University of Amsterdam
Informatics Institute

Contact Tracing with Differential Privacy Guarantees

Protect your score

Rob Romijnders, Christos Louizos, Yuki M. Asano, Max Welling
AAAI 2024

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

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****"*

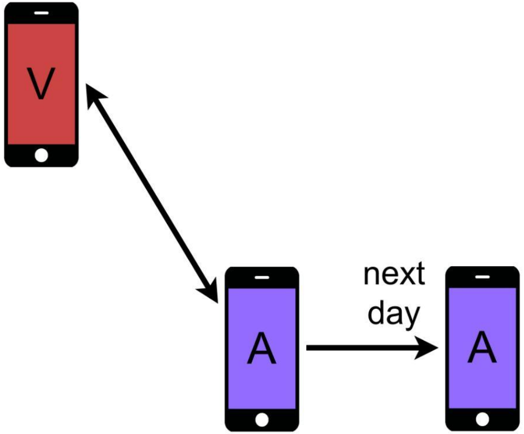
Research question

Low peak infection rate

under reasonable differential privacy

Attack Scenario

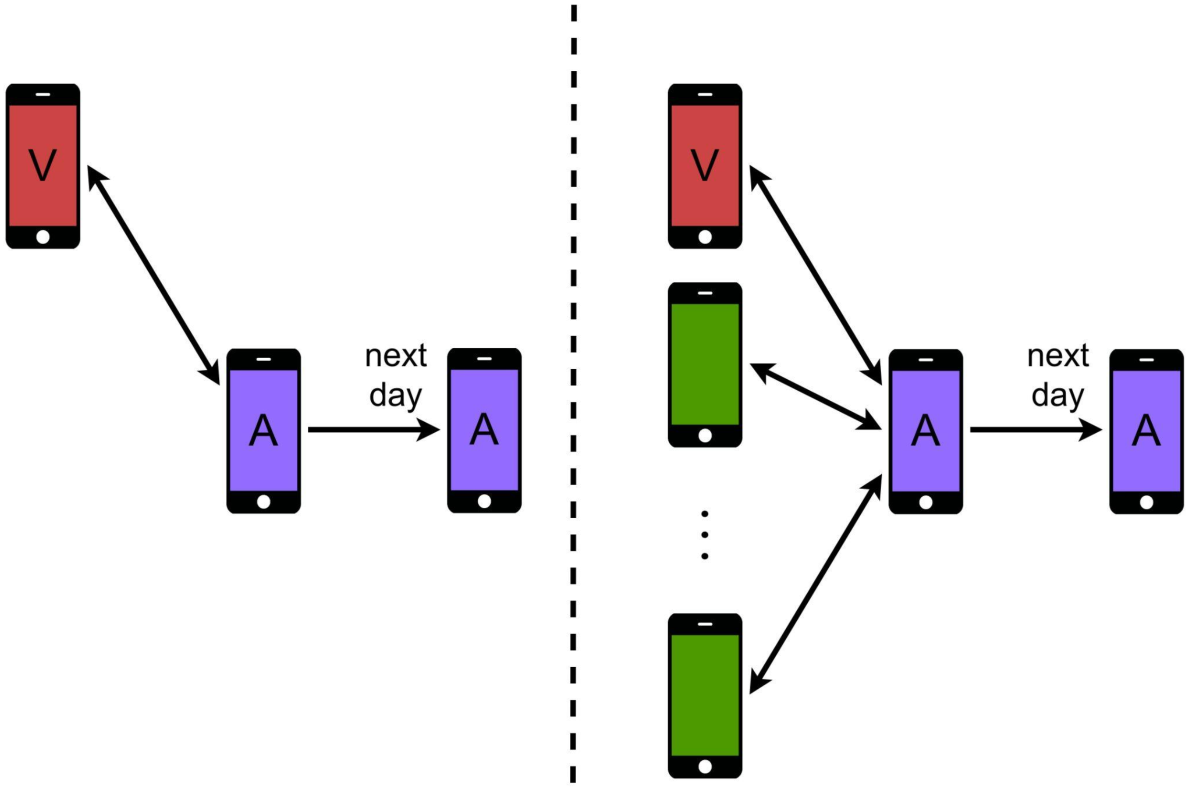
Privacy with respect to released covidscore



