University of Amsterdam
Informatics Institute

DNA: Differential Privacy Neural Augmentation for Contact Tracing

ICLR Private ML workshop

Rob Romijnders, Christos Louizos, Yuki M. Asano, Max Welling

ELSEVIER
Brain, Behavior, and Immunity
Volume 89, October 2020, Pages 531-542

Review Article

COVID-19 pandemic and mental health consequences: Systematic review of the current evidence

Nina Vindegaard, Michael Eriksen Benros

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.bbi.2020.05.048 Get rights and content

ELSEVIER
Best Practice & Research Clinical Anaesthesiology
Volume 35, Issue 3, October 2021, Pages 293-306

2

Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives

Alan D. Kaye MD, PhD (Provost & Vice Chancellor of Academic Affairs) [a],
Chikezie N. Okeagu MD (Assistant Professor) [b], Alex D. Pham MD (Resident Physician) [c],
Rayce A. Silva (Medical Student) [d], Joshua J. Hurley MD, PGY-1 (Resident Physician) [e],
Brett L. Arron MD (Associate Professor) [f], Noeen Sarfraz MD MPH (Resident Physician) [g],
Hong N. Lee MD (Assistant Professor) [h], G.E. Ghali DDS, MD, FACS, FRCS(Ed) (Chancellor) [i],
Jack W. Gamble (Professor and Chairman) [i], Henry Liu MD (Professor) [j],
Richard D. Urman MD (Associate Professor) [k],
Elyse M. Cornett PhD (Assistant Professor) [h]

Show more

Covid-19: Cities fear 'huge' economic impact of restrictions

29 September 2020

BBC, Sept 2020

Covid had negative impact on children's reading - Estyn

4 May

BBC, May 2023

This interactive tool tracks covid-19 travel restrictions by country

Skyscanner's detailed travel map is color-coded in stoplight-style green, yellow and red

Washington Post, December 2020

Privacy is important

"The top reasons against app use were as follows: mistrusting the government, concerns about data security and **privacy**, and doubts about efficacy." Jones et al. 2021

Privacy is important

"The top reasons against app use were as follows: mistrusting the government, concerns about data security and **privacy**, and doubts about efficacy." Jones et al. 2021

"The most cited reasons for not downloading were related to **data** (...) **concerns**" Gao et al. 2022

Privacy is important

"The top reasons against app use were as follows: mistrusting the government, concerns about data security and **privacy**, and doubts about efficacy." Jones et al. 2021

"The most cited reasons for not downloading were related to **data** (...) **concerns**" Gao et al. 2022

"The main reasons for not downloading and using the app were (...) **worries about privacy**" Walrave et al. 2022

The Lancet, 2020

"most of the applications in use or under consideration have an impact on individual privacy that democratic societies would normally consider to be ***unacceptably high****"*

