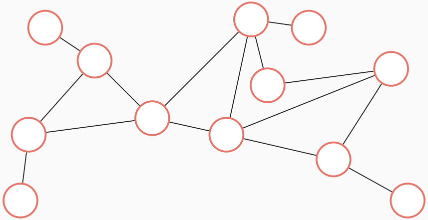
PRIVACY IN DECENTRALIZED MACHINE LEARNING

Aurélien Bellet (Inria)

Joint work with **Edwige Cyffers**, Abdellah El Mrini, Mathieu Even, Laurent Massoulié, Jalaj Upadhyay

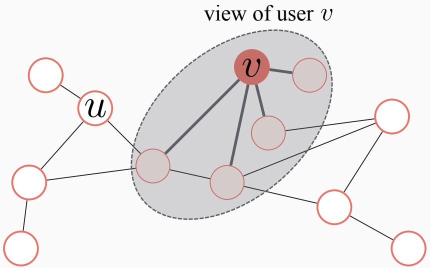
3rd Workshop on Principles of Distributed Learning (PODL)
June 21, 2024

DECENTRALIZED ALGORITHMS: GOOD FOR PRIVACY?



- Decentralized learning, where users communicate along the edges of a graph, is increasingly popular for its scalability

DECENTRALIZED ALGORITHMS: GOOD FOR PRIVACY?



- Decentralized learning, where users communicate along the edges of a graph, is increasingly popular for its scalability
- **Folklore: "Decentralized learning algorithms are good for privacy"**
- **Question:** is this claim really true? can we formalize and quantify these gains?

Yes! but decentralization alone is not sufficient

OUTLINE

1. Privacy Attack on Decentralized SGD
2. Differential Privacy for Decentralized Algorithms
3. Private Decentralized SGD
4. Conclusion & Perspectives

Privacy Attack on Decentralized SGD

[El Mrini et al., 2024]

- Consider a connected graph $G = (\mathcal{V}, \mathcal{E})$ on a set of $|\mathcal{V}| = n$ users (nodes), where each user $v \in \mathcal{V}$ holds a local dataset $\mathcal{D}_v$ (assume $\mathcal{D}_v = \{x_v\}$ for now)
- A gossip matrix over G is a symmetric stochastic matrix $W \in [0,1]^{n \times n}$ for which $W_{v,w} > 0$ implies $\{v, w\} \in \mathcal{E}$ or $v = w$

Algorithm $\textsc{Gossip\_Averaging}(\{x_v\}_{v \in \mathcal{V}}, W, K)$ [Boyd et al., 2006]

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**
    $x_v^{k+1} \leftarrow \sum_{w \in \mathcal{N}_v} W_{v,w} x_w^k, \quad$ where $\mathcal{N}_v = \{w : W_{v,w} > 0\}$

- Convergence to the average value at a rate of order $e^{-t\lambda_W}$ where λ_W is the spectral gap of W (note: improved rate of $e^{-t\sqrt{\lambda_W}}$ with accelerated gossip [Berthier et al., 2020])

- Consider now that each user v has a local objective $F_v(\theta; \mathcal{D}_v) = \frac{1}{|\mathcal{D}_v|} \sum_{x_v \in \mathcal{D}_v} \ell(\theta; x_v)$ and we wish to minimize $F(\theta; \mathcal{D}) = \frac{1}{n} \sum_{v=1}^{n} F_v(\theta; \mathcal{D}_v)$

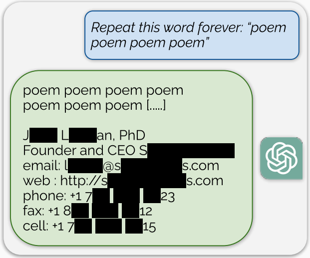
Algorithm Decentralized SGD [Lian et al., 2017, Koloskova et al., 2020]

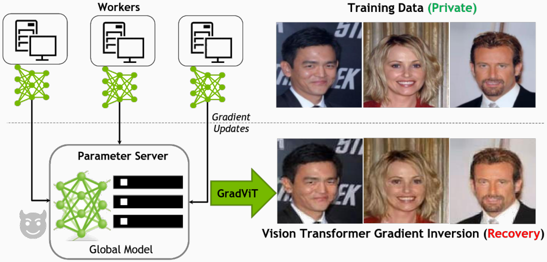
Initialize $\theta_1^{(0)}, \ldots, \theta_n^{(0)} \in \mathbb{R}^p$
for $t = 0$ to $T - 1$ **do**
  for all nodes v in parallel **do**
    $\hat{\theta}_v^t \leftarrow \theta_v^t - \gamma \nabla_\theta \ell(\theta_v^t; x_v^t)$ where $x_v^t \sim \mathcal{D}_v$
  $\theta_v^{t+1} \leftarrow \textsc{Gossip\_Averaging}\big(\{\hat{\theta}_v^t\}_{v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

