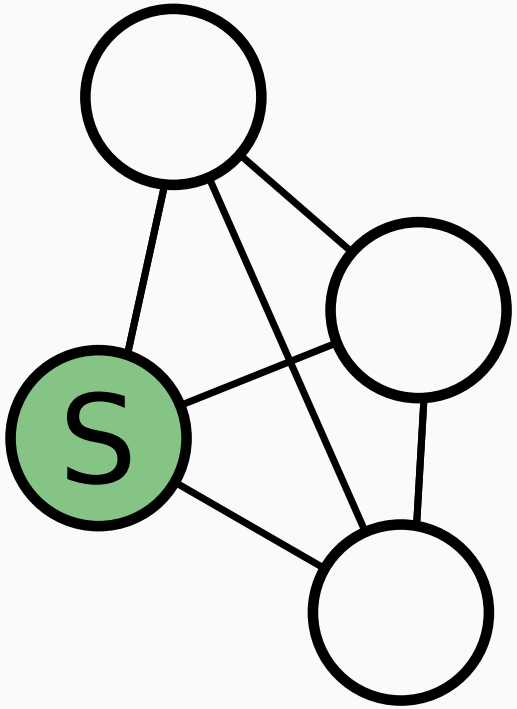
WHO STARTED THIS RUMOR? QUANTIFYING THE NATURAL DIFFERENTIAL PRIVACY GUARANTEES OF GOSSIP PROTOCOLS

Aurélien Bellet (INRIA, Magnet team)

Joint work with Hadrien Hendrikx (INRIA/MSR-INRIA) and Rachid Guerraoui (EPFL)

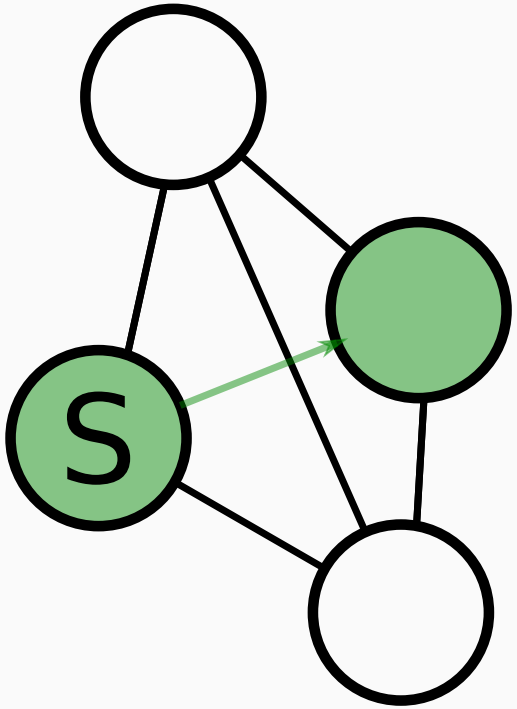
APVP 2019

GOSSIP RUMOR SPREADING



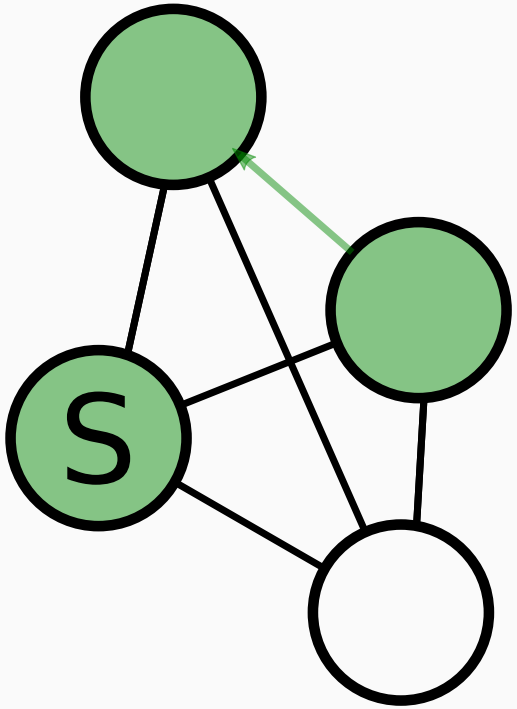
- A source node wants to propagate a rumor in a complete graph
- Nodes that know the rumor can tell another random peer
- Well studied problem [Pittel, 1987, Karp et al., 2000, Acan et al., 2017]