Our problem

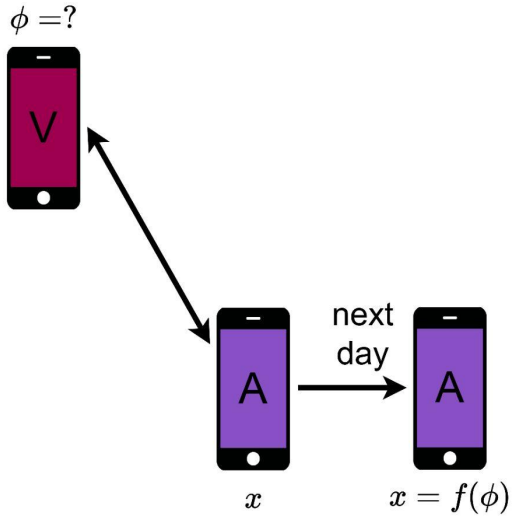
Low peak infection rate

under reasonable differential privacy

Attack Scenario

Privacy with respect to released covidscore

V is victim, A is attacker

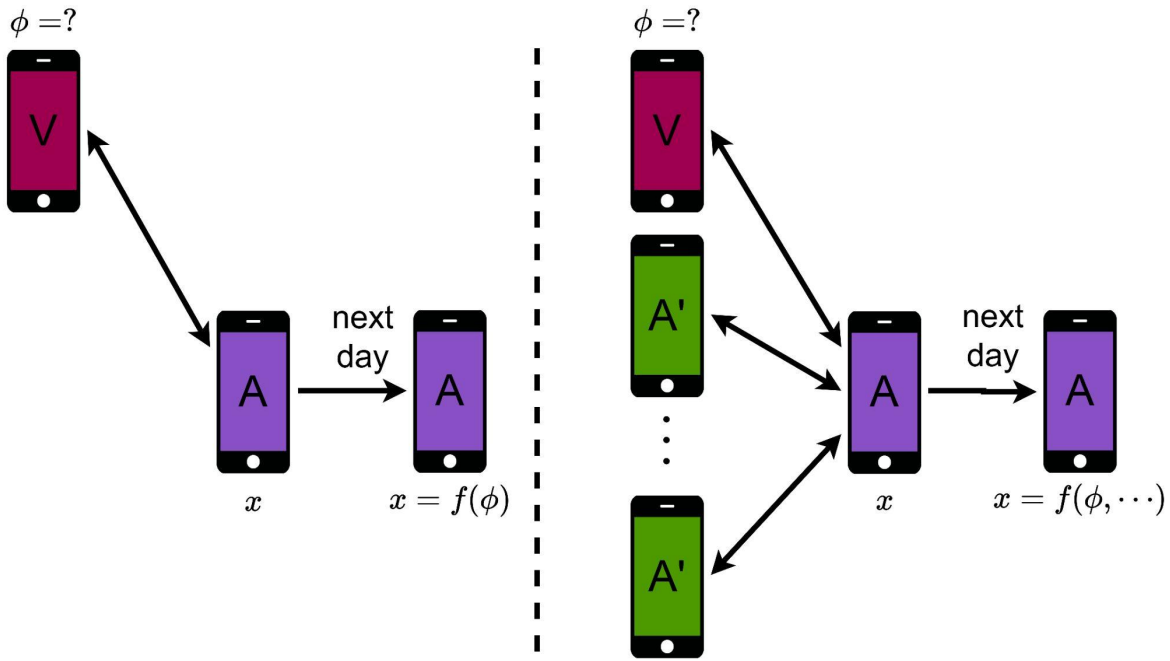


Attack Scenario

Privacy with respect to released covidscore

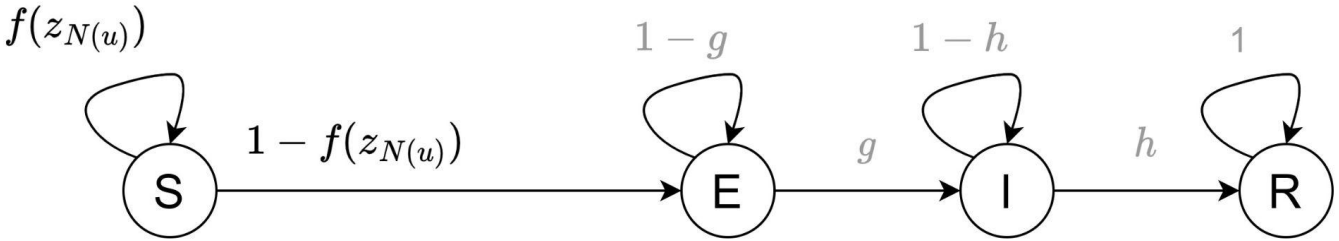
V is victim, A is attacker

Green phones, A' are co-attackers



SEIR transitions are a Markov chain

Susceptible - Exposed - Infected - Recovered



$$f(z_{N(u)}) = (1 - p_0)(1 - p_1)^{|\{z \in z_{N(u)}:\ z = I\}|}$$

Dynamics model

Susceptible

Exposed

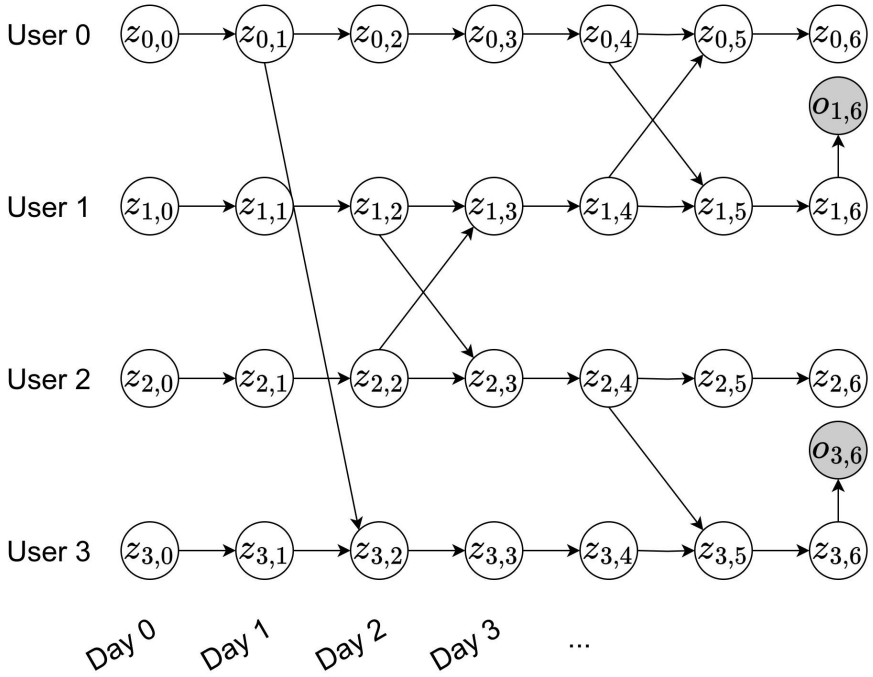
Infected

Recovered

$$P(z_{u,t+1}|\mathcal{Z}_t) = \begin{cases} f(u,t,\mathcal{Z}_t) & \text{if } z_t = S \wedge z_{t+1} = S \\ 1 - f(u,t,\mathcal{Z}_t) & \text{if } z_t = S \wedge z_{t+1} = E \\ 1 - g & \text{if } z_t = E \wedge z_{t+1} = E \\ g & \text{if } z_t = E \wedge z_{t+1} = I \\ 1 - h & \text{if } z_t = I \wedge z_{t+1} = I \\ h & \text{if } z_t = I \wedge z_{t+1} = R \\ 1 & \text{if } z_t = R \wedge z_{t+1} = R \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$f(u,t,\mathcal{Z}_t) = (1-p_0)(1-p_1)^{|\{(v,u,t)\in\mathcal{D}:z_{v,t}=I\}|} \quad (2)$$