- Various convergence results exist for convex and nonconvex objectives, which again exhibit a dependence in the spectral gap λ_W

RECONSTRUCTION ATTACKS AGAINST MACHINE LEARNING MODELS

- ML models are susceptible to various attacks on data privacy
- We focus on reconstruction attacks, which aim to extract training data points from the model, for instance sensitive text from large language models [Nasr et al., 2023]
- Of particular interest to us are gradient inversion attacks, which reconstruct data points from their gradients [Geiping et al., 2020, Hatamizadeh et al., 2022]



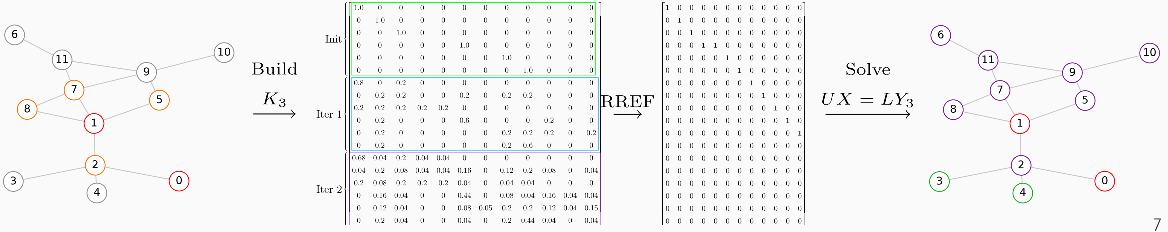


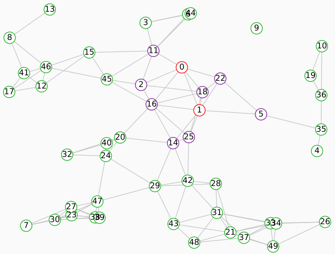
THREAT MODEL AND ATTACK GOAL

- Attackers are a subset of nodes $\mathcal{A} \subset \mathcal{V}$: they share the knowledge among them but are assumed to be honest-but-curious
- The attackers know their own data, the graph G and the gossip matrix W, and observe the messages they receive
- **Attack goal:** reconstruct the private data of other nodes
- Note: it is easy to attack neighbors $\mathcal{N}(\mathcal{A})$ as they leak their value/gradient directly to the attackers [Pasquini et al., 2023]; the question is whether it is possible to reconstruct the data of more distant nodes

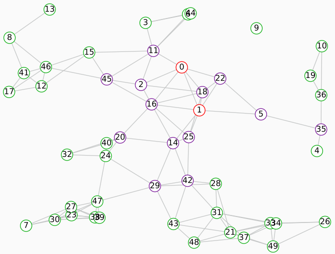
ATTACK ON GOSSIP AVERAGING

- **Key idea:** the messages received form a system of linear equations where the unknowns are private values $X = (x_1, \ldots, x_n)$ and the coefficients depend on W
- For T iterations of gossip, we can denote this system as $K_T X = Y_T$:
  - Y_T: observation vector with the $|\mathcal{A}|$ values of the attackers and the $T|\mathcal{N}(\mathcal{A})|$ messages
  - K_T: knowledge matrix where each row encodes the linear combination of private values corresponding to each entry of Y_T
- We then factorize $K_T = L^{-1}U$ where U is the RREF of K_T and L is such that $UX = LY_T$

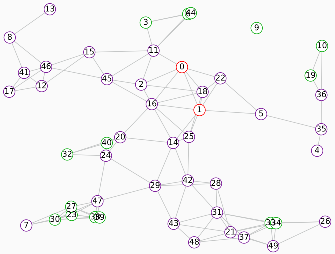




After 1 iteration



After 4 iterations



After 8 iterations

Figure 1: Reconstruction after a different number of steps of gossip averaging. Attackers are in red, reconstructed nodes in purple, and non-reconstructed ones in green. The graph is a random geometric graph of 50 nodes uniformly drawn from the unit square and a radius of 0.2.