GOSSIP RUMOR SPREADING



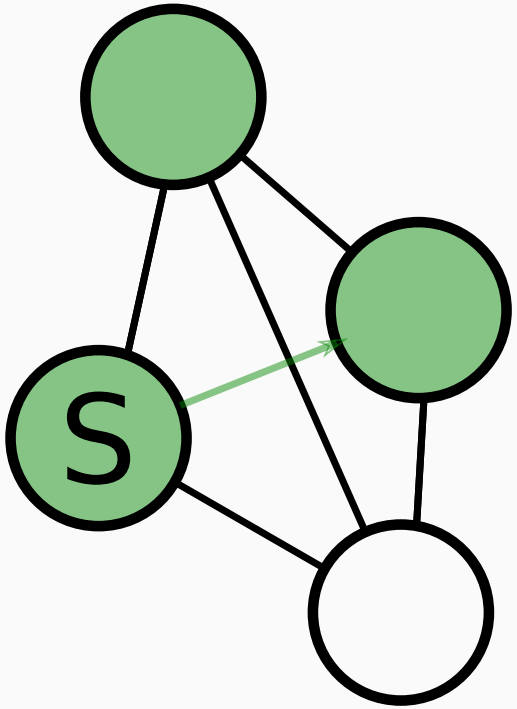
- A source node wants to propagate a rumor in a complete graph
- Nodes that know the rumor can tell another random peer
- Well studied problem [Pittel, 1987, Karp et al., 2000, Acan et al., 2017]

GOSSIP RUMOR SPREADING



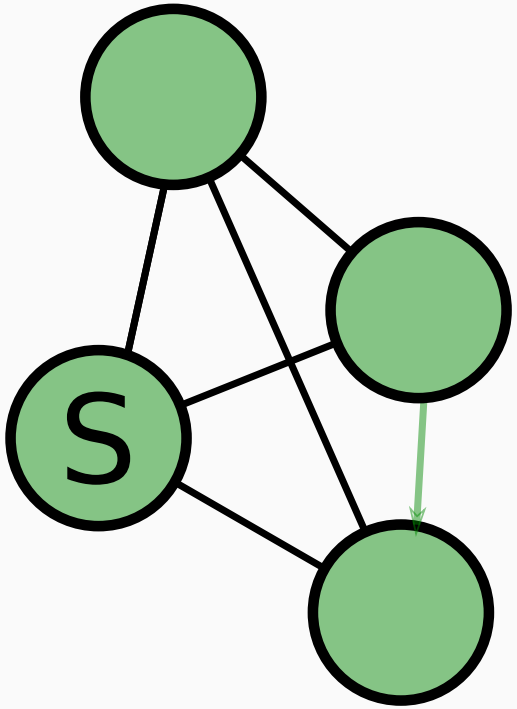
- A source node wants to propagate a rumor in a complete graph
- Nodes that know the rumor can tell another random peer
- Well studied problem [Pittel, 1987, Karp et al., 2000, Acan et al., 2017]

GOSSIP RUMOR SPREADING



- A source node wants to propagate a rumor in a complete graph
- Nodes that know the rumor can tell another random peer
- Well studied problem [Pittel, 1987, Karp et al., 2000, Acan et al., 2017]

GOSSIP RUMOR SPREADING



- A source node wants to propagate a rumor in a complete graph
- Nodes that know the rumor can tell another random peer
- Well studied problem [Pittel, 1987, Karp et al., 2000, Acan et al., 2017]

PRIVACY IN RUMOR SPREADING

- How to share information without being identified?
  - Copyright infringement when sharing music, whistle-blowers, freedom of speech
- Converse problem: how to accurately locate the source?
  - Copyright infringement when sharing music, computer or biological virus diffusion, fake news
- Folklore belief: gossip protocols provide some source anonymity
- Crucial to understand the fundamental limits of this claim
- Existing work [Pinto et al., 2012, Fanti et al., 2017] has focused on specific protocols and attacks, without prior information

