacy **and** efficiency!

Privacy-utility analysis

- Recall central DP achieves $\frac{\alpha\Delta^2}{2n^2\varepsilon}$ and local DP achieves $\frac{\alpha\Delta^2}{2n\varepsilon}$
- Combining the two previous results and setting the mean privacy loss $\overline{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v)$ to satisfy $\overline{\varepsilon} = \max_{v\in\mathcal{V}} \overline{\varepsilon}_v \leq \varepsilon$, we get (ignoring log terms):

Graph	Arbitrary	Complete	Ring	D-Torus	Expander
Utility	$\frac{\alpha\Delta^2 d}{n^2\varepsilon\sqrt{\lambda_W}}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2}{n\varepsilon}$	$\frac{\alpha\Delta^2 D}{n^{2-1/D}\varepsilon}$	$\frac{\alpha\Delta^2}{n^2\varepsilon}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor (d: max degree)
- Some graphs (e.g., expanders) make this constant: we get privacy **and** efficiency!
- Note: we also have extensions to time-varying graphs and randomized gossip

Gossip-based Decentralized SGD

Theorem ([Cyffers et al. 2022])

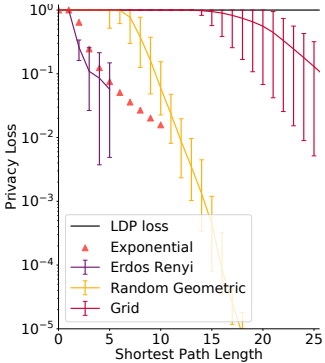
Let F be μ-strongly convex, F_v be L-smooth and $\mathbb{E}[\|\nabla \ell(\theta^\star; x_v, y_v) - \nabla F(\theta^\star)\|^2] \leq \rho_v^2$. Let $\bar{\rho}^2 = \frac{1}{n}\sum_{v\in\mathcal{V}} \rho_v^2$. For any $\varepsilon > 0$, and appropriate choices of T and K, there exists f such that the algorithm is (α, f)-PNDP, with:

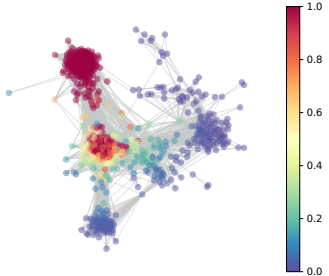
$$\forall v \in \mathcal{V}\,, \quad \bar{\varepsilon}_v \leq \varepsilon \quad \text{and} \quad \mathbb{E}[F(\bar{\theta}^{1:T}) - F(\theta^\star)] \leq \tilde{\mathcal{O}}\left(\frac{\alpha \Delta^2 d_v}{n\mu\varepsilon\sqrt{\lambda_W}} + \frac{\bar{\rho}^2}{nL}\right),$$

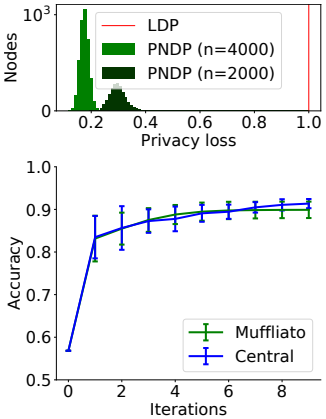
where d_v is the degree of node v and λ_W is the spectral gap associated with W.

- The term $\frac{\bar{\rho}^2}{nL}$ is privacy-independent and dominated by the first term
- The first term has the same form as before, so same conclusions apply

Muffliato – experimental results







Left: pairwise NDP loss on synthetic graphs Middle: Privacy loss of *Muffliato* from a node chosen at random on a Facebook ego graph Right: Privacy loss and utility of *Muffliato*-GD compared to a baseline based on a trusted aggregator.

Conclusion & Perspectives

Take-home message

- Decentralized optimization can amplify differential privacy guarantees, providing a new incentive for using such approaches beyond the usual motivation of scalability

Conclusion & Perspectives

Take-home message

- Decentralized optimization can amplify differential privacy guarantees, providing a new incentive for using such approaches beyond the usual motivation of scalability

Future work

- Privacy and utility guarantees for random walk-based decentralized SGD on arbitrary graphs [B. Johansson, Rabi, and M. Johansson 2009], possibly with multiple parallel walks [**Hendrikx2022a**]
- Capturing the redundancy in gossip-based communication (i.e., correlated noise) to further improve privacy guarantees
- Fair comparison between random-walk and gossip algorithms
- Lower bounds on privacy loss.