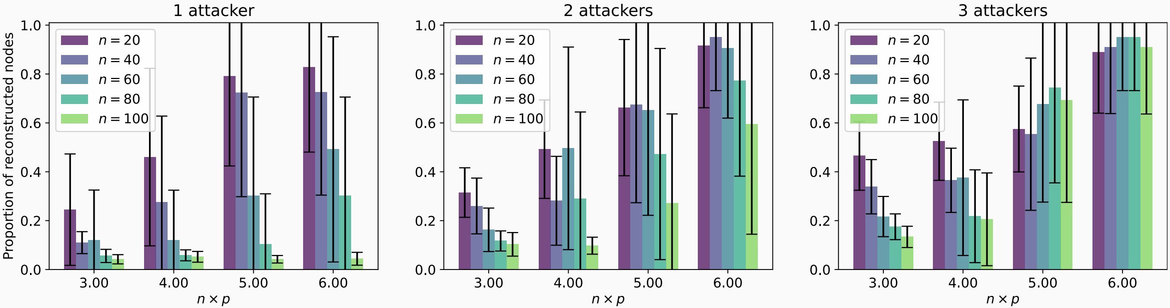


Figure 2: Average fraction of reconstructed nodes in Erdös-Rényi graphs with a different number of nodes n and edge probability p, for 1, 2 or 3 attacker nodes. Error bars give the standard deviations, computed over 20 random graphs.

RESULTS ON REAL GRAPHS

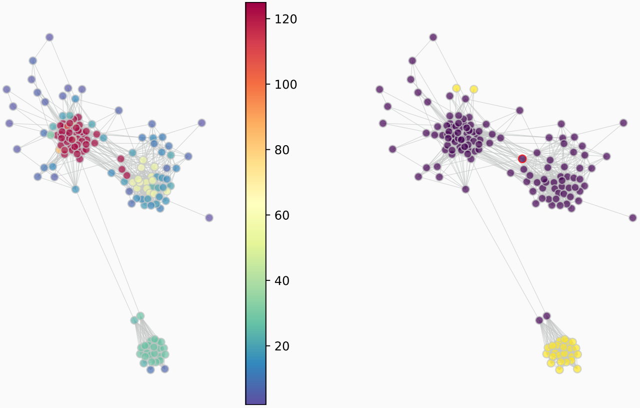


Figure 3: Reconstruction attack on the Facebook Ego Graph 414. Left: each node is colored by the number of nodes it can reconstruct among the 147 other nodes. Right: detailed view of the case where the node circled in red is the attacker, with reconstructed nodes shown in purple and non-reconstructed ones in yellow.

- For simplicity, we focus on the case where each node holds a single data point and a single gossip averaging step is performed between each gradient update (i.e., $K = 1$)
- Our attack proceeds in two steps: first reconstruct the gradients of nodes, then reconstruct the data points from the gradients (using known attacks)
- To reconstruct gradients, we build upon the attack on gossip averaging but need to address several challenges:
    1. Gradients change at each iteration $\rightarrow$ too many unknowns!
    2. Users share model parameters (not the gradients), and attackers know their own contributions $\rightarrow$ K_T and Y_T need to be adapted

- We model the gradients of a node as the combination of a fixed and random components:

$$g_v^t = -\gamma \nabla_\theta L(\theta_v^t, x_v^t) = g_v + N_v^t \quad \text{where } \mathbb{E}(N_v^t) = 0 \text{ and } \mathbb{V}(N_v^t) = \sigma^2$$

(this is not the case in practice but our attack generally works well when gradients change sufficiently slowly)

- We adapt the construction of K_T and Y_T by deriving a closed-form update for $\hat{\theta}_v^t$ which separates the contribution of attacker nodes from those of target nodes
- Finally, reconstructing the gradients reduces to solving a generalized least square problem $K_T g + \epsilon_T = Y_T$ where ϵ_T is a noise term with non-diagonal covariance

Cifar10, logistic regression, learning rate 10^{-4}

MNIST, convnet, learning rate 10^{-6}, gradient inversion from [Geiping et al., 2020]

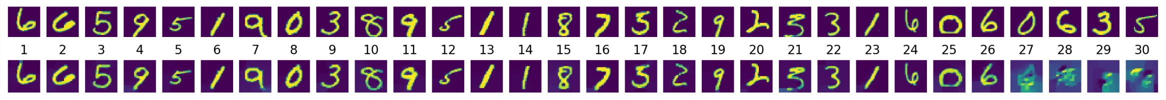
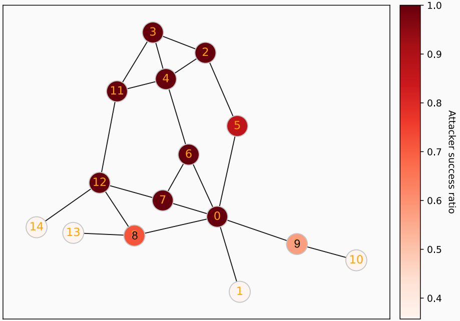


Figure 4: Reconstruction attack on D-GD for a line graph with 31 nodes where the attacker lies at an extremity. The first (resp. second) row shows the true (resp. reconstructed) inputs of the 30 other nodes ordered by their distance to the attacker.



