EPIDEMIC MITIGATION BY STATISTICAL INFERENCE FROM CONTACT TRACING DATA

LENKA ZDEBOROVÁ (EPFL)

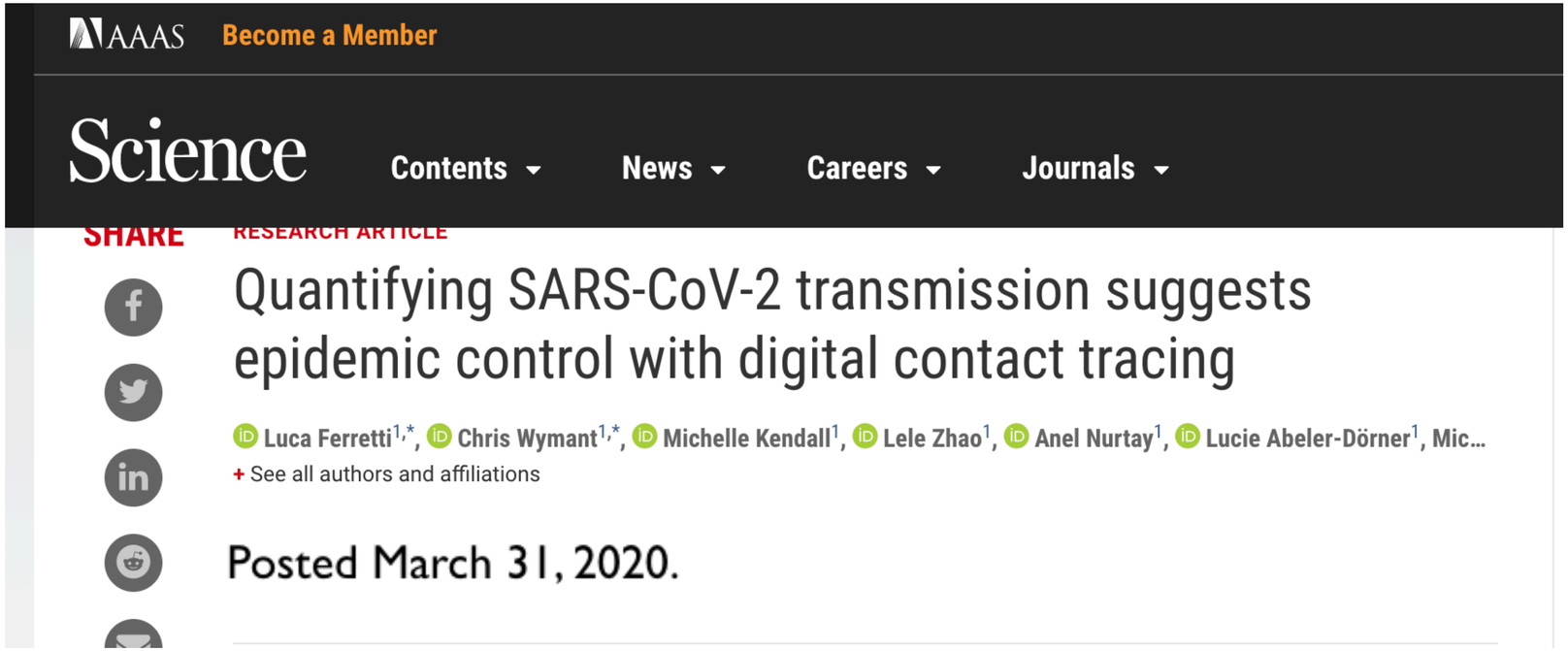
IN COLLABORATION WITH:

A. BAKER[A], I. BIAZZO[B], A. BRAUNSTEIN[B,C,D], G. CATANIA[B], L. DALL'ASTA[B,D], A. INGROSSO[C], F. KRZAKALA[A,E], F. MAZZA[B], M. MÉZARD[A], A. P. MUNTONI[B], M, REFINETTI[A], S, SARAO MANNELLI[F]

ARXIV:2009.09422

IRS 2021, 26/1/2021

We need more efficient strategies to mitigate an epidemic.



Main conclusion: If tests are reported within a day and adoption is good (~30–50%) contact tracing can reduce significantly the epidemic.

CONTACT TRACING DATA

- Information about individuals (stored on the phone of the individual):
  Age, syndromes, health related-risks, results of tests, etc.
- Information about contacts (stored on the phone of the two individuals):
  Time, duration, distance during the contact, barrier-measures used (mask etc.).