CONTRIBUTIONS

1. Propose an information-theoretic model of anonymity in gossip protocols based on differential privacy
2. Prove matching upper and lower bounds on privacy
3. Analyze the trade-off between privacy and dissemination speed

PROPOSED MODEL

CLASS OF GOSSIP PROTOCOLS

- A set of n nodes labeled from 0 to $n-1$
- Let I the set of informed nodes
- The primitive `tell_gossip`(i, I) allows an informed node $i \in I$ to tell the rumor to another node j chosen uniformly at random

Definition (Gossip protocols)

A gossip protocol on a complete graph is one that (a) terminates, (b) ensures that at the end of its execution, the set of informed nodes $I = \{0, \ldots, n-1\}$, and (c) can modify I only through calls to the `tell_gossip` primitive.

ADVERSARY MODEL

- The attacker is a set $\mathcal{C}$ of curious nodes of size f
- A gossip algorithm generates an ordered random sequence S_{omni} of triplets (t, i, j) of executions of `tell_gossip`
- The attacker gathers a random subsequence $S \subset S_{omni}$:

$$S = ((i, j) \mid (t, i, j) \in S_{omni}, j \in \mathcal{C})$$

- Ratio f/n: probability for the attacker to observe a communication (thought of as a constant independent of n)

PRIVACY DEFINITIONS

- We adapt (ϵ, δ)-differential privacy (DP) [Dwork et al., 2006] to our case where the source node defines the input
- $p_i(E)$: probability of event E if node i is the source
- $\mathcal{S}$: set of all possible sequences observed by the adversary

Definition (Differential privacy)

We say that a gossip protocol is (ϵ, δ)-differentially private if

$$p_i(S) \leq e^{\epsilon} p_j(S) + \delta, \quad \forall S \subset \mathcal{S}, \ \forall i, j \in \{0, ..., n-1\}$$

- Source indistinguishability: any output is almost as likely regardless of who started the rumor
- Robustness to side information (e.g., source is among k nodes)
- Composition properties (e.g., sequence of related rumors)

- When $\delta > 0$, (ϵ, δ)-DP allows the protocol to release the source identity with small probability
- We will favor DP protocols that also guarantee a notion of prediction uncertainty

Definition (Prediction uncertainty)

A gossip algorithm guarantees prediction uncertainty with $c > 0$ if for a uniform prior $P(I_0)$ on the source node and any $i \in \{0, ..., n-1\}$:

$$\frac{p(I_0 \neq \{i\}|S)}{p(I_0 = \{i\}|S)} \geq c, \quad \forall S \subset \mathcal{S},\ p_i(S) > 0$$

- No output can reveal the source with large probability
- Bounds the probability of success of *any* attack by $1/(1+c)$
- $(\epsilon, 0)$-DP with $\epsilon > 0$ implies PU but converse is not true

OPTIMAL PRIVACY

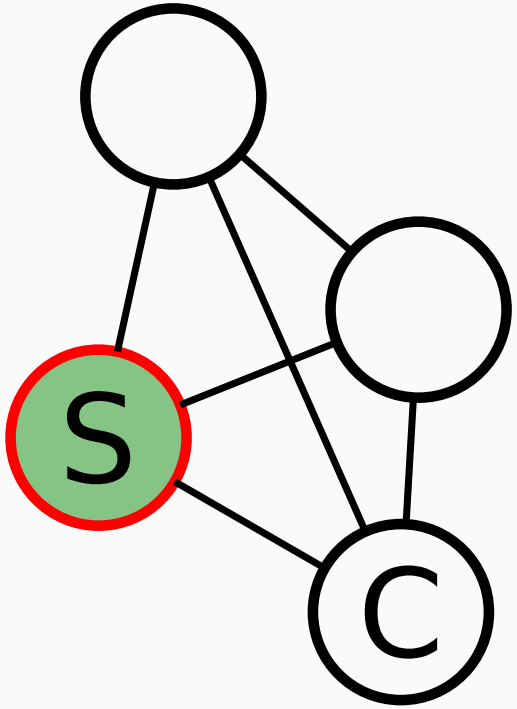
MATCHING LOWER AND UPPER BOUNDS

Theorem ([Bellet et al., 2019])