Example of probabilistic model of four users in six days



Exact inference is intractable!

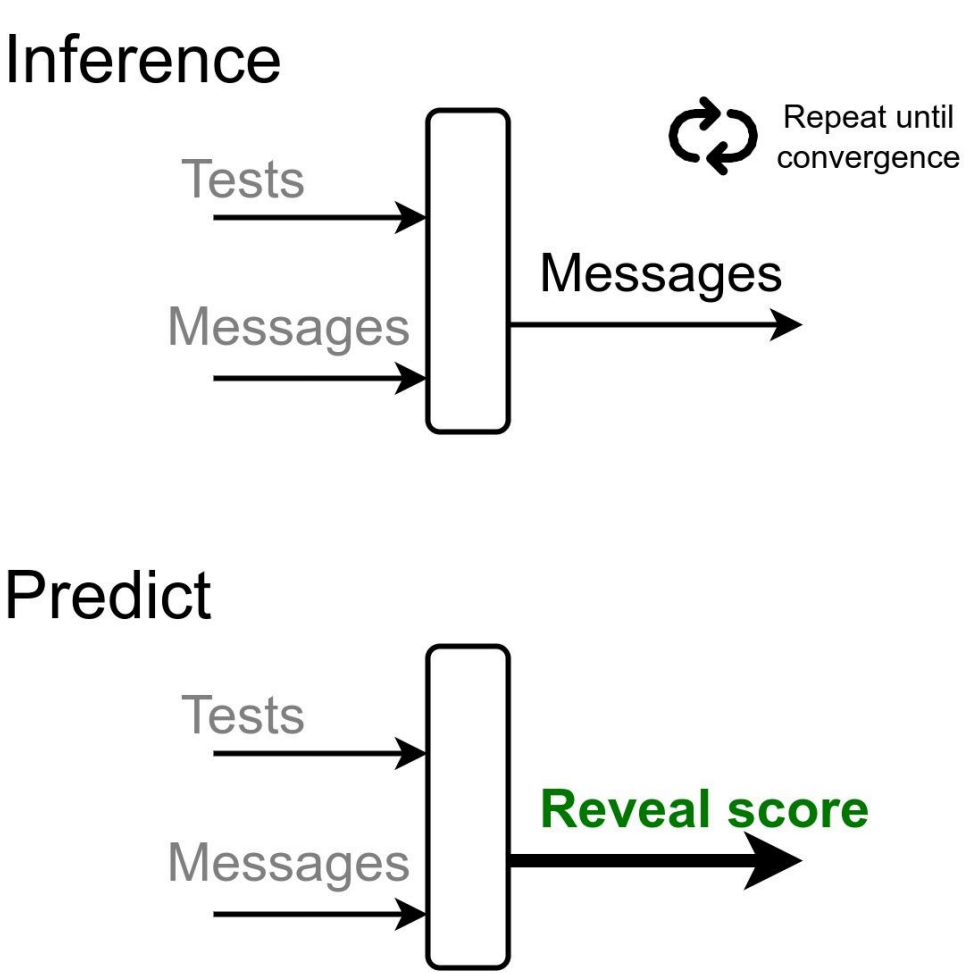
$$
\begin{aligned}
p(\text{joint distribution}) = \\
& p(z_{0,0}) \cdot p(z_{0,1}|z_{0,0}) \quad \cdot p(z_{0,2}|z_{0,1}) \quad \cdot p(z_{0,3}|z_{0,2}) \cdot p(z_{0,4}|z_{0,3}) \cdot p(z_{0,5}|z_{0,4}, z_{1,4}) \cdot p(z_{0,6}|z_{0,5}) \cdot \\
& p(z_{1,0}) \cdot p(z_{1,1}|z_{1,0}) \quad \cdot p(z_{1,2}|z_{1,1}) \cdot p(z_{1,3}|z_{1,2}, z_{2,2}) \cdot p(z_{1,4}|z_{1,3}) \cdot p(z_{1,5}|z_{1,4}, z_{0,4}) \cdot p(z_{1,6}|z_{1,5}) \cdot \\
& p(z_{2,0}) \cdot p(z_{2,1}|z_{2,0}) \quad \cdot p(z_{2,2}|z_{2,1}) \cdot p(z_{2,3}|z_{2,2}, z_{1,2}) \cdot p(z_{2,4}|z_{2,3}) \quad \cdot p(z_{2,5}|z_{2,4}) \cdot p(z_{2,6}|z_{2,5}) \cdot \\
& p(z_{3,0}) \cdot p(z_{3,1}|z_{3,0}) \cdot p(z_{3,2}|z_{3,1}, z_{0,1}) \quad \cdot p(z_{3,3}|z_{3,2}) \cdot p(z_{3,4}|z_{3,3}) \cdot p(z_{3,5}|z_{3,4}, z_{2,4}) \cdot p(z_{3,6}|z_{3,5}) \cdot \\
& p(o_{3,6}|z_{3,6}) \quad \cdot p(o_{1,6}|z_{1,6})
\end{aligned}
\tag{146}
$$