CONTACT TRACING: CURRENTLY

- Wikipedia: “COVID-19 apps”
- Contact tracing as mostly implemented currently (Google & Apple, DP3T, etc.): Upon a positive test of an individual, his/her recent, sufficiently close, and long contacts are contacted and advised to be tested or to self-isolate.
- Current status: Test results communicated with several-day delays, and very low adoption rate, very small rate of notifications is tested positive. Resulting in currently negligible influence on the epidemic.

AAAS Become a Member

Science Contents News Careers Journals

SHARE

RESEARCH ARTICLE

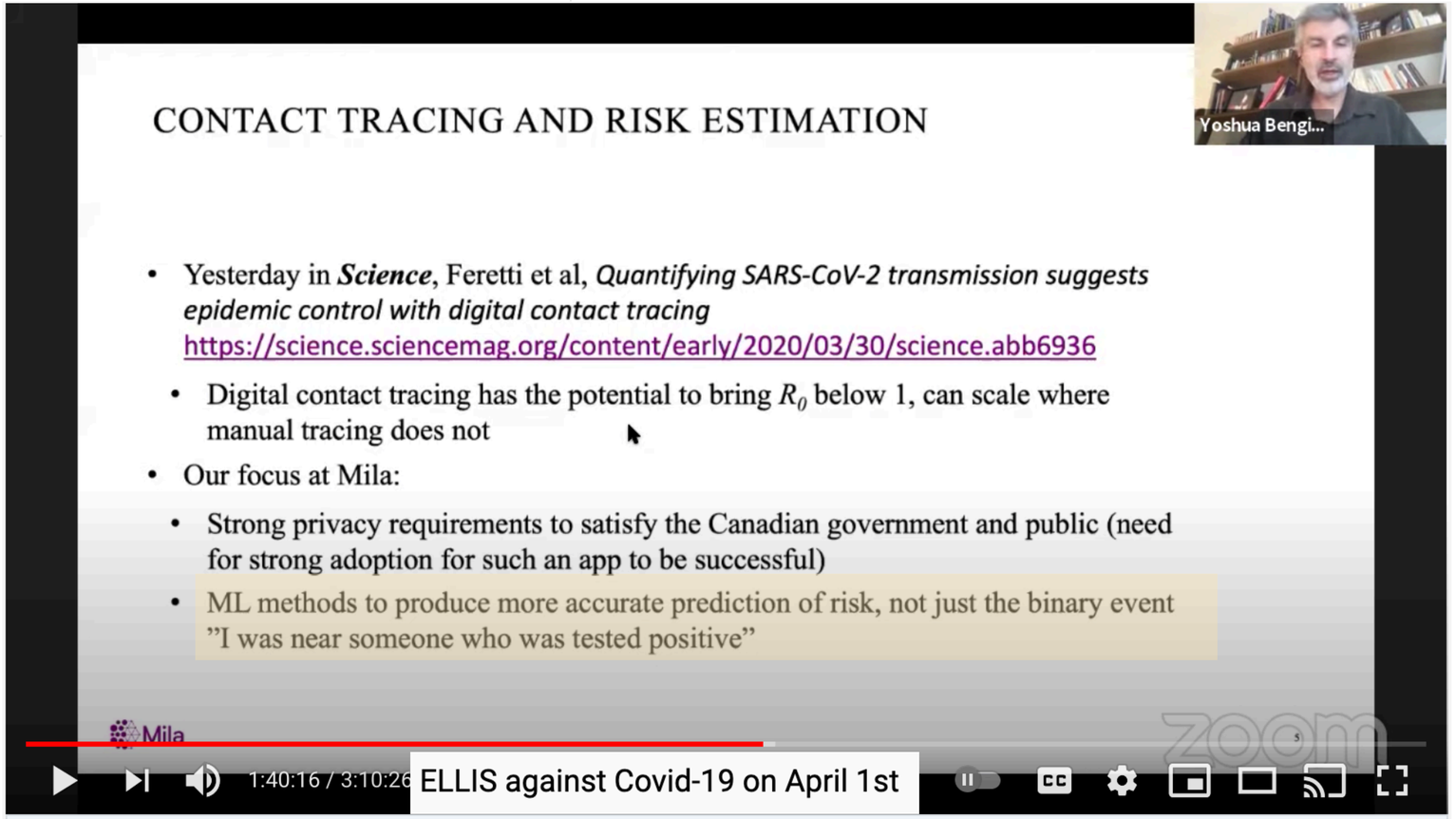
Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing

Luca Ferretti[1,*], Chris Wymant[1,*], Michelle Kendall[1], Lele Zhao[1], Anel Nurtay[1], Lucie Abeler-Dörner[1], Mic…

+ See all authors and affiliations

Posted March 31, 2020.