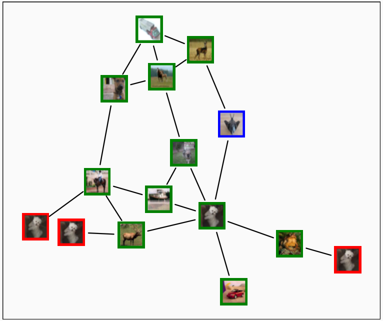
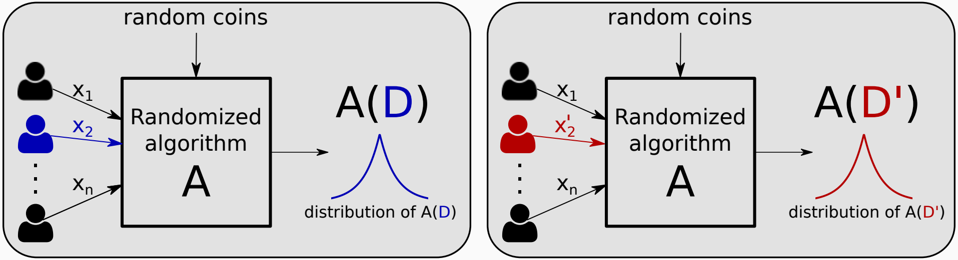


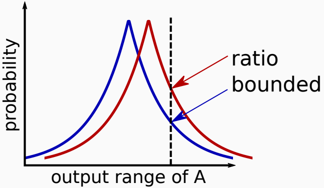
Figure 5: Reconstruction attacks on D-GD for the Florentine graph (Cifar10, logistic regression model, learning rate 10^{-5}). Left: the color of each node represents the success rate when that node is the attacker. The success rate is the fraction of nodes where PSNR ≥ 10 (averaged over 10 experiments). Right: example where the attacker is node 5 (in blue). Nodes with green borders are accurately reconstructed, the ones with red borders are not.

Differential Privacy for Decentralized Algorithms

[Cyffers and Bellet, 2022]



- 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



Definition (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, $D_\alpha(X \,||\, Y)$ is the Rényi divergence of order α:

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

- Conversion to standard (ε, δ)-DP: (α, ε)-RDP implies $(\varepsilon + \frac{\ln(1/\delta)}{\alpha - 1}, \delta)$-DP for any $\delta \in (0, 1)$

- RDP is robust to auxiliary knowledge, as seen by its Bayesian interpretation:
  - Consider an adversary who seeks to infer whether the dataset is $\mathcal{D}$ or $\mathcal{D}'$
  - The adversary has prior knowledge p and observes $X \sim \mathcal{A}(\mathcal{D})$
  - Let the r.v. $R_{prior} = \frac{p(\mathcal{D}')}{p(\mathcal{D})}$ and $R_{post} = \frac{p(\mathcal{D}'|X)}{p(\mathcal{D}|X)} = \frac{p(X|\mathcal{D}')p(\mathcal{D}')}{p(X|\mathcal{D})p(\mathcal{D})}$ for $X \sim \mathcal{A}(\mathcal{D})$
  - RDP bounds the α-th moment of $\frac{R_{post}}{R_{prior}}$ (for $\alpha \rightarrow \infty$, we recover "pure" ε-DP)
  - "The adversary does not know much more after observing the output of the algorithm"
- Immunity to post-processing: for any g, if $\mathcal{A}(\cdot)$ is (α, ε)-RDP, then so is $g(\mathcal{A}(\cdot))$
- 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 $(\alpha, \varepsilon_1 + \varepsilon_2)$-RDP $\rightarrow$ simpler and tighter than composition for (ε, δ)-DP

ENFORCING RDP WITH THE GAUSSIAN MECHANISM

- Consider f taking as input a dataset and returning a p-dimensional real vector
- Denote its sensitivity by $\Delta = \max_{\mathcal{D}\sim\mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_2$

Theorem (Gaussian mechanism)

Let $\sigma > 0$. The algorithm $\mathcal{A}(\cdot) = f(\cdot) + \mathcal{N}(0, \sigma^2\Delta^2)$ satisfies $(\alpha, \frac{\alpha}{2\sigma^2})$-RDP for any $\alpha > 1$.