$$
p(z_{2,6}|o_{1,6}, o_{3,6}) = \frac{p(\text{joint distribution})}{\sum_{z_{0,0}} \sum_{z_{0,1}} \sum_{z_{0,2}} \sum_{z_{0,3}} \sum_{z_{0,4}} \sum_{z_{0,5}} \sum_{z_{0,6}} \sum_{z_{1,0}} \sum_{z_{1,1}} \sum_{z_{1,2}} \sum_{z_{1,3}} \sum_{z_{1,4}} \sum_{z_{1,5}} \sum_{z_{1,6}} \sum_{z_{2,0}} \sum_{z_{2,1}} \sum_{z_{2,2}} \sum_{z_{2,3}} \sum_{z_{2,4}} \sum_{z_{2,5}} \sum_{z_{2,6}} \sum_{z_{3,0}} \sum_{z_{3,1}} \sum_{z_{3,2}} \sum_{z_{3,3}} \sum_{z_{3,4}} \sum_{z_{3,5}} \sum_{z_{3,6}} p(\text{joint distribution})}
\tag{147}
$$

Modular view

Rules for approximate inference are given by either Belief Propagation or Factorized Neighbors

$$b_u(z_u) = \sum_{z_{N(u)}} P(z_u|z_{N(u)}, \mathcal{O}) B_{N(u)}(z_{N(u)})$$
$$= E_{B_{N(u)}(z_{N(u)})}[P(z_u|z_{N(u)}, \mathcal{O})].$$



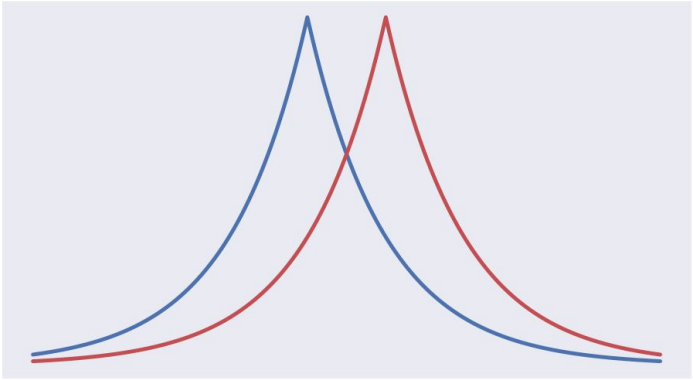
Differential privacy

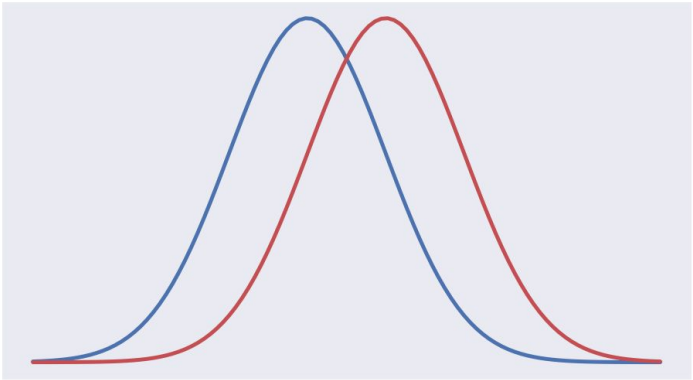
Definition of (ε, δ) differential privacy (Dwork and Roth 2014): for every $\varepsilon > 0$, $\delta \in [0, 1)$, a mechanism $f(\cdot)$, for any outcome Φ in the range of $f(\cdot)$, and any two adjacent data sets D, D' that differ in at most one element, satisfies the constraint:

$$p(f(D) \in \Phi) \leq e^{\varepsilon} p(f(D') \in \Phi) + \delta$$

Gaussians:

$$\sigma > \frac{\Delta}{\varepsilon}\left(2\log(\frac{1.25}{\delta})\right)^{\frac{1}{2}}$$



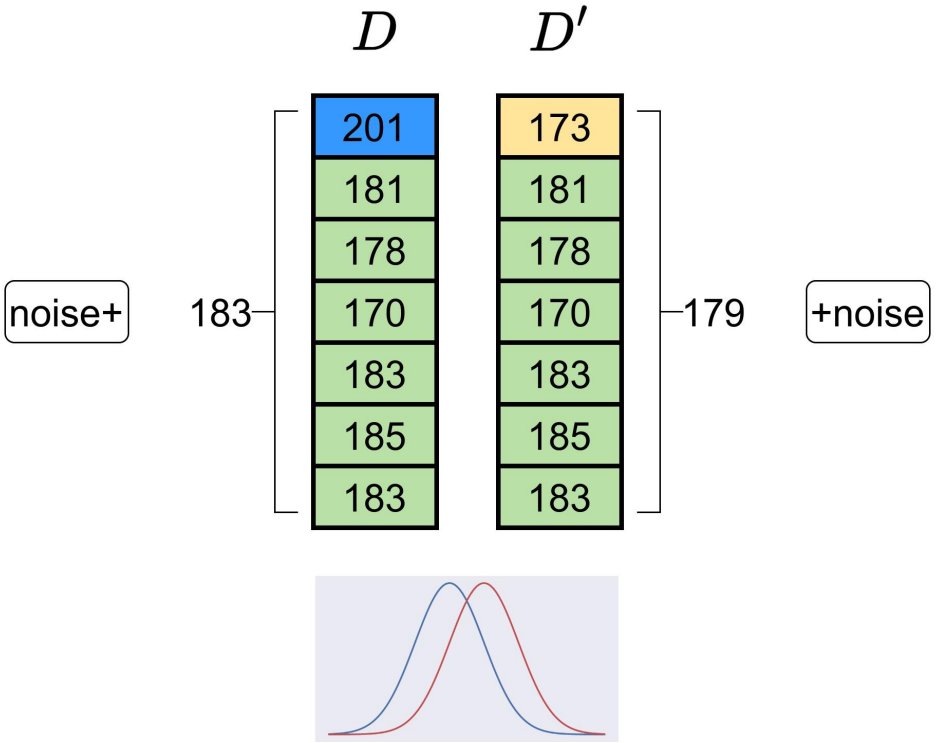


Differential privacy

Gaussian Mechanism:

Add noise according to ‘sensitivity’

$$\sigma > \frac{\Delta}{\varepsilon}\left(2\log(\frac{1.25}{\delta})\right)^{\frac{1}{2}}$$



Privacy bound

Dataset: a collection of messages, mu, on timesteps t

$$D = \{(\mu_i, t_i)\}_{i=1}^{C}$$

Sensitivity: the maximal change with respect to the change in one message

$$\Delta = \max_{\mu_1, \mu_1' \in [0, \gamma_u]} |F(\ (\mu_1, t_1) \cup D\ ) - F(\ (\mu_1', t_1) \cup D\ )| \leq p_1 \gamma_u \qquad \forall\ D.$$

Typically, p1 takes a value around 0.05, and gamma around 0.7

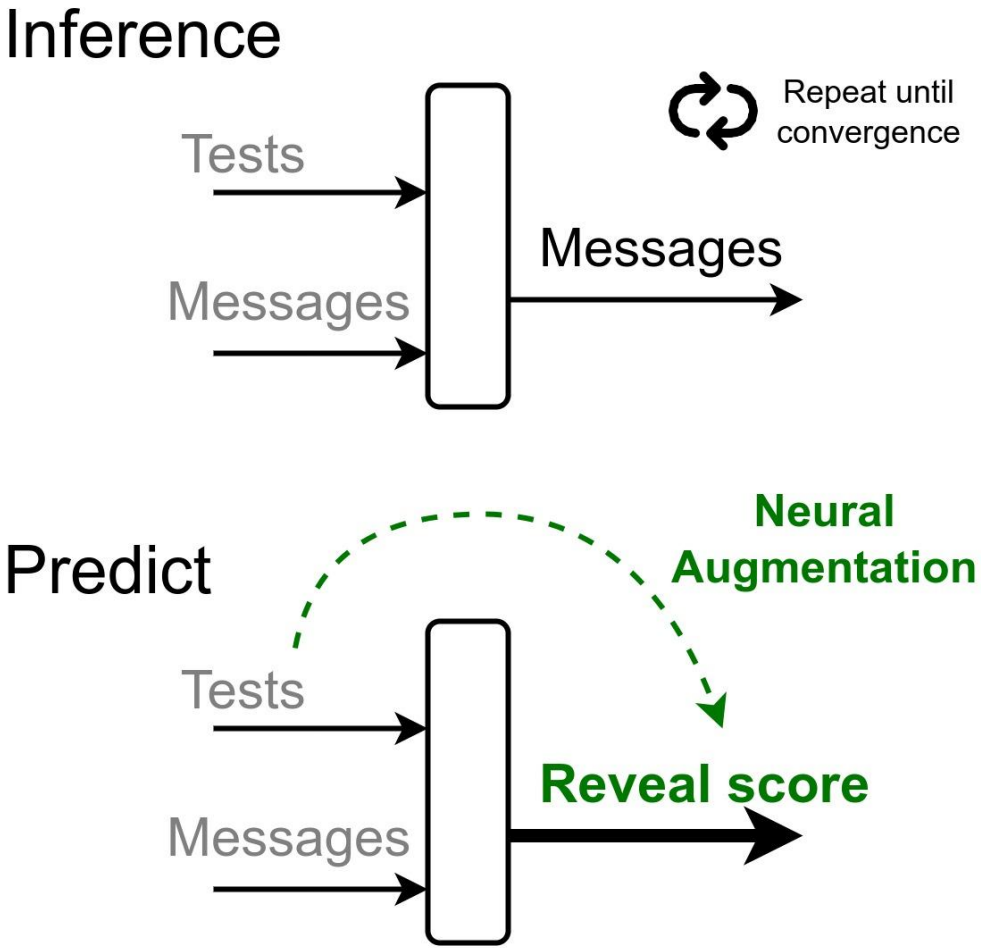
Renyi Differential Privacy between log-normals

$$\sigma^2 \geq \frac{a}{2C\rho}\Big(\log(1-\gamma_u p_1) - \log(1-\gamma_l p_1)\Big)^2$$

Neural Augmentation

Neural augmentation known from:

- MRI reconstruction
  (Lønning et al. Medical image analysis, 2019)
- Enhanced belief propagation
  (Satorras et al., AISTATS 2021)
- Fast sparse coding
  (Gregor et al. ICML 2010)



Lipschitz-bounded Neural Network

$$\phi = G_\theta(\{(\mu_i, t_i)\}_{i=1}^{C_T}) = g_\theta^{(2)}\left(\frac{1}{C} \sum_i g_\theta^{(1)}([\mu_i, t_i]^T) \right)$$

During training: estimate Lipschitz constant with power iterations O(p^2)
During testing: calculate Spectral norm exactly once O(p^3)

"Deep sets" Zaheer et al. NeurIPS 2017
"Spectral Normalization.." Miyato et al. ICLR 2018

Make Lipschitz function DP with Gaussian noise

Algorithm 1 DNA: Differentially private Neural Augmentation

Require: Dataset $D = \{(\mu_i, t_i)\}_{i=1}^{C_T}$, constants $p_1, \gamma_u \in (0, 1)$;

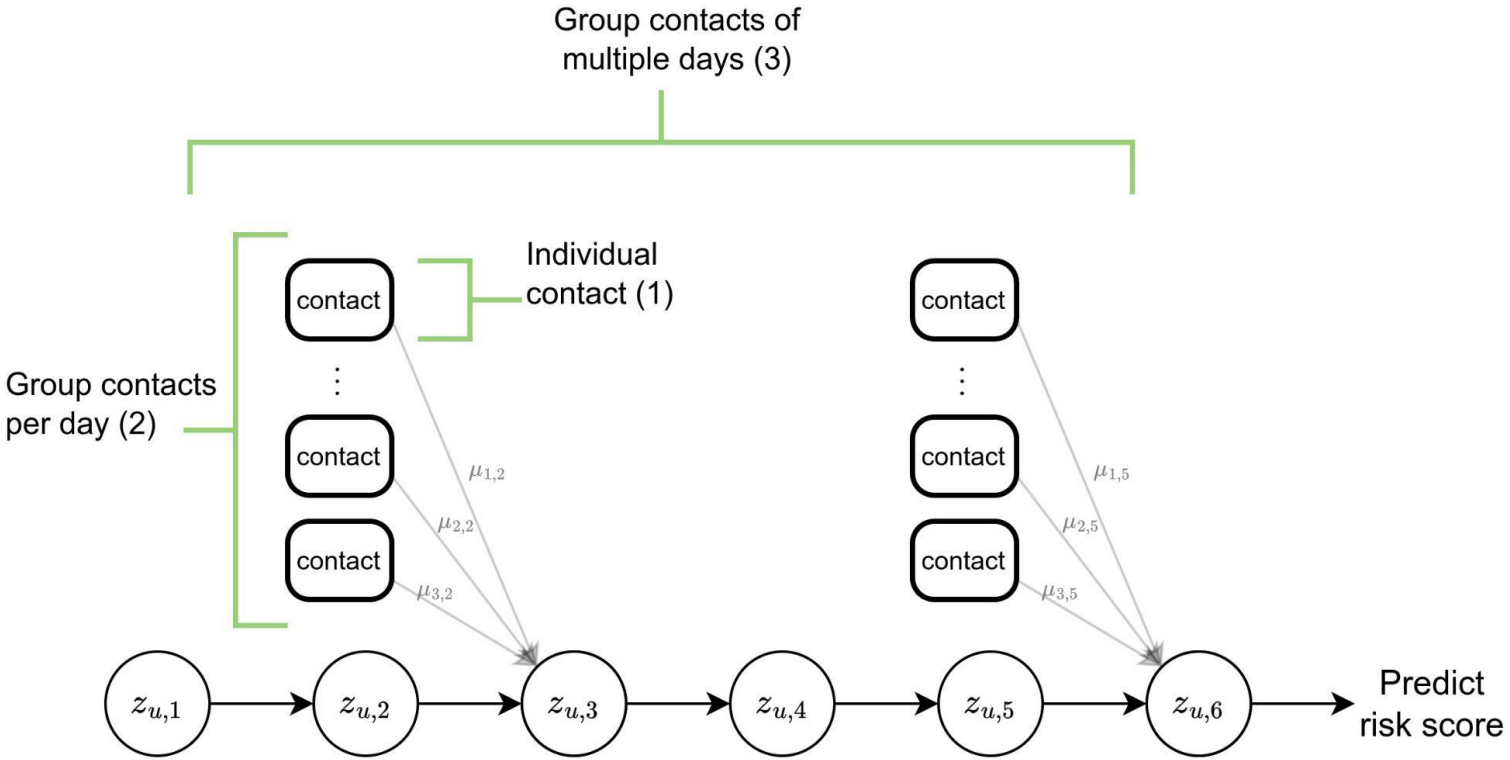
$$\mu_i \leftarrow \min(\mu_i, \gamma_u)$$

$$\bar{\phi} \leftarrow F(\{(\mu_i, t_i)\}_{i=1}^{C_T}) + p_1 \times G_\theta(\{(\mu_i, t_i)\}_{i=1}^{C_T})$$

$$\phi \leftarrow \bar{\phi} + \mathcal{N}(0, \tfrac{2}{\varepsilon^2}(\gamma_u p_1(1 + \tfrac{1}{C_T}))^2 \log(\tfrac{5}{4\delta}))$$

Jha et al. SIAM Journal on computing 2013

Privacy hierarchy



Different algorithms to compare on simulator

- Traditional contact tracing (Baker et al. 2021)
- Per-message, level 1 (Romijnders et al. 2023)
- DPFN, level 2 (Romijnders et al. 2024)
- DPFN-S, level 3 (Ours)
- DNA, level 3+ (Ours)

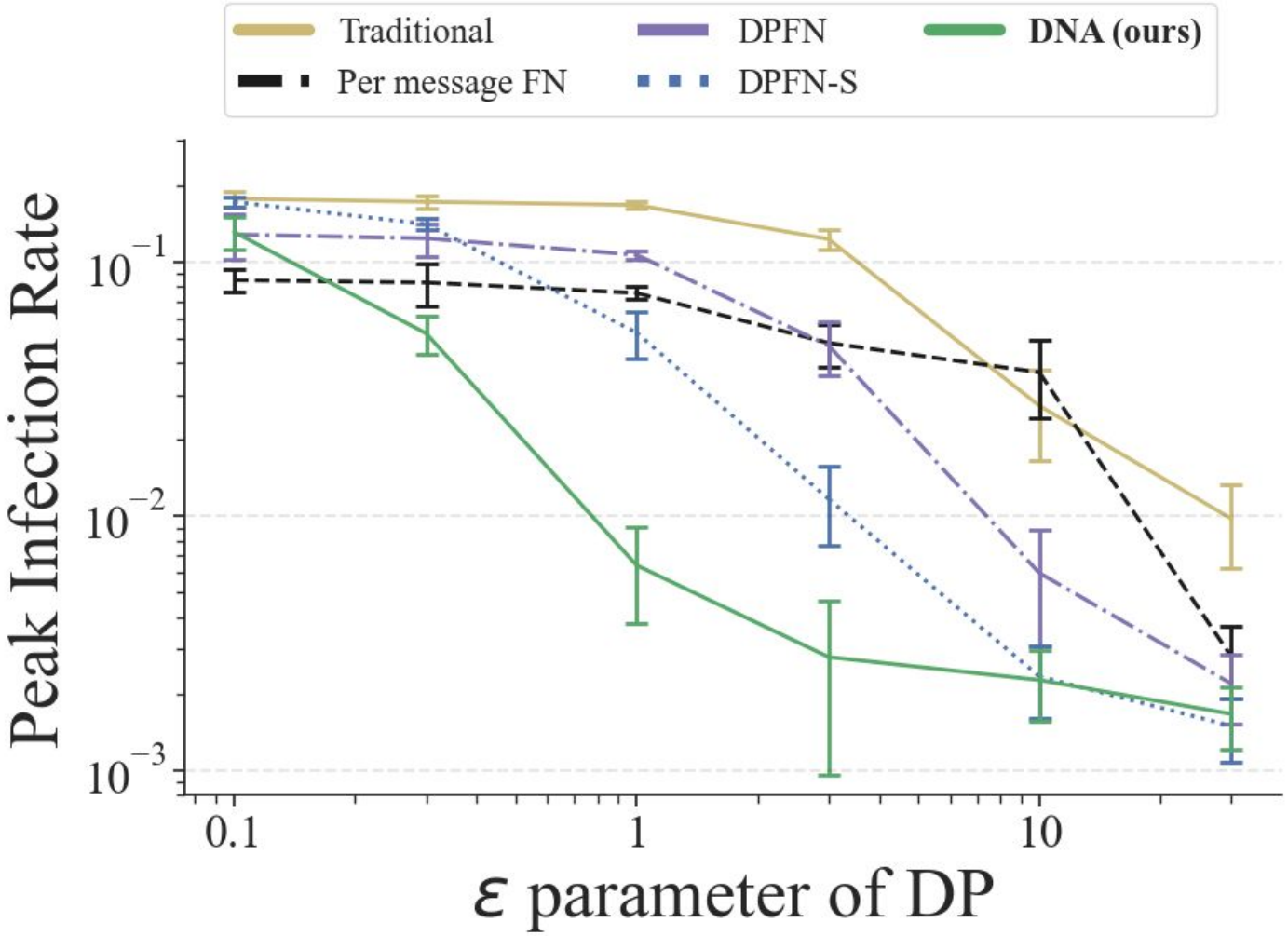
Simulator for experiments

Need simulator as better predictions interact with agents

OpenABM (Hinch et al. 2021)

- Stratifying for
  - 9 age categories
  - 3 occupations
  - 6 household types
- In total 150 parameters calibrated against a typical city in the UK

Experimental results



DNA has better utility under various noise scenarios

Even when up to 50% of the agents don't follow the protocol, or when the tests become more noisy, the DNA method still achieves lower Peak infection rate, compared to the same method without neural augmentation

	DPFN-S (‰)	**DNA** (‰)
Follow protocol		
100%	$52.7_{\pm 10.9}$	$6.4_{\pm 2.6}$
80%	$60.4_{\pm 9.6}$	$6.4_{\pm 2.2}$
50%	$100.1_{\pm 4.4}$	$27.2_{\pm 8.6}$
Noisy tests		
FPR 1%, FNR .1%	$52.7_{\pm 10.9}$	$6.4_{\pm 2.6}$
FPR 10%, FNR 1%	$81.3_{\pm 2.6}$	$19.5_{\pm 2.5}$
FPR 25%, FNR 3%	$130.4_{\pm 1.5}$	$81.3_{\pm 1.8}$

Units are number of infections per thousand agents, ± standard deviation

FPR/FNR = False Positive/Negative Rate

Conclusion

- Novel view of Lipschitz Neural Augmentation as providing Differential Privacy w.r.t. input

- This neural augmentation increases sensitivity, but compares favourably with better predictions

- Future work:
  - Decentralized reinforcement learning, partial adoption

University of Amsterdam
Informatics Institute

DNA: Differential Privacy Neural Augmentation for Contact Tracing

Questions

r.romijnders@uva.nl; romijndersrob@gmail.com
github.com/robromijnders/dna