Main conclusion: If tests are reported within a day and adoption is good (~30–50%) contact tracing can reduce significantly the epidemic.



BETTER THAN CURRENT TRACING: INFERENCE OF RISK

- Risk can be estimated more accurately than listing contacts with infected individuals. Individual should account for increased risks of their neighbours and spread the information to their neighbours.
- What is needed from the app? Communication between individuals who have been in contact (in an encrypted manner, only small bandwidth needed). Exchange of simple messages (probabilities) when in contact.
- Related works:
  - Covi white paper, by Bengio & MILA: 2005.08502
  - ViraTrace (I. Bestvina): https://github.com/ViraTrace/InfectionModel.
  - CRISP: A Probabilistic Model for Infection Risk Estimation 2006.04942

OUR WORK: DEVELOPMENT OF ALGORITHMS FOR RISK INFERENCE

- Belief propagation on trajectories, probabilistic model that conditions the SIR dynamics to the observations. (builds on Altarelli, Braunstein, Dall'Asta et al, PRL'14, Braunstein, Ingrosso Sci. Rep.'16)
- Mean-field risk estimation (builds on Lokhov, Mézard, Ohta, LZ, PRE '14 & PRE '15):

ARXIV:2009.09422

Inferring the origin of an epidemic with a dynamic message-passing algorithm

Andrey Y. Lokhov[1], Marc Mézard[1,2], Hiroki Ohta[1], and Lenka Zdeborová[3]
[1] *LPTMS, Université Paris-Sud and CNRS-UMR 8626, 91405 Orsay, France,*
[2] *Ecole Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France, and*
[3] *IPhT, CEA Saclay and CNRS-URA 2306, 91191 Gif-sur-Yvette, France.*
(Dated: July 3, 2014)

We study the problem of estimating the origin of an epidemic outbreak: given a contact network and a snapshot of epidemic spread at a certain time, determine the infection source. This problem is important in different contexts of computer or social networks. Assuming that the epidemic spread follows the usual susceptible-infected-recovered model, we introduce an inference algorithm based on dynamic message-passing equations and we show that it leads to significant improvement of performance compared to existing approaches. Importantly, this algorithm remains efficient in the case where the snapshot sees only a part of the network.

PACS numbers: 89.75.Hc, 05.20.-y, 02.50.Tt

I. INTRODUCTION

Understanding and controlling the spread of epidemics on networks of contacts is an important task of today's science. It has far-reaching applications in mitigating the results of epidemics caused by infectious diseases, computer viruses, rumor spreading in social media and others. In the present article we address the problem of

more detailed information about the epidemic than just a snapshot at a given time [10]. Note, however, that all the present methods are limited, for instance none of them makes an efficient use of the information about the nodes to which the epidemic did not spread.

In this paper we introduce a new algorithm for the estimation of the origin of an SIR epidemic from the knowledge of the network and the snapshot of some nodes at

STRUCTURE OF THE PROBLEM

- Individual i at time t is in a state $x_i^t \in \{S, I, R\}$; the whole trajectory $\mathbf{x}_i$.
- Prior $P(\{\mathbf{x}_i\}_{i=1}^N)$ is given by a spreading model, its parameters and an initial condition.
- Observations: individual i reports results of tests of symptoms at time t.

- Goal: Compute the probability $P_A^i(t) = P(x_i^t = A), A \in \{S, I, R\}$ over all trajectories compatible with the observations. .

Prior (spreading model) & observations => risk estimation

"PRIOR" FOR INFERENCE: SUSCEPTIBLE–INFECTED–RECOVERED (SIR) AGENT-BASED MODEL

- Population of N individuals
- Spreading of a virus
  - Susceptible individuals (S) → Can be infected
  - Infected individuals (I) → Can infect others
  - Removed individuals (R) → Cannot spread or be infected

$$x_i^t = S \rightarrow \begin{cases} x_i^{t+1} = I \ \text{ w.p. } \ 1 - \prod_{j \in \partial i(t), x_j^t = I} (1 - \lambda_{ij}(t)) \\ x_i^{t+1} = S \ \text{ w.p. } \ \prod_{j \in \partial i(t), x_j^t = I} (1 - \lambda_{ij}(t)) \end{cases}$$

$$x_i^t = I \rightarrow \begin{cases} x_i^{t+1} = I \ \text{ w.p. } 1 - \mu_i \\ x_i^{t+1} = R \ \text{ w.p. } \ \mu_i \end{cases} \qquad x_i^t = R \rightarrow x_i^{t+1} = R$$

Parameters:

- $\lambda_{ij}(t)$ attack rate = probability that if susceptible i meets infected j, j infects i. Depends on the duration and distance of contact, the barrier measures etc
- μ_i: Recovery rate = probability of person i becoming removed in one time-step. Depends on the individual (age, health, etc)

BAYESIAN INFERENCE

- Individual i at time t is in a state $x_i^t \in \{S, I, R\}$; the whole trajectory $\mathbf{x}_i$.
- Prior:

$$P(\{\mathbf{x}_i\}_{i=1}^N) = \prod_{i=1}^{N} \left[p(x_i^{t=0}) \prod_{t=1}^{T} p(x_i^t \mid \partial x_i^{t-1}, x_i^{t-1}) \right]$$

- Including the observations:

$$P(\{\mathbf{x}_i\}_{i=1}^N \mid \mathcal{O}) = \frac{1}{Z(\mathcal{O})} P(\{\mathbf{x}_i\}_{i=1}^N) \prod_{i=1}^{N} p(\mathcal{O}_i \mid \mathbf{x}_i)$$

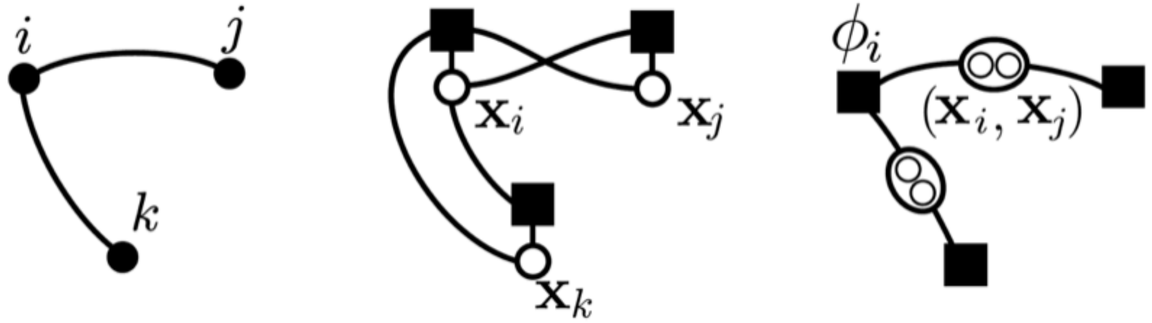
$$P(\{\mathbf{x}_i\}_{i=1}^N \mid \mathcal{O}) = \frac{1}{Z(\mathcal{O})} \prod_{i=1}^{N} \left[p(x_i^{t=0}) p(\mathcal{O}_i \mid \mathbf{x}_i) \prod_{t=1}^{T} p(x_i^t \mid \partial x_i^{t-1}, x_i^{t-1}) \right]$$

- Goal: Compute the probability $P_A^i(t) = P(x_i^t = A), A \in \{S, I, R\}$ as the marginal on the posterior.

GRAPHICAL MODEL

$$P(\{\mathbf{x}_i\}_{i=1}^N \,|\, \mathcal{O}) = \frac{1}{Z(\mathcal{O})} \prod_{i=1}^{N} \left[p(x_i^{t=0}) p(\mathcal{O}_i \,|\, \mathbf{x}_i) \prod_{t=1}^{T} p(x_i^t \,|\, \partial x_i^{t-1}, x_i^{t-1}) \right]$$

- Transformation avoiding short loops:



- Belief Propagation iterative update for probabilities of trajectories (each trajectory at most 2 change points):

$$m_{i\to j}^{n+1}(\mathbf{x}_i, \mathbf{x}_j) = \mathscr{F}_{BP}(\{m_{k\to i}^{n}(\mathbf{x}_k, \mathbf{x}_i)\}_{k\in\partial i})$$

- This is conjectured to give the exact marginals on large random tree-like graphs with independent evolution of contacts and observations.

ALTARELLI, BRAUNSTEIN, DALL'ASTA ET AL, PRL'14, BRAUNSTEIN, INGROSSO SCI. REP.'16, ARXIV:2009.09422

OUR WORK: DEVELOPMENT OF ALGORITHMS FOR RISK INFERENCE

- Belief propagation on trajectories, probabilistic model that conditions the SIR dynamics to the observations. (builds on Altarelli, Braunstein, Dall'Asta et al, PRL'14, Braunstein, Ingrosso Sci. Rep.'16)

- Mean-field risk estimation (builds on Lokhov, Mézard, Ohta, LZ, PRE '14 & PRE '15):

ARXIV:2009.09422

"PRIOR" FOR INFERENCE: SUSCEPTIBLE–INFECTED–RECOVERED (SIR) AGENT-BASED MODEL

- Population of N individuals
- Spreading of a virus
  - Susceptible individuals (S) → Can be infected
  - Infected individuals (I) → Can infect others
  - Removed individuals (R) → Cannot spread or be infected