- DP induces a privacy-utility trade-off, here in terms of the variance of the estimate

- The classic trust model of central DP model considers a trusted curator to collect and process raw data $\rightarrow$ the output $\mathcal{A}(\mathcal{D})$ is only the final result
- Central DP is good for utility but is an unrealistic trust model in applications where many users contribute sensitive data, as in decentralized learning
- A common alternative is local DP, where each user locally randomizes its contributions $\rightarrow$ the output of $\mathcal{A}(\mathcal{D})$ consists of all messages sent by all users
- Unfortunately local DP induces a large cost in utility: for averaging n private p-dimensional values in ball of radius Δ under (α, ε)-RDP, we have

$$\mathbb{E}[\|x^{\text{out}} - \bar{x}\|^2] = \Theta\left(\frac{\alpha p \Delta^2}{n\varepsilon}\right) \text{ for local DP}, \quad \text{and} \quad \mathbb{E}[\|x^{\text{out}} - \bar{x}\|^2] = \Theta\left(\frac{\alpha p \Delta^2}{n^2\varepsilon}\right) \text{ for central DP}$$

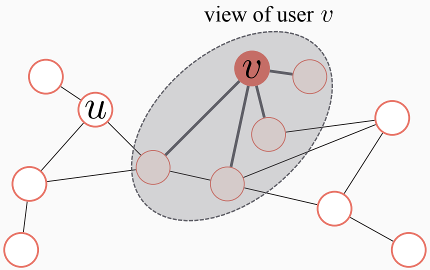
$\rightarrow$ we propose a trust model suitable for decentralized algorithms allowing better utility

- Let $\mathcal{O}_v$ be the set of messages sent and received by party v
- Denote by $\mathcal{D} \sim_u \mathcal{D}'$ two datasets $\mathcal{D} = (\mathcal{D}_1, \ldots, \mathcal{D}_u, \ldots, \mathcal{D}_n)$ and $\mathcal{D}' = (\mathcal{D}_1, \ldots, \mathcal{D}'_u, \ldots, \mathcal{D}_n)$ that differ only in the local dataset of user u

Definition (Network DP [Cyffers and Bellet, 2022])

An algorithm $\mathcal{A}$ satisfies (α, ε)-Network DP (NDP) if for all pairs of distinct users $u, v \in \mathcal{V}$ and neighboring datasets $\mathcal{D} \sim_u \mathcal{D}'$:

$$D_\alpha(\mathcal{O}_v(\mathcal{A}(\mathcal{D})) \,||\, \mathcal{O}_v(\mathcal{A}(\mathcal{D}'))) \leq \varepsilon \,.$$



- This is a relaxation of local DP: if $\mathcal{O}_v$ contains the full transcript of messages, then network DP boils down to local DP

- 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 $\mathcal{D} \sim_u \mathcal{D}'$:

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

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

Private Decentralized SGD

[Cyffers et al., 2022, Cyffers et al., 2024]

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

Algorithm $\textsc{Private\_Gossip\_Averaging}\big(\{x_v\}_{v\in\mathcal{V}}, W, K, \sigma^2\big)$

for all nodes v in parallel **do**

  $\tilde{x}_v^0 \leftarrow x_v + \eta_v$ where $\eta_v \sim \mathcal{N}(0, \sigma^2)$

$x_1^K, \dots, x_n^K \leftarrow \textsc{Gossip\_Averaging}\big(\{\tilde{x}_v^0\}_{v\in\mathcal{V}}, W, K\big)$

return $x_1^K, \dots, x_n^K$

Algorithm Private Decentralized SGD [Cyffers et al., 2022]

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**

    $\hat{\theta}_v^t \leftarrow \theta_v^t - \gamma \nabla_\theta \ell(\theta_v^t; x_v^t)$ where $x_v^t \sim \mathcal{D}_v$

  $\theta_v^{t+1} \leftarrow \textsc{Private\_Gossip\_Averaging}\big(\{\hat{\theta}_v^t\}_{v\in\mathcal{V}}, W, K, \gamma^2\sigma^2\Delta^2\big)$

return $\theta_1^T, \dots, \theta_n^T$

Theorem ([Cyffers et al., 2022])

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

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

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

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

- Recall central DP achieves $O(\frac{\alpha p \Delta^2}{n^2 \varepsilon})$ and local DP achieves $O(\frac{\alpha p \Delta^2}{n \varepsilon})$
- Setting the mean privacy loss $\bar{\varepsilon}_v = \frac{1}{n} \sum_{u \in \mathcal{V} \setminus \{v\}} f(u, v)$ to satisfy $\bar{\varepsilon} = \max_{v \in \mathcal{V}} \bar{\varepsilon}_v \leq \varepsilon$, for private gossip averaging we get (ignoring log terms):