V is victim, A is attacker

Attack Scenario

Privacy with respect to released covidscore

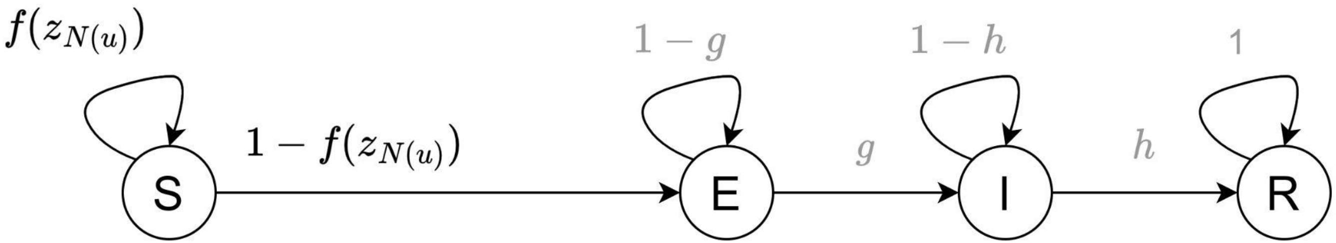


V is victim, A is attacker

Green phones are agents with ‘known’ score

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

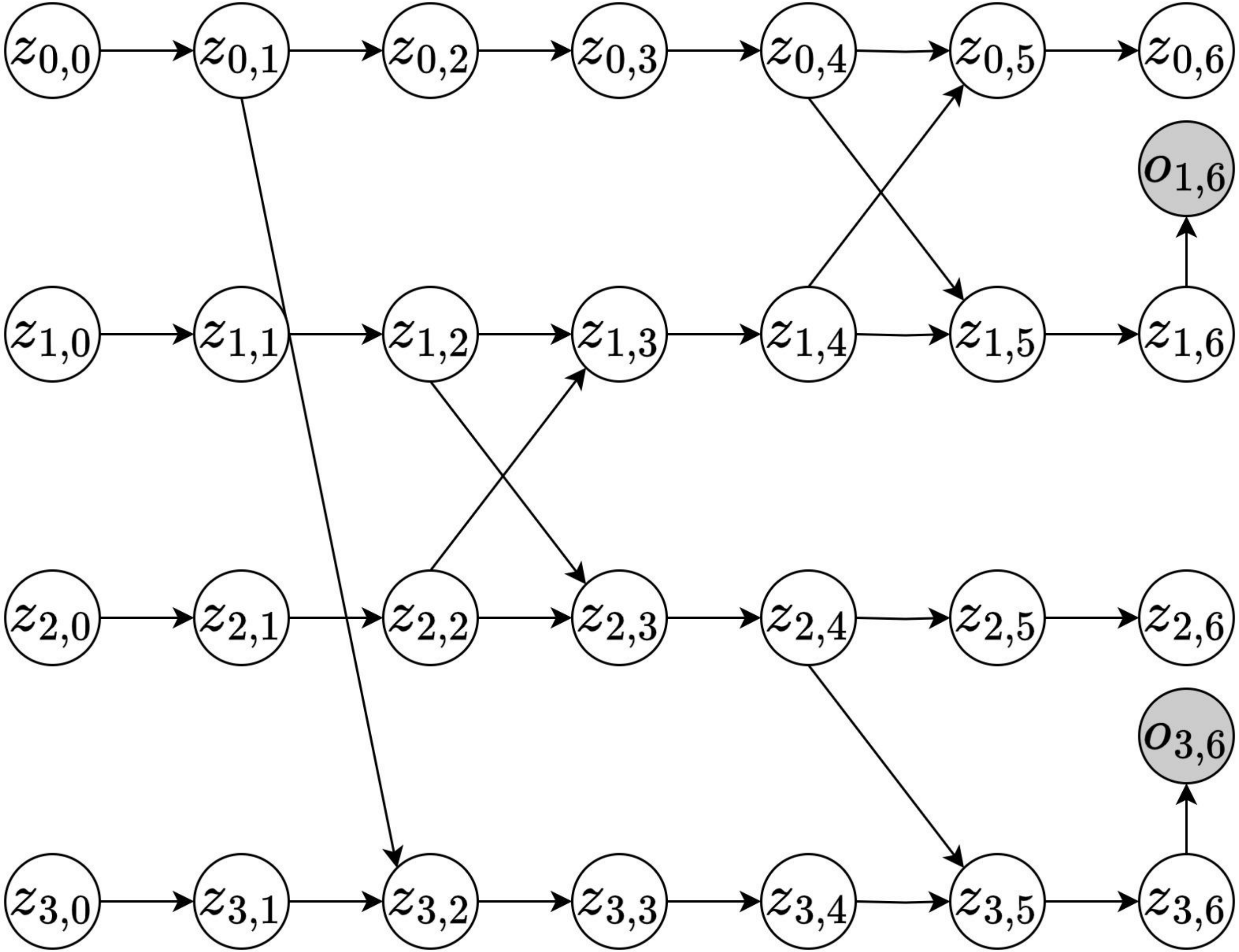
Exposed

Infected

Recovered

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

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


$z_{0,0}$
$z_{0,1}$
$z_{0,2}$
$z_{0,3}$
$z_{0,4}$
$z_{0,5}$
$z_{0,6}$
$o_{1,6}$
$z_{1,0}$
$z_{1,1}$
$z_{1,2}$
$z_{1,3}$
$z_{1,4}$
$z_{1,5}$
$z_{1,6}$
$z_{2,0}$
$z_{2,1}$
$z_{2,2}$
$z_{2,3}$
$z_{2,4}$
$z_{2,5}$
$z_{2,6}$
$o_{3,6}$
$z_{3,0}$
$z_{3,1}$
$z_{3,2}$
$z_{3,3}$
$z_{3,4}$
$z_{3,5}$
$z_{3,6}$

Exact inference is intractable!

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