Parameters:

- $\lambda_{ij}(t)$ attack rate = probability that if susceptible i meets infected j, j infects i. Depends on the duration and distance of contact, the barrier measures etc
- μ_i: Recovery rate = probability of person i becoming removed in one time-step. Depends on the individual (age, health, etc)

- What is the probability of person i to be in state S,I or R at time t? $P_S^i(t)$, $P_I^i(t)$, $P_R^i(t)$

DYNAMICAL MESSAGE PASSING (DMP)

- Given an initial conditions $\{P_S^i(0), P_I^i(0), P_R^i(0)\}_{i=1}^N$ + parameters $\{\mu_i(t), \lambda_{ij}(t)\}_{i=1}^N$

- Lokhov, Mézard, Ohta, LZ, PRE '14 & PRE '15 gave dynamical message passing algorithm to give $\{P_S^i(t), P_I^i(t), P_R^i(t)\}_{i=1}^N$ that are (conjectured to be) asymptotically exact on tree-like graphs as $N \to \infty$

No observations (test results + symptoms) included!!!

MEAN-FIELD MESSAGE PASSING (SIMPLIFICATION & SMALL λ LIMIT OF DMP)

- time evolution equations for $P_S^i(t)$, $P_I^i(t)$, and $P_R^i(t)$

$$P_S^i(t+1) = P_S^i(t)\left(1 - \sum_{j \in \partial i(t)} P_I^j(t)\lambda_{ij}(t)\right)$$

$$P_R^i(t+1) = P_R^i(t) + \mu_i P_I^i(t)$$

$$P_I^i(t+1) = P_I^i(t) + P_S^i(t) \sum_{j \in \partial i(t)} P_I^j(t)\lambda_{ij}(t) - \mu_i P_I^i(t)$$

Parameters:

- $\lambda_{ij}(t)$: Probability that if susceptible i meets infected j, j infects i:
  - depends on the individuals: barrier measures etc
  - depends on time: duration and distance of contact
- μ_i: Recovery probability of person i:
  - depends on the individual (age, health, etc)
- $\partial i(t)$: Sum over ALL the individuals i was in contact with at time t:
  - Tracked with App

- Given an initial conditions $\{P_S^i(0), P_I^i(0), P_R^i(0)\}$ + parameters ➡ $\{P_S^i(t), P_I^i(t), P_R^i(t)\}$

Include the observations (test results + symptoms)

FEED-BACK LOOP: USE TEST RESULTS

$\{P_S^i(0), P_I^i(0), P_R^i(0)\}$ → (MF equations) → $\{P_S^i(t), P_I^i(t), P_R^i(t)\}$

$\{P_S^i(t-\tau), P_I^i(t-\tau), P_R^i(t-\tau)\}$ ← Use Information to update probabilities at t-Δt

$\{P_S^i(t-\tau), P_I^i(t-\tau), P_R^i(t-\tau)\}$ → (MF equations) → $\{P_S^i(t), P_I^i(t), P_R^i(t)\}$

Get better estimates of the probabilities at time t

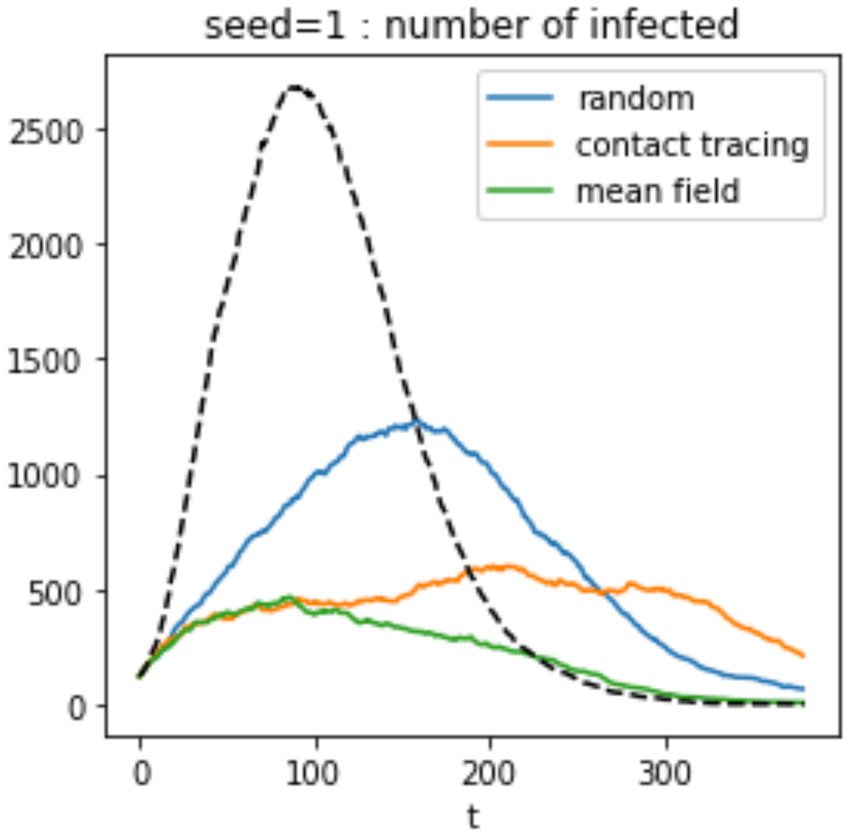
- An infected person has been infectious for the time τ: $P_I^i(t') = 1 \text{ for } t - \tau \leq t' \leq t$
- A recovered person cannot change state: $P_R^i(t') = 1 \text{for} t' \geq t$
- A susceptible individual has not yet been infected: $P_S^i(t') = 1 \text{for} t' \leq t$