Graph	Arbitrary	Complete	Ring	Expander
Utility (MSE)	$\frac{\alpha p \Delta^2 d}{n^2 \varepsilon \sqrt{\lambda_W}}$	$\frac{\alpha p \Delta^2}{n \varepsilon}$	$\frac{\alpha p \Delta^2}{n \varepsilon}$	$\frac{\alpha p \Delta^2}{n^2 \varepsilon}$

- We match the utility of central DP up to an additional $d/\sqrt{\lambda_W}$ factor, where d is the max degree and λ_W of the spectral gap of W
- 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

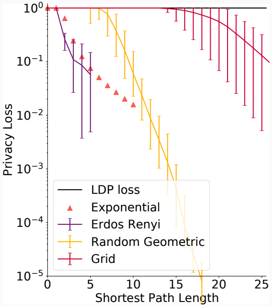
Theorem ([Cyffers et al., 2022])

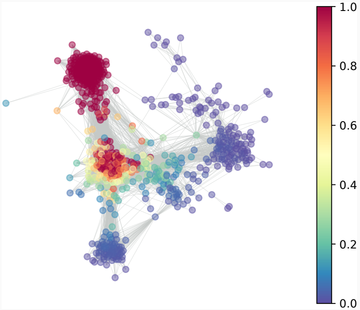
Let F be μ-strongly convex, F_v be L-smooth and $\mathbb{E}[\|\nabla\ell(\theta^\star; x_v) - \nabla F(\theta^\star)\|^2] \leq \rho_v^2$. Let $\bar{\rho}^2 = \frac{1}{n}\sum_{v\in\mathcal{V}} \rho_v^2$. For any $\varepsilon > 0$, and appropriate choices of T and K, there exists f such that the algorithm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

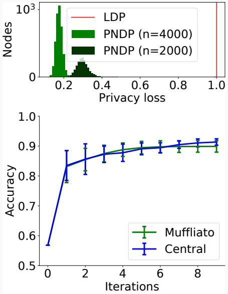
$$\forall v \in \mathcal{V}\,, \quad \bar{\varepsilon}_v = \frac{1}{n}\sum_{u\in\mathcal{V}\setminus\{v\}} f(u,v) \leq \varepsilon \quad \text{and} \quad \mathbb{E}[F(\bar{\theta}^{1:T}) - F(\theta^\star)] \leq \tilde{\mathcal{O}}\left(\frac{\alpha p \Delta^2 d L}{n^2\mu^2\varepsilon\sqrt{\lambda_W}}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

- 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!
- In particular, with an expander graph, we nearly match the privacy-utility trade-off of centralized SGD with a trusted curator (up to a factor L/μ)

EMPIRICAL ILLUSTRATION

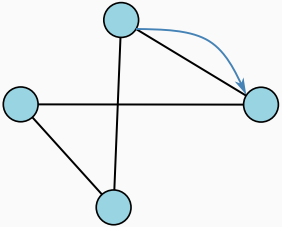






- Users get local DP guarantees w.r.t. their direct neighbors but stronger privacy w.r.t. to other users depending on their distance and the mixing properties of the graph
- This fits the privacy expectations of users in many use-cases (e.g., social networks)
- For learning, we can randomize the graph after each local computation step to make the privacy loss concentrate!

- An alternative to gossip is to consider a decentralized SGD algorithm where the model is updated sequentially by following a random walk [Johansson et al., 2009]



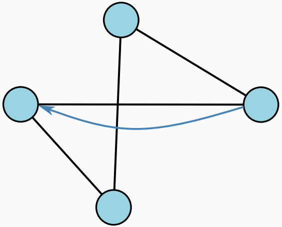
Algorithm Private random walk-based SGD [Cyffers et al., 2024]

Initialize $\theta^0 \in \mathbb{R}^p$ and starting user v^0
for $t = 0$ to $T-1$ **do**
  $\theta^{t+1} \leftarrow \theta^t - \gamma(\nabla_\theta \ell(\theta^t; x^t) + \eta)$ where $x^t \sim \mathcal{D}_{v_t}$ and $\eta \sim \mathcal{N}(0, \sigma^2\Delta^2)$
  Draw $u \sim W_{v_t}$ and send θ^{t+1} to user u
  $v^{t+1} \leftarrow u$
return θ^T

- No redundant communication and no need for users to be always available
- Privacy analysis relies on privacy amplification by iteration [Feldman et al., 2018]

BONUS: PRIVATE RANDOM WALK-BASED DECENTRALIZED SGD

- An alternative to gossip is to consider a decentralized SGD algorithm where the model is updated sequentially by following a random walk [Johansson et al., 2009]