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

Inference

Gibbs sampling

$$p(z_u | \hat{z}_{\neg u}, \mathcal{O}).$$

Belief Propagation

$$\mu_{f_s \to z_{u,t}}(z_{u,t}) = \sum_{z_s} f_s(z_s, z_{u,t}) \prod_{k \in \mathrm{Nb}(f_s) \setminus z_{u,t}} \mu_{z_k \to f_s}$$

$$\mu_{z_{u,t} \to f_s}(z_{u,t}) = \prod_{k \in \mathrm{Nb}(z_{u,t}) \setminus f_s} \mu_{f_k \to z_{u,t}}$$

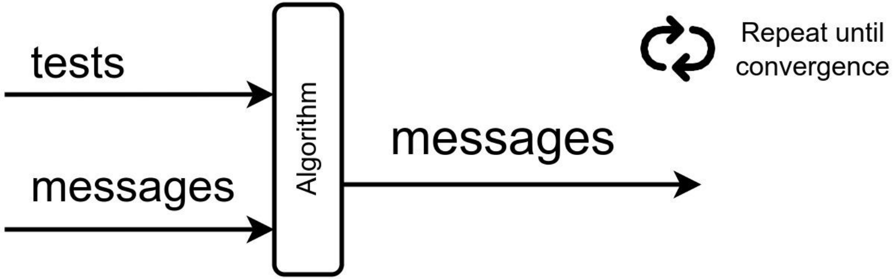
Factorised neighbours

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

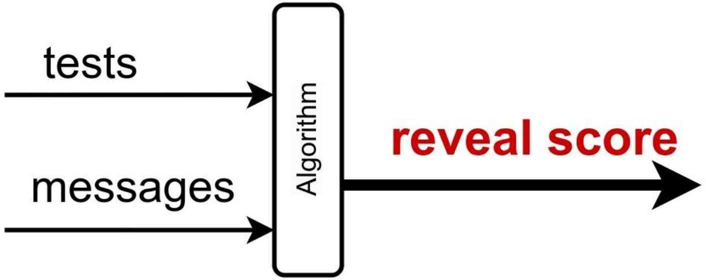
Rosen-zvi et al. “The dlr hierarchy of approximate inference.” 2012

Modular view of contact tracing

Inference



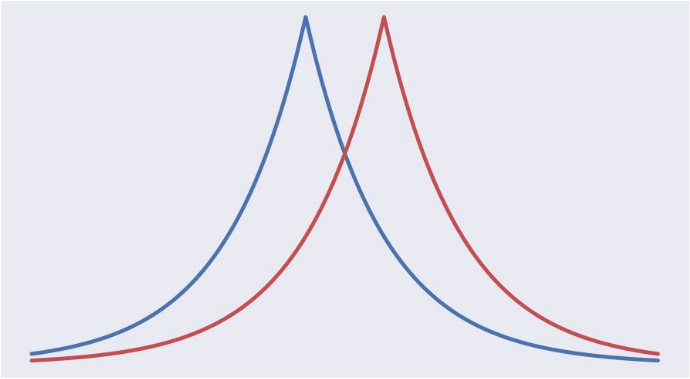
Reveal

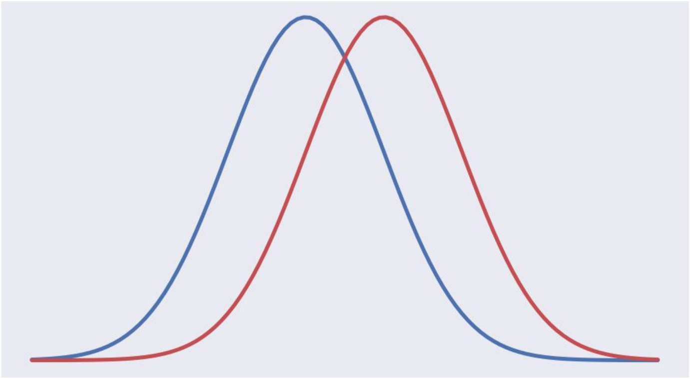


Differential privacy

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:

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



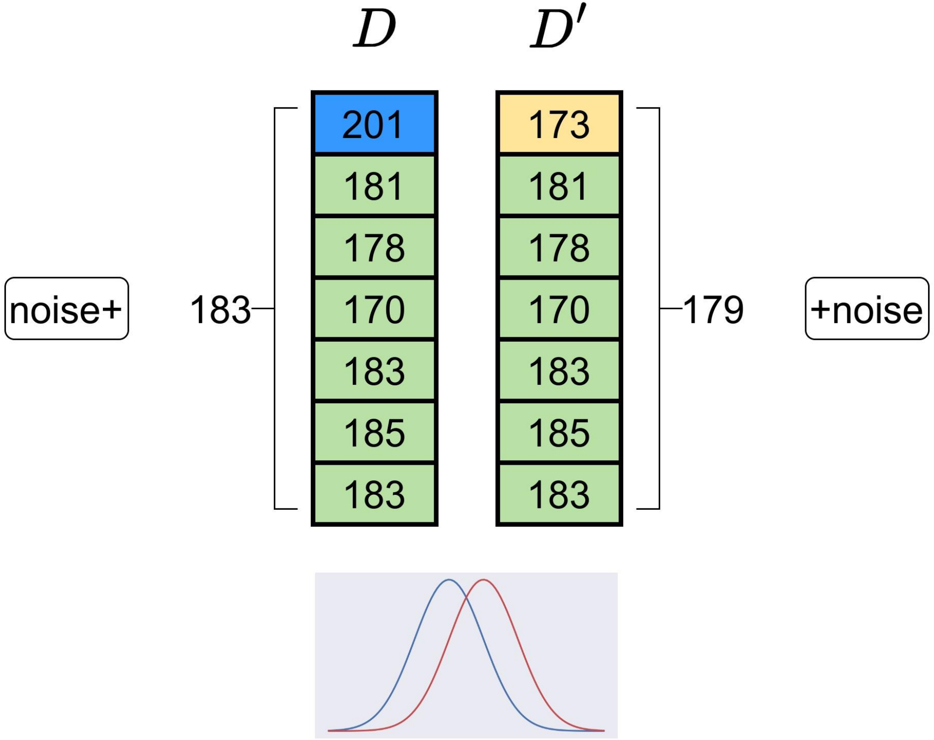


Differential privacy

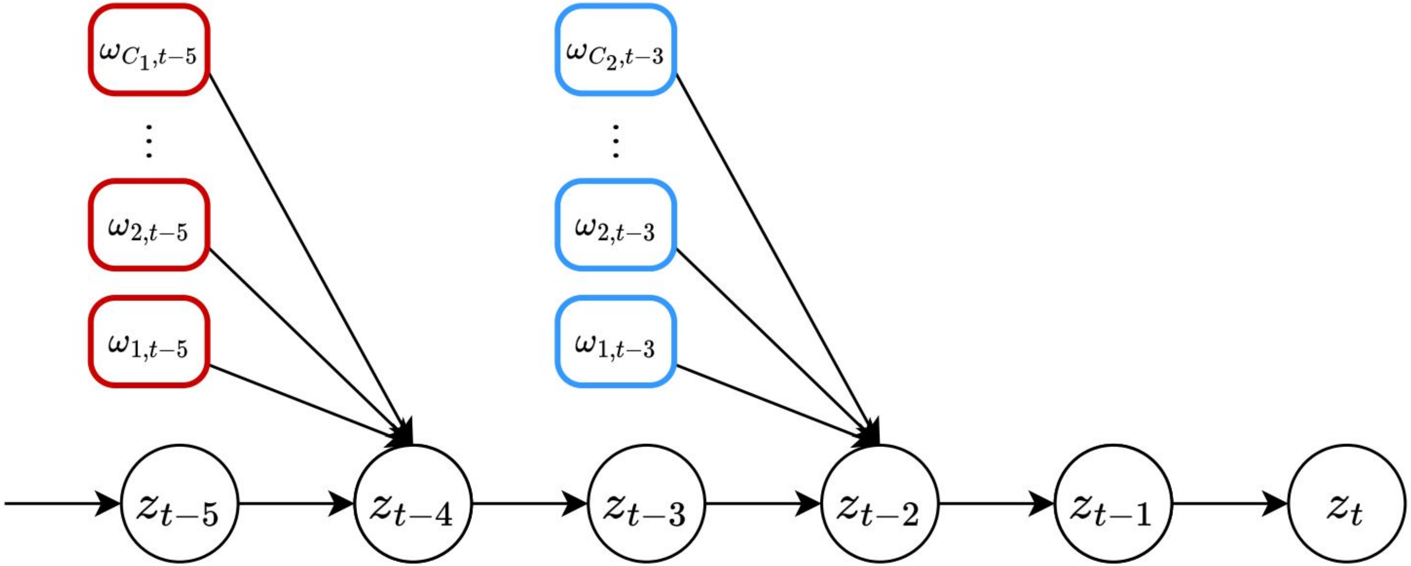
Gaussian Mechanism:

Add noise according to ‘sensitivity’

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



Example graph for a user over seven days



FN can be written as function of product factors

$$\phi_{u,t} = F_1(\omega_{1,t-5}, \omega_{2,t-5}, \cdots, \omega_{C_1,t-5},
\omega_{1,t-3}, \omega_{2,t-3}, \cdots, \omega_{C_2,t-3})$$

"No time to waste", Romijnders et al., AISTATS 2023

FN can be written as function of product factors

$$\phi_{u,t} = F_1(\omega_{1,t-5}, \omega_{2,t-5}, \cdots, \omega_{C_1,t-5}, \\ \omega_{1,t-3}, \omega_{2,t-3}, \cdots, \omega_{C_2,t-3})$$

$$\phi_{u,t} = F_2(\prod_{i=1}^{C_1} \omega_{i,t-5}, \quad \prod_{i=1}^{C_2} \omega_{i,t-3})$$

"No time to waste", Romijnders et al., AISTATS 2023

Log-normals have closed form Renyi divergence

$$D_a(p_u|p_v) = \underbrace{\log\left(\frac{\sigma_v}{\sigma_u}\right) + \frac{1}{2(a-1)}\log\left(\frac{\sigma_v^2}{\sigma_*^2}\right)}_{\text{equals } 0}$$

$$+ \frac{a}{2\sigma_*^2} \cdot (\mu_u - \mu_v)^2$$

$$= \frac{a}{2\sigma_*^2}(\mu_u - \mu_v)^2 = \frac{a}{2C\sigma^2}(\mu_u - \mu_v)^2$$

Optimize a,rho from RDP to DP

Optimization-problem 1.

$$\min_{a,\rho} \frac{a}{\rho}$$

Such that:

$$\rho + \frac{\log \frac{1}{\delta}}{a-1} - \varepsilon = 0$$

$$a = 1 + \frac{d + \sqrt{d(d+\varepsilon)}}{\varepsilon}$$

Renyi Differential Privacy between log-normals

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

Methods to compare with in experimental results

- Traditional contact tracing (Baker et al. 2021)
- Gibbs sampling (Herbrich et al. 2020; Wang et al. 2015)
- Per-message noising of FN (Romijnders et al. 2023)
- DPFN (ours)