→ Repeat the process at every time step → Final Inference

RESULTS

EXPERIMENTS ON CONTROL OF EPIDEMIC, TESTING HIGHEST-RISK NODES

Uncontrolled epidemic
Randomly test & isolate.
Trace & test & isolate.
MF-trace & test & isolate

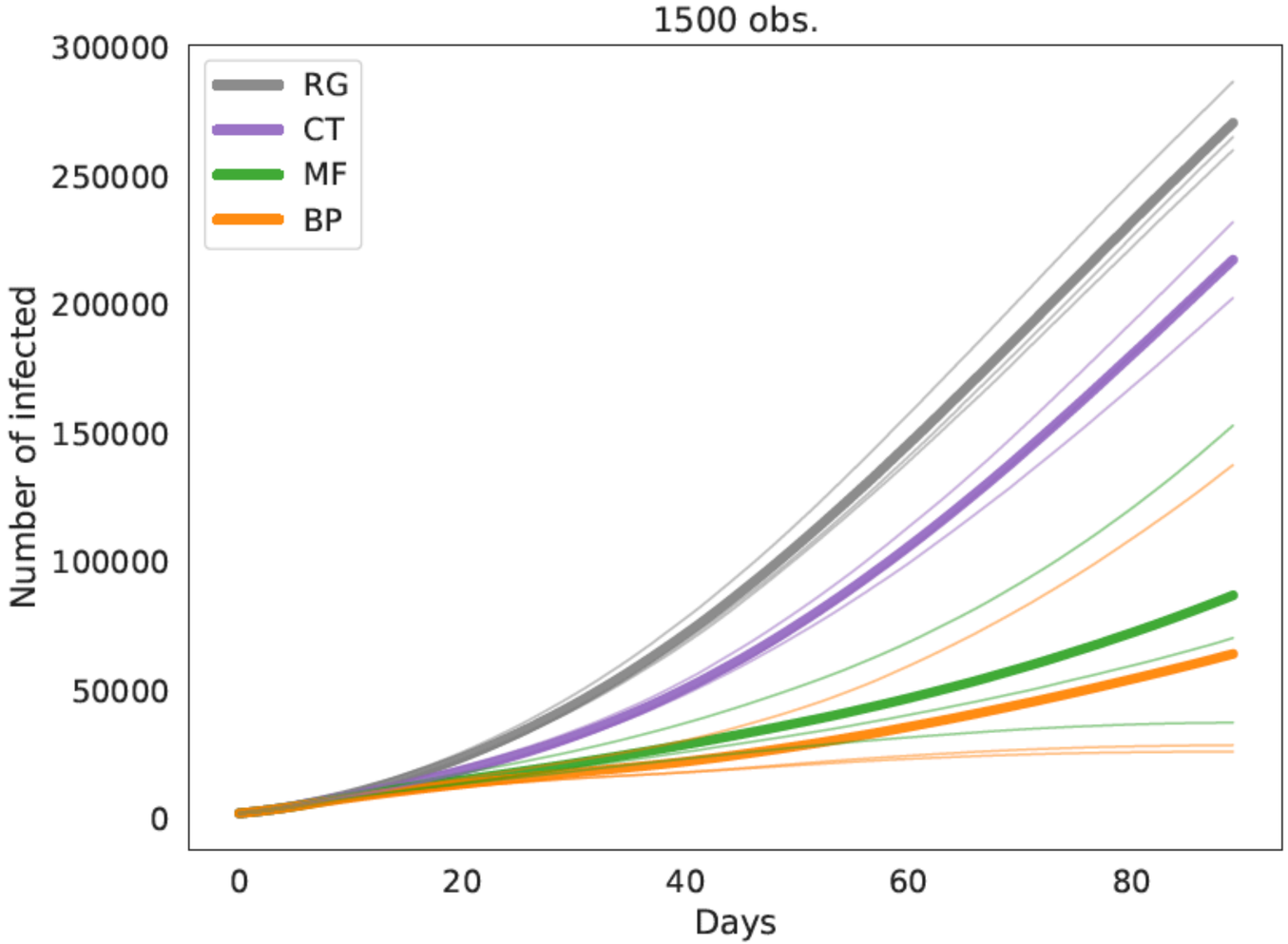


Random geometric contact graph in 2D, scale 1.1, daily on average 7.4 contacts. Population size= 10000,
λ=0.02 , μ=0.03. Initially 20 infected + 10 time steps of uncontrolled evolution,
Tests: symptomatic = 50% of all infected, and 21 from ranking.

EXPERIMENTS ON CONTROL OF EPIDEMIC, TESTING HIGHEST-RISK NODES

Randomly test & isolate.
Trace & test & isolate.
MF-trace & test & isolate.
BP-trace & test & isolate.

MF & BP are scalable and mitigate the epidemic more efficiently than classical tracing.