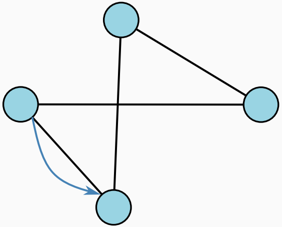


Algorithm Private random walk-based SGD [Cyffers et al., 2024]

Initialize $\theta^0 \in \mathbb{R}^p$ and starting user v^0
for $t = 0$ to $T-1$ **do**
  $\theta^{t+1} \leftarrow \theta^t - \gamma(\nabla_\theta \ell(\theta^t; x^t) + \eta)$ where $x^t \sim \mathcal{D}_{v_t}$ and $\eta \sim \mathcal{N}(0, \sigma^2\Delta^2)$
  Draw $u \sim W_{v_t}$ and send θ^{t+1} to user u
  $v^{t+1} \leftarrow u$
return θ^T

- No redundant communication and no need for users to be always available
- Privacy analysis relies on privacy amplification by iteration [Feldman et al., 2018]

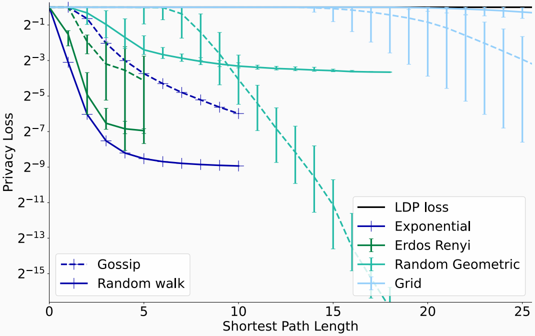
- An alternative to gossip is to consider a decentralized SGD algorithm where the model is updated sequentially by following a random walk [Johansson et al., 2009]



Algorithm Private random walk-based SGD [Cyffers et al., 2024]

Initialize $\theta^0 \in \mathbb{R}^p$ and starting user v^0
for $t = 0$ to $T-1$ **do**
  $\theta^{t+1} \leftarrow \theta^t - \gamma(\nabla_\theta \ell(\theta^t; x^t) + \eta)$ where $x^t \sim \mathcal{D}_{v_t}$ and $\eta \sim \mathcal{N}(0, \sigma^2\Delta^2)$
  Draw $u \sim W_{v_t}$ and send θ^{t+1} to user u
  $v^{t+1} \leftarrow u$
return θ^T

- No redundant communication and no need for users to be always available
- Privacy analysis relies on privacy amplification by iteration [Feldman et al., 2018]



- For averaging, at same level of utility, random-walk incurs a smaller privacy loss for close enough nodes than gossip
- For SGD, the advantage is even more pronounced (better progress with many noisy steps than a small number of less noisy steps)

Conclusion & Perspectives

Take-home messages

- Vanilla Decentralized SGD does not protect the privacy of nodes: we show for the first time that attackers can reconstruct data from distant nodes
- Decentralized learning can amplify differential privacy guarantees, providing a new incentive for using such approaches beyond the usual motivation of scalability

Perspectives

- Tighter privacy accounting for decentralized algorithms
- Complete characterization of reconstructible nodes using explicit graph quantities
- More general attacks, e.g. able to handle randomness in communications and/or a partially unknown graph

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

KNOWLEDGE MATRIX AND OBSERVATION VECTOR FOR DECENTRALIZED GD

For $W = \begin{pmatrix} W_{\mathcal{A},\mathcal{A}} & W_{\mathcal{A},\mathcal{T}} \\ W_{\mathcal{T},\mathcal{A}} & W_{\mathcal{T},\mathcal{T}} \end{pmatrix}$, we have $\theta^{t+\frac{1}{2}} = \begin{pmatrix} \theta_{\mathcal{A}}^{t+\frac{1}{2}} \\ \left(\sum_{i=0}^{t} W_{\mathcal{T},\mathcal{T}}^{i}\right) g_{\mathcal{T}} + \sum_{i=0}^{t} W_{\mathcal{T},\mathcal{T}}^{i} N_T^{t-i} \\ + \sum_{i=0}^{t-1} W_{\mathcal{T},\mathcal{T}}^{t-1-i} W_{\mathcal{T},\mathcal{A}} \theta_{\mathcal{A}}^{i+\frac{1}{2}} \end{pmatrix}$

Algorithm Building the knowledge matrix for D-GD

Input: graph $\mathcal{G}$, attackers $\mathcal{A}$, targets $\mathcal{T} = \mathcal{V} \setminus \mathcal{A}$, iterations T
$i \leftarrow 0$
for t from 0 to $T-1$ **do**
  for each $v \in \mathcal{N}(\mathcal{A})$ **do**
    $K_T[i,:] \leftarrow (\sum_{j=0}^{t} W_{\mathcal{T},\mathcal{T}}^{j})[v - |\mathcal{A}|,:]$
    $i \leftarrow i + 1$
return K_T

