Epidemic mitigation by statistical inference from contact tracing data

Baker A, Biazzo I, Braunstein A, Catania G, Dall'Asta L, Ingrosso A, Krzakala F, Mazza F, Mézard M, Muntoni AP, Refinetti M, Mannelli S, Zdeborová L

Anna Paola Muntoni

Italian Institute for Genomic Medicine - Politecnico di Torino

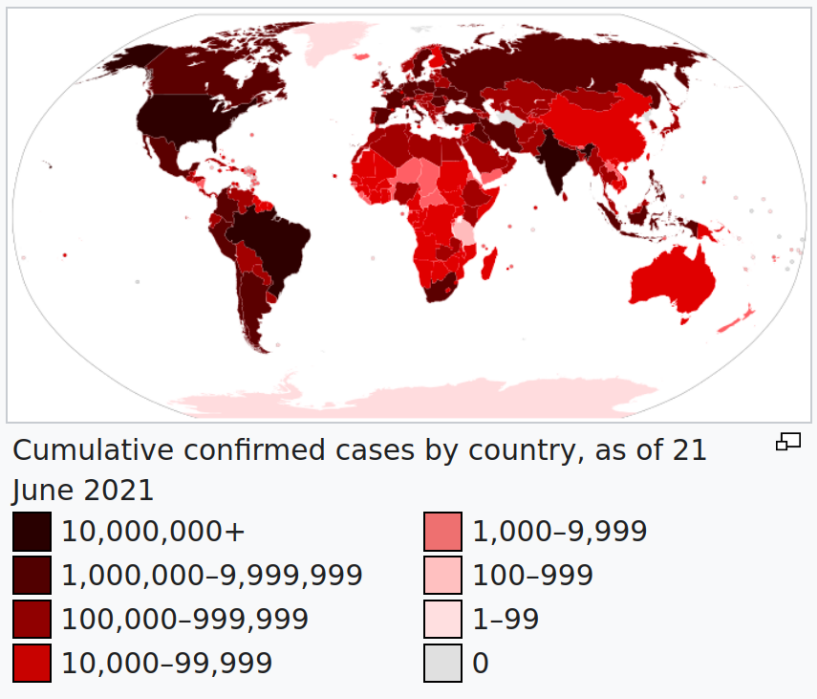
XXV National Conference on Statistical Physics and Complex Systems
Parma, 24th June 2021

Introduction

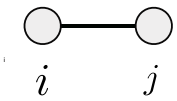
SARS-CoV-2 is the virus responsible for COVID-19 disease.

The virus spreads through both symptomatic and asymptomatic individuals: it is hard to mitigate the contagion.

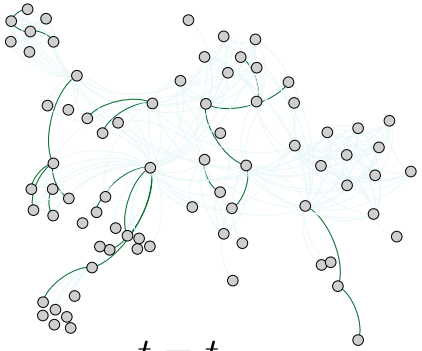
The most effective strategy to contrast the epidemic spreading consists in tracing, testing, and isolating the individuals that have been recently in contact with confirmed cases.

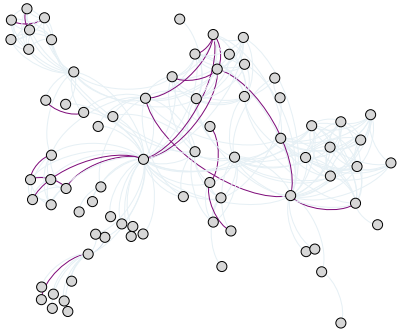


Manual contact tracing

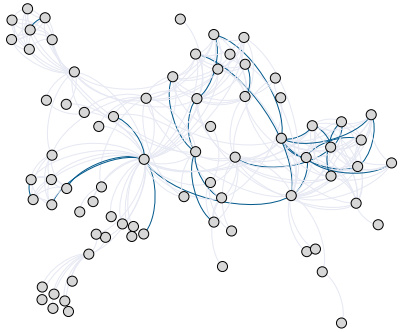