If a gossip algorithm satisfies (ϵ, δ)-differential privacy for any $\epsilon \geq 0$ and c-prediction uncertainty, than we have $\delta \geq \frac{f}{n}(1 - \frac{e^{\epsilon}-1}{f})$ and $c \leq \frac{n}{f+1} - 1$. Furthermore, these bounds are matched by a protocol.

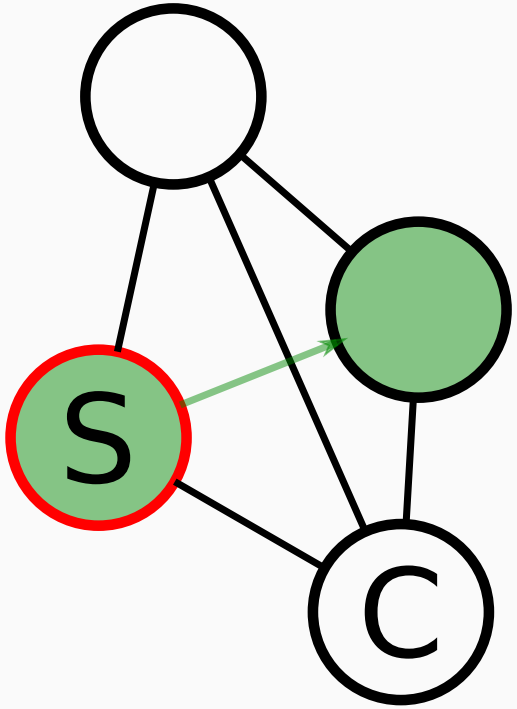
- Proof technique: consider the set of sequences where the first node to communicate with a curious node is the source
- In the regime $\epsilon = 0$: δ cannot be smaller than the proportion of curious nodes (intuitive)
- Remarkably, one can achieve $\delta < f/n$ by trading-off with ϵ (can even get pure ϵ-DP for $\epsilon \approx \log f$)
- PU guarantee: attackers always have a high probability of making a mistake

OPTIMALLY PRIVATE PROTOCOL



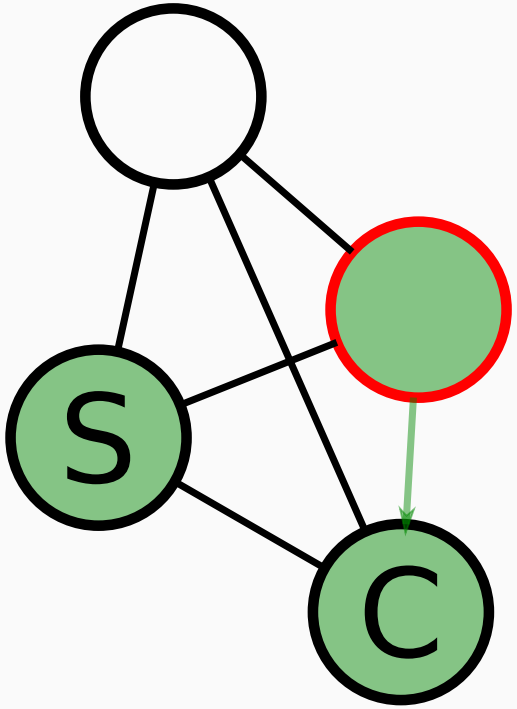
- Nodes communicate the rumor to exactly one node
- The rumor does not die while the state of the system after first communication is independent from the source
- All nodes (including the source) follow the same behavior

OPTIMALLY PRIVATE PROTOCOL



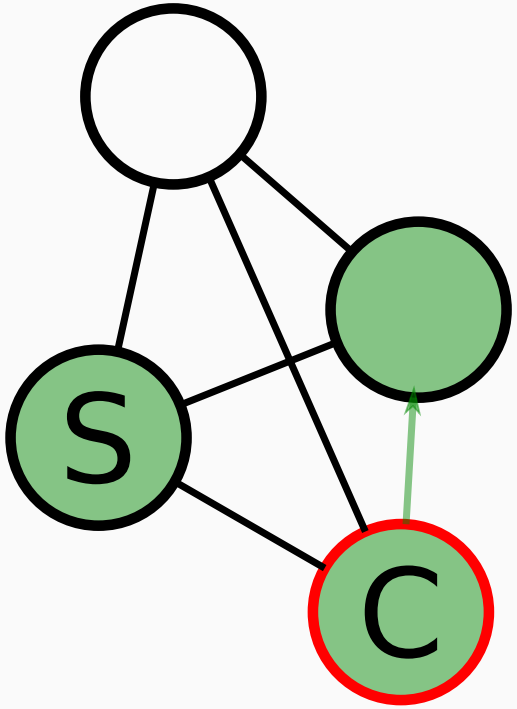
- Nodes communicate the rumor to exactly one node
- The rumor does not die while the state of the system after first communication is independent from the source
- All nodes (including the source) follow the same behavior

OPTIMALLY PRIVATE PROTOCOL



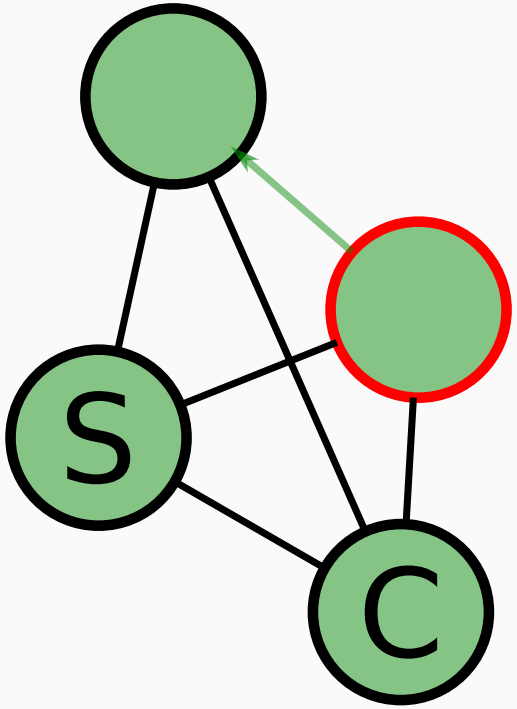
- Nodes communicate the rumor to exactly one node
- The rumor does not die while the state of the system after first communication is independent from the source
- All nodes (including the source) follow the same behavior

OPTIMALLY PRIVATE PROTOCOL



- Nodes communicate the rumor to exactly one node
- The rumor does not die while the state of the system after first communication is independent from the source
- All nodes (including the source) follow the same behavior

OPTIMALLY PRIVATE PROTOCOL



- Nodes communicate the rumor to exactly one node
- The rumor does not die while the state of the system after first communication is independent from the source
- All nodes (including the source) follow the same behavior

OPTIMALLY PRIVATE PROTOCOL: DIFFUSION SPEED

- This algorithm achieves optimal DP but it is slow
- It takes $O(n \log n)$ rounds to inform all nodes with probability at least $1 - 1/n$
- In contrast, the standard gossip ("push") algorithm needs only $O(\log n)$ rounds
- Can we design a gossip algorithm which is both fast and private?

FAST AND PRIVATE GOSSIP PROTOCOLS

A PARAMETERIZED GOSSIP PROTOCOL

Algorithm 1

Input: Number of nodes n, source node k, probability $s \in [0, 1]$
$I \leftarrow \{k\}, A \leftarrow \{k\}$
while $|I| < n$ **do**
    for each $i \in A$ **do**
        $j, I \leftarrow \texttt{tell\_gossip}(i, I)$
        $A \leftarrow A \cup \{j\}$
        $A \leftarrow A \setminus \{i\}$ with probability $1 - s$
    end for
end while

- Parameter s controls the trade-off between privacy and speed
- $s = 0$ recovers the optimal but slow protocol
- $s = 1$ recovers the standard "push" gossip algorithm

Theorem ([Bellet et al., 2019])

For $0 < s < 1$, Algorithm 1 with parameter s guarantees $(0, \delta)$-differential privacy with:

$$\delta \leq \min_{r \in \mathbb{N}} 1 - (1 - s^r)\left(1 - \frac{f}{n}\right)^r$$

- When s is not too high ($s \leq 0.5$), this is very close to the lower bound (within a factor 2)
- For $s = 1$ we have an impossibility result: if the algorithm satisfies DP for all n then $\delta = 1$

Theorem ([Bellet et al., 2019])

Algorithm 1 guarantees prediction uncertainty with

$$c = \left(1 - \frac{f+1}{n}\right)(1-s).$$

- Prediction uncertainty holds, unlike for $s = 1$
- Note: some gap with optimal guarantee for $s = 0$ (most likely due to our proof technique)

DIFFUSION SPEED

- For simplicity, consider the synchronous version of Algorithm 1 in which a round corresponds to iterating over the full set A

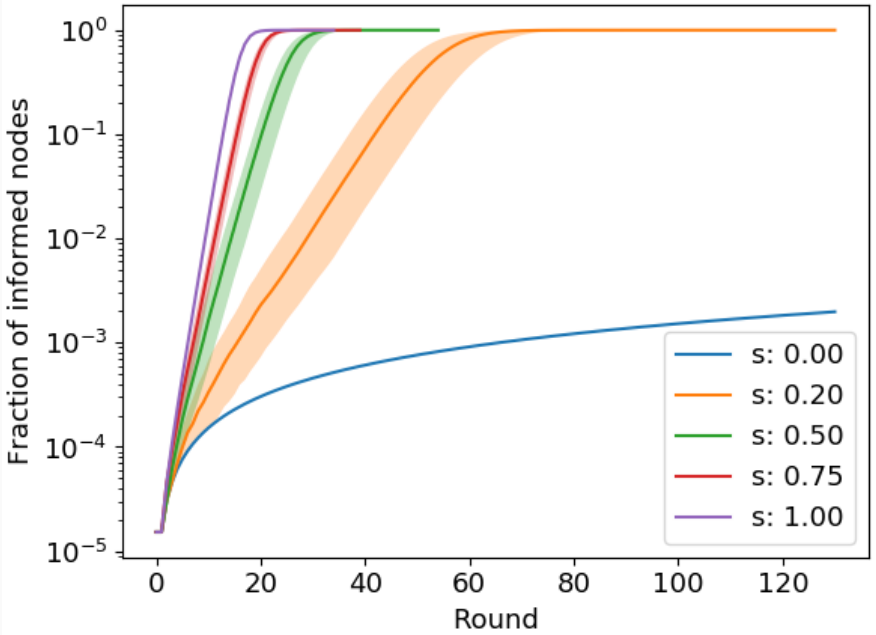
Theorem ([Bellet et al., 2019])

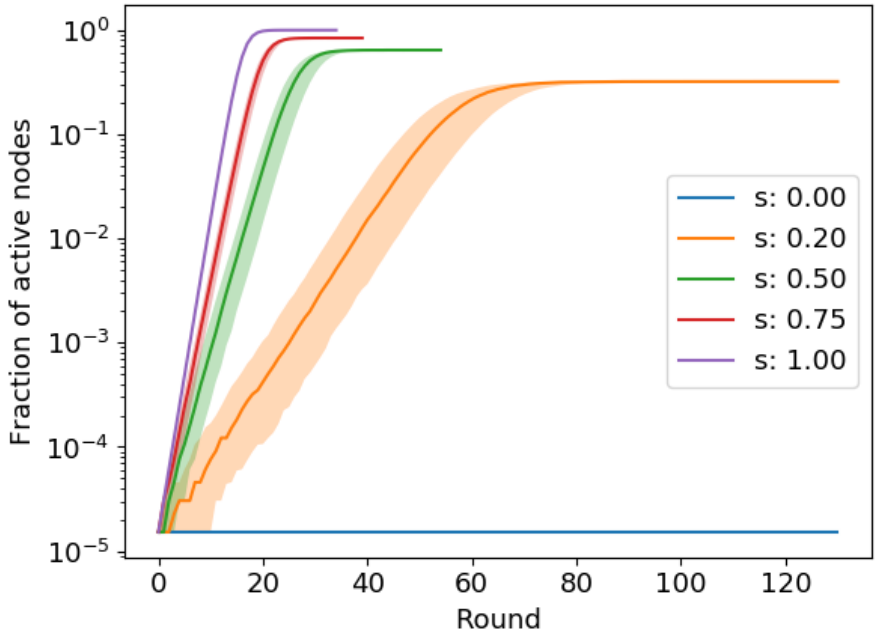
For a given $s > 0$, there exists $\alpha > 0$ such that for all $C > 0$, there exists n large enough such that the gossip protocol with parameter s needs at most $C\alpha^{-1} \log n$ rounds to ensure that all nodes are informed with probability at least $1 - \frac{1}{n}$.

- Logarithmic spreading time is preserved for $s > 0$
- Intuition: for $s > 0$ there is an exponential growth of the number of actives nodes (it takes approximately $1/s$ rounds to double)

DIFFUSION SPEED

- The previous result is asymptotic, but the empirical behavior is consistent even for networks of moderate size (below: $n = 2^{16}$)





FUTURE WORK

FUTURE WORK

- Extensions to more general graphs
- Requires to relax DP: cf Pufferfish [Kifer and Machanavajjhala, 2014], metric-based DP [Chatzikokolakis et al., 2013]
- "Free" amplification of DP for the content of the message: cf recent work on the shuffle model [Balle et al., 2019]
- Strong potential implications for privacy-preserving decentralized machine learning [Bellet et al., 2018]

Thank you for your attention!

Questions?

REFERENCES I

[Acan et al., 2017] Acan, H., Collevecchio, A., Mehrabian, A., and Wormald, N. (2017).
On the push&pull protocol for rumor spreading.
SIAM Journal on Discrete Mathematics, 31(2):647–668.

[Balle et al., 2019] Balle, B., Bell, J., Gascon, A., and Nissim, K. (2019).
The Privacy Blanket of the Shuffle Model.
Technical report, arXiv:1903.02837.

[Bellet et al., 2019] Bellet, A., Guerraoui, R., and Hendrikx, H. (2019).
Who started this rumor? Quantifying the natural differential privacy guarantees of gossip protocols.
Technical report, 1902.07138.

[Bellet et al., 2018] Bellet, A., Guerraoui, R., Taziki, M., and Tommasi, M. (2018).
Personalized and Private Peer-to-Peer Machine Learning.
In *AISTATS*.

[Chatzikokolakis et al., 2013] Chatzikokolakis, K., Andrés, M. E., Bordenabe, N. E., and Palamidessi, C. (2013).
Broadening the Scope of Differential Privacy Using Metrics.
In *Proceedings of the International Symposium on Privacy Enhancing Technologies (PETS)*.

References II

[Dwork et al., 2006] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006).
Our data, ourselves: Privacy via distributed noise generation.
In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503.

[Fanti et al., 2017] Fanti, G., Kairouz, P., Oh, S., Ramchandran, K., and Viswanath, P. (2017).
Hiding the rumor source.
IEEE Transactions on Information Theory.

[Karp et al., 2000] Karp, R., Schindelhauer, C., Shenker, S., and Vocking, B. (2000).
Randomized rumor spreading.
In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 565–574. IEEE.

[Kifer and Machanavajjhala, 2014] Kifer, D. and Machanavajjhala, A. (2014).
Pufferfish: A framework for mathematical privacy definitions.
ACM Transactions on Database Systems (TODS), 39(1):3.

[Pinto et al., 2012] Pinto, P. C., Thiran, P., and Vetterli, M. (2012).
Locating the source of diffusion in large-scale networks.
Physical review letters, 109(6):068702.

[Pittel, 1987] Pittel, B. (1987).
On spreading a rumor.
SIAM Journal on Applied Mathematics, 47(1):213–223.

REFERENCES III