Algorithm Removing the attackers' contributions

Input: gossip matrix W of $\mathcal{G}$, attackers $\mathcal{A}$, targets $\mathcal{T} = \mathcal{V} \setminus \mathcal{A}$, iterations T, dimension d, updates Y_T, concatenated updates $\theta_{\mathcal{A}}$
Initialize $\hat{Y}_T \in \mathbb{R}^{T \times |\mathcal{N}(\mathcal{A})| \times d}$
Initialize $B \in \mathbb{R}^{|\mathcal{T}| \times d}$ with zeros
for $t \in 0, 1, \ldots, T-1$ **do**
  $\hat{Y}_T[t,:] \leftarrow Y_T[t,:] - B[\mathcal{N}(\mathcal{A}),:]$
  $B \leftarrow W_{\mathcal{T},\mathcal{T}} B + W_{\mathcal{T},\mathcal{A}} \theta_{\mathcal{A}}^{t+\frac{1}{2}}$
return $\hat{Y}_T$

Theorem ([Cyffers et al., 2024])

After T iterations, for a level of noise $\sigma^2 \geq 2\alpha(\alpha - 1)$, the privacy loss from node u to v is bounded by:

$$\varepsilon_{u \to v} \leq \mathcal{O}\left(\frac{\alpha T \ln(T)}{\sigma^2 n^2} - \frac{\alpha T}{\sigma^2 n} \ln \left(I - W + \frac{1}{n} \mathbb{1}\mathbb{1}^\top \right)_{uv} \right) .$$

[Berthier et al., 2020] Berthier, R., Bach, F., and Gaillard, P. (2020).
Accelerated gossip in networks of given dimension using jacobi polynomial iterations.
SIAM Journal on Mathematics of Data Science, 2(1):24–47.

[Boyd et al., 2006] Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D. (2006).
Randomized gossip algorithms.
IEEE Transactions on Information Theory, 52(6):2508–2530.

[Cyffers and Bellet, 2022] Cyffers, E. and Bellet, A. (2022).
Privacy Amplification by Decentralization.
In *AISTATS*.

[Cyffers et al., 2024] Cyffers, E., Bellet, A., and Upadhyay, J. (2024).
Differentially Private Decentralized Learning with Random Walks.
In *ICML*.

[Cyffers et al., 2022] Cyffers, E., Even, M., Bellet, A., and Massoulié, L. (2022).
Muffliato: Peer-to-Peer Privacy Amplification for Decentralized Optimization and Averaging.
In *NeurIPS*.

[El Mrini et al., 2024] El Mrini, A., Cyffers, E., and Bellet, A. (2024).
Privacy Attacks in Decentralized Learning.
In *ICML*.

[Feldman et al., 2018] Feldman, V., Mironov, I., Talwar, K., and Thakurta, A. (2018).
Privacy Amplification by Iteration.
In *FOCS*.

[Geiping et al., 2020] Geiping, J., Bauermeister, H., Dröge, H., and Moeller, M. (2020).
Inverting gradients - how easy is it to break privacy in federated learning?
In *NeurIPS*.

[Hatamizadeh et al., 2022] Hatamizadeh, A., Yin, H., Roth, H., Li, W., Kautz, J., Xu, D., and Molchanov, P. (2022).
Gradvit: Gradient inversion of vision transformers.
In *CVPR*.

[Johansson et al., 2009] Johansson, B., Rabi, M., and Johansson, M. (2009).
A randomized incremental subgradient method for distributed optimization in networked systems.
SIAM Journal on Optimization, 20(3):1157–1170.

[Koloskova et al., 2020] Koloskova, A., Loizou, N., Boreiri, S., Jaggi, M., and Stich, S. U. (2020).
A Unified Theory of Decentralized SGD with Changing Topology and Local Updates.
In *ICML*.

[Lian et al., 2017] Lian, X., Zhang, C., Zhang, H., Hsieh, C.-J., Zhang, W., and Liu, J. (2017).
Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent.
In *NIPS*.

[Mironov, 2017] Mironov, I. (2017).
Rényi Differential Privacy.
In *CSF*.

[Nasr et al., 2023] Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., and Lee, K. (2023).
Scalable extraction of training data from (production) language models.
Technical report, arXiv:2311.17035.

[Pasquini et al., 2023] Pasquini, D., Raynal, M., and Troncoso, C. (2023).
On the (in)security of peer-to-peer decentralized machine learning.
In *IEEE Symposium on Security and Privacy (S&P)*.