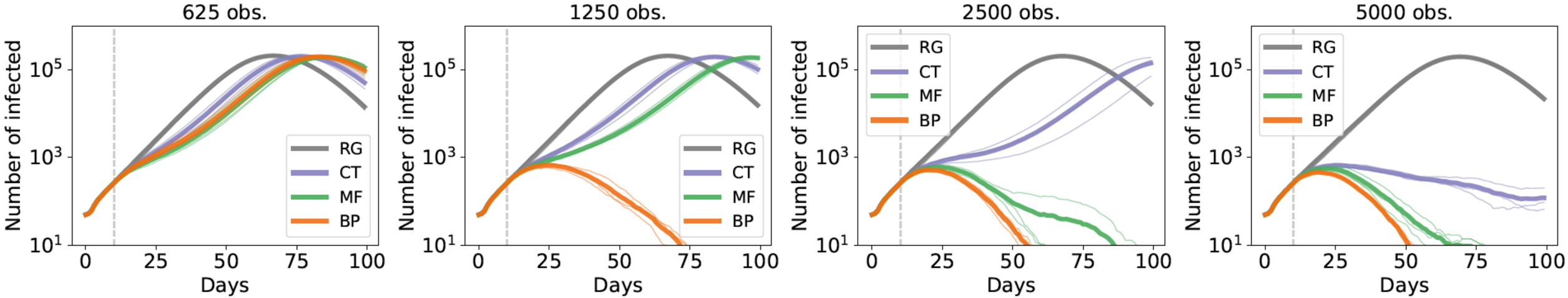


Parameters: N=500 000 individuals, contact network is random geometric graph with 6 contacts a day. Epidemic spread from SIR model with $\lambda = 0.05, \mu = 0.02$, 200 patients zero. Uncontrolled epidemic for first 10 days, then every day we test 50% infected 5 days after their infection, and 1500 tests according to the ranking given by the risk estimation algorithm. Positive individuals are isolated.

ROBUSTNESS EVALUATIONS

- Epidemic spreading model and contact network more realistic and not matching the prior. (OpenABM https://github.com/BDI-pathogens/OpenABM-Covid19 by Hinch et al.)
- Partial usage/adoption of the tracing application.
- False positive & negative tests.

EXPERIMENTS ON CONTROL OF EPIDEMIC: OPEN_ABM

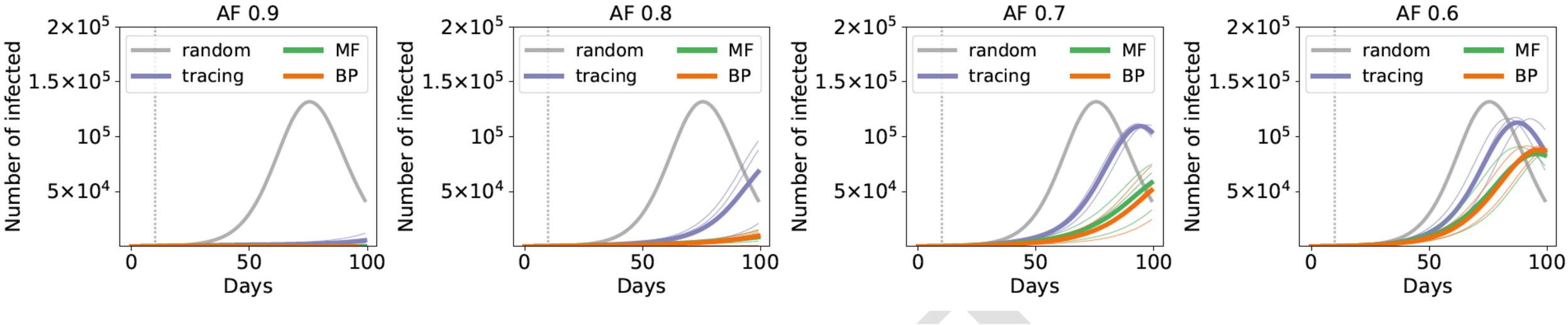


Randomly test & isolate.
Trace & test & isolate.
MF-trace & test & isolate.
BP-trace & test & isolate.

Parameters: N=500 000 individuals, contact network and spread from OpenABM. 50 patients zero. Uncontrolled epidemic for first 10 days, then every day we test 50% infected 5 days after their infection, and X tests according to the ranking given by the risk estimation algorithm. Positive individuals are isolated.

- Key point: Even though the MF/BP inference procedures do not capture most of the details and complexity of the Oxford OPEN_ABM model, they still work and provide large improvement over competing current contact tracing methods.

EXPERIMENTS ON CONTROL OF EPIDEMIC: PARTIAL ADOPTION OF TRACING APP

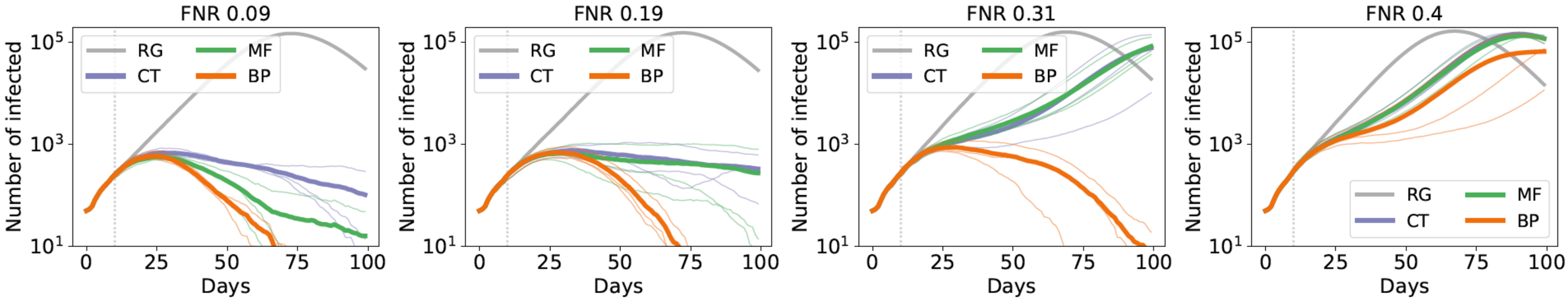


Randomly test & isolate.
Trace & test & isolate.
MF-trace & test & isolate.
BP-trace & test & isolate.

Parameters: N=500 000 individuals, contact network and spread from OpenABM. 50 patients zero. Uncontrolled epidemic for first 10 days, then every day we test 50% infected 5 days after their infection, and 2500 tests according to the ranking given by the risk estimation algorithm. Positive individuals and their households are isolated.

- Point: For these parameters, the performance deteriorates at <70% adoption.

EXPERIMENTS ON CONTROL OF EPIDEMIC: FALSE TESTS



Randomly test & isolate.
Trace & test & isolate.
MF-trace & test & isolate.
BP-trace & test & isolate.

Parameters: N=500 000 individuals, contact network and spread from OpenABM. 50 patients zero. Uncontrolled epidemic for first 10 days, then every day we test 50% infected 5 days after their infection, and 5000 tests according to the ranking given by the risk estimation algorithm. Positive individuals and their households are isolated.

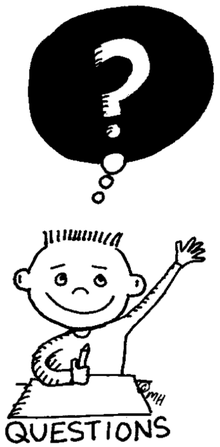
- Point: For these parameters, even 20% false negative tests are supported.

CONCLUSION:

- Probabilistic estimation of risks allows more efficient control of the epidemic.

ONGOING WORK:

- Embedding in other realistic agent-based simulators.
- Learn parameters of the algorithm from observed data (neural-enhanced risk estimation).



ARXIV:2009.09422