Two simulators

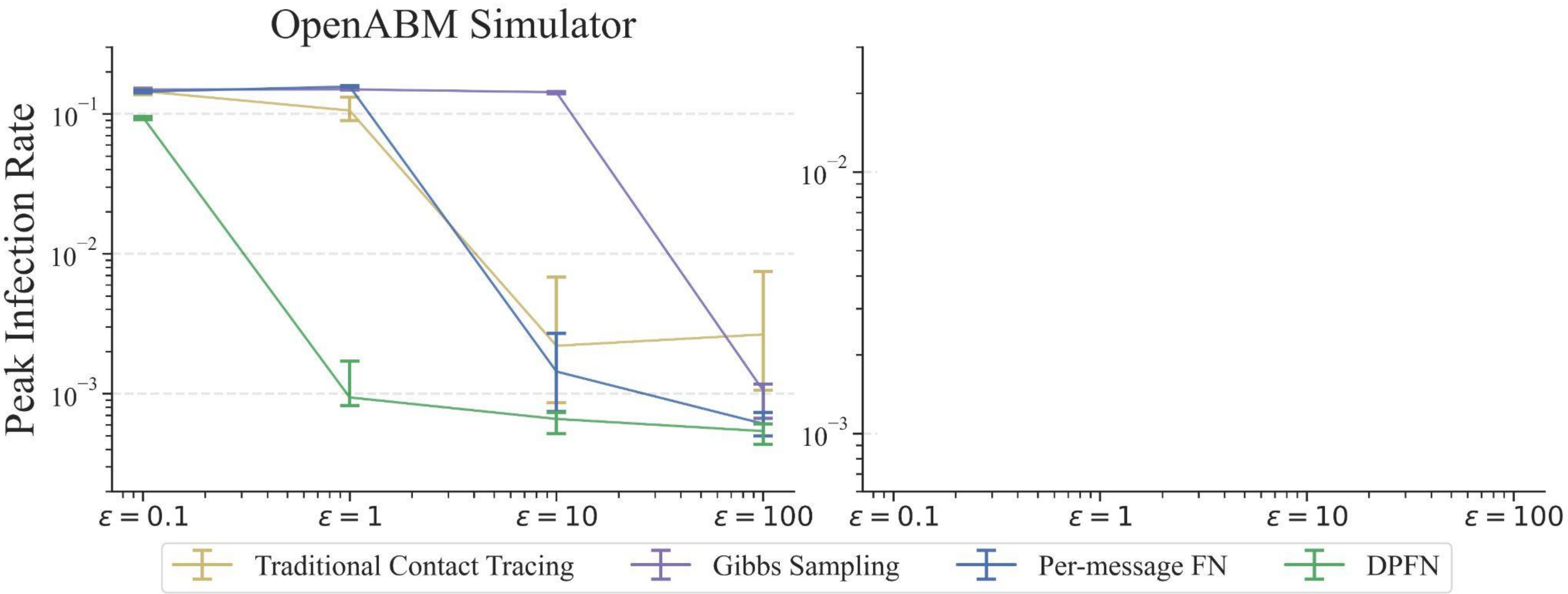
OpenABM (Hinch et al. 2021)

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

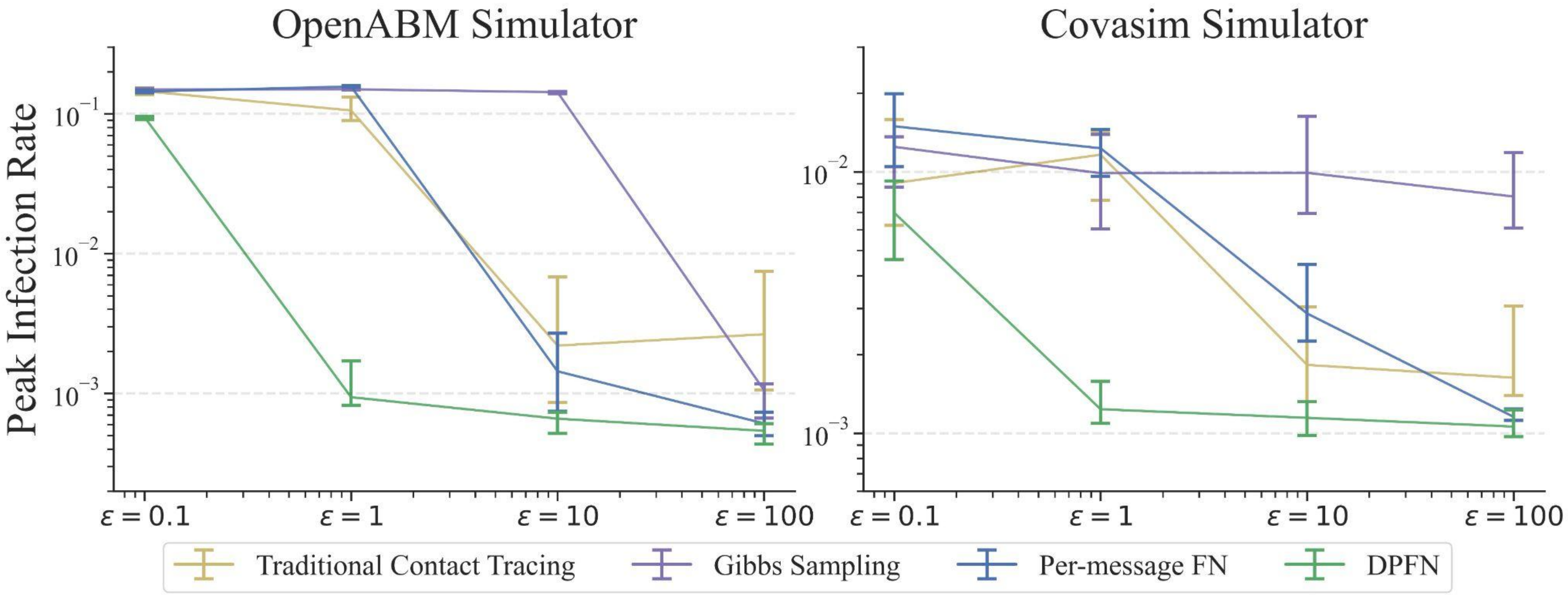
COVASIM (Kerr et al. 2021)

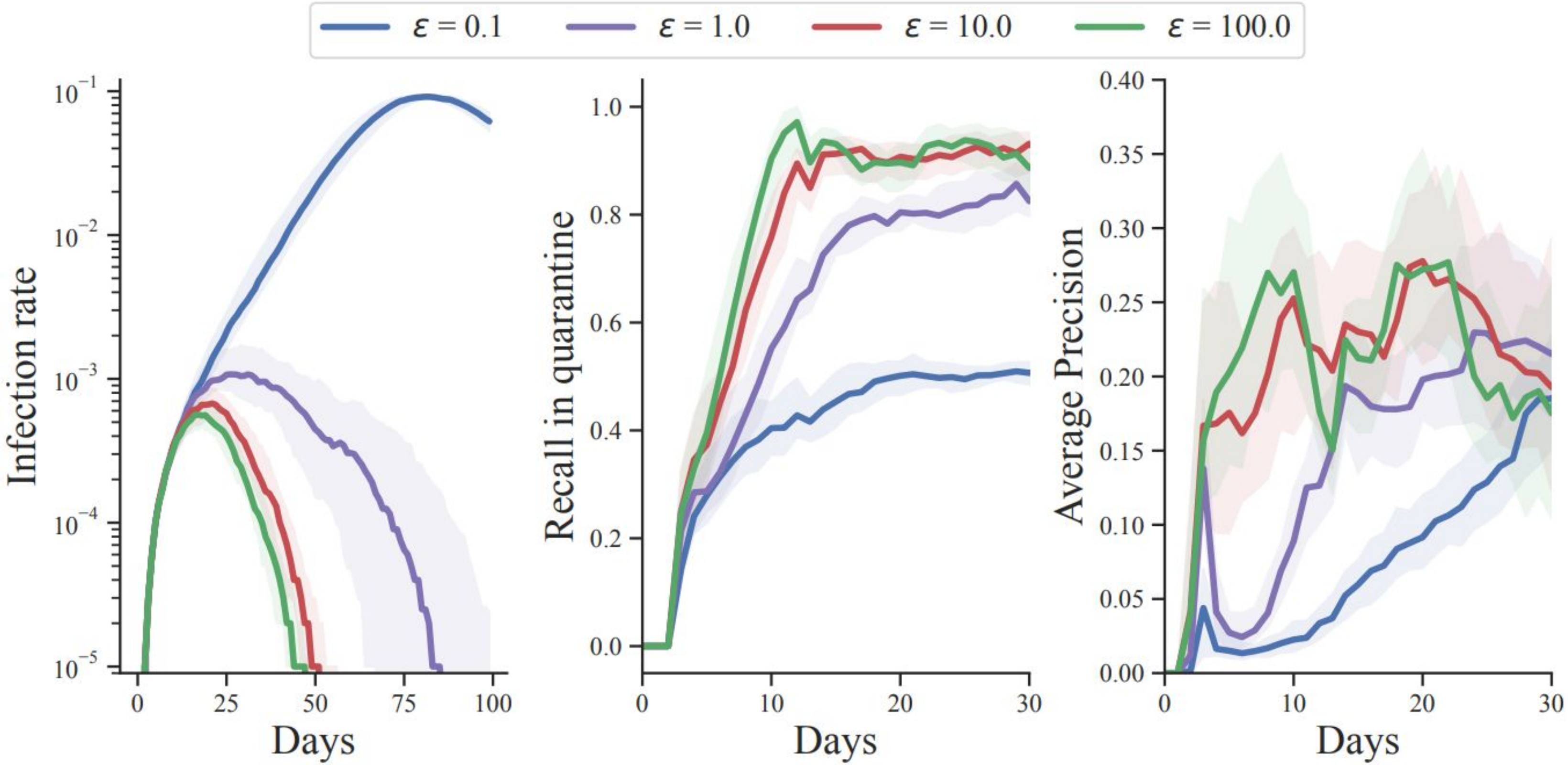
- Contact patterns in layers like households, schools, workplaces, and social communities.
- Calibrated against a typical city in the USA

Results on two widely used simulators

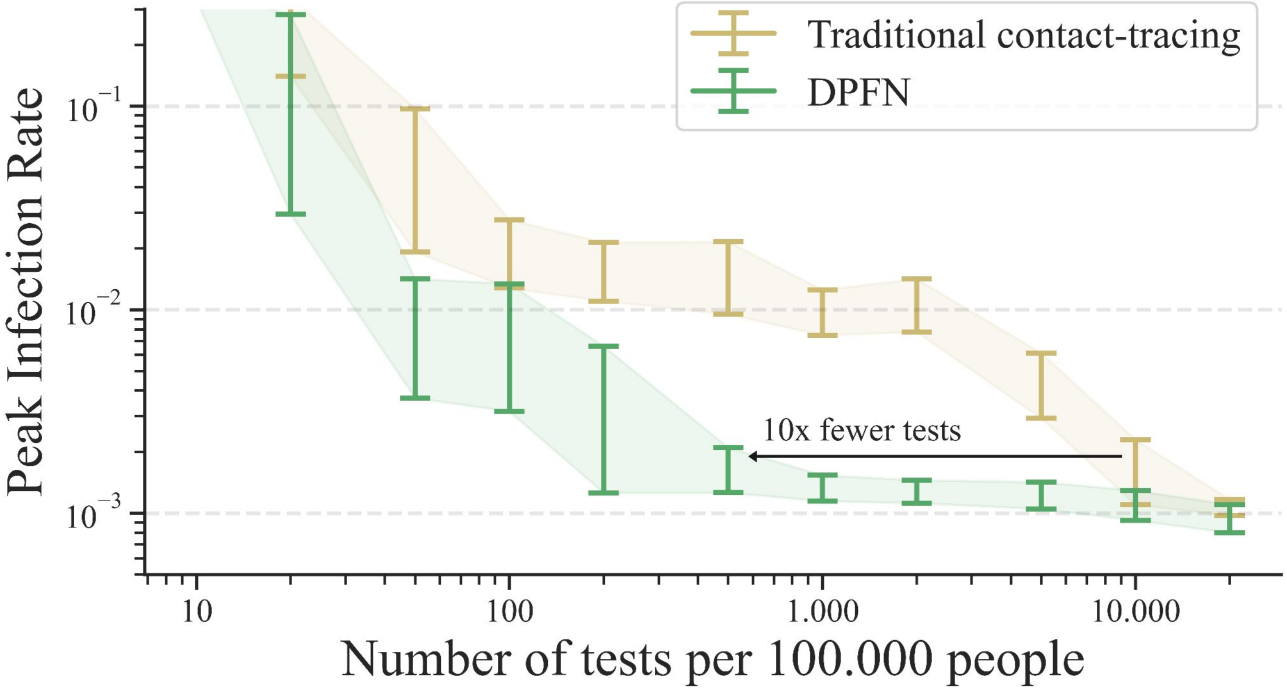


Results on two widely used simulators




ε = 0.1
ε = 1.0
ε = 10.0
ε = 100.0
Infection rate
Days
Recall in quarantine
Days
Average Precision
Days

Our method has low PIR with fewer tests



More testing

Noise effects of DP can be counteracted with additional testing budget

Test setup	No privacy	DPFN	DPFN+
(fpr 0.0%; fnr 0.0%)	0.5 [0.5,0.6]	1.1 [0.7,1.4]	0.6 [0.5,0.7]
(fpr 1%; fnr 0.1%)	0.5 [0.4,0.6]	1.1 [0.9,1.7]	0.6 [0.5,0.6]
(fpr 10%; fnr 1%)	0.6 [0.5,0.8]	17.6 [11.4,20.4]	0.9 [0.7,1.0]
(fpr 25%; fnr 3%)	0.6 [0.5,0.8]	46.6 [40.4,48.0]	0.7 [0.6,0.8]
No testing		200 [190,212]	

Conclusion

- Attack model against contact tracing
- Novel decentralised, differentially private algorithm
- Pareto optimal; save 10x and 2.5x testing budget at low PIR
- Future work:
  - Partial adoption, decentralized reinforcement learning, prosocial participation

UNIVERSITY OF AMSTERDAM
Informatics Institute

Protect Your Score: Contact Tracing with Differential Privacy Guarantees

github.com/robromijnders/dpfn_aaai

r.romijnders@uva.nl; romijndersrob@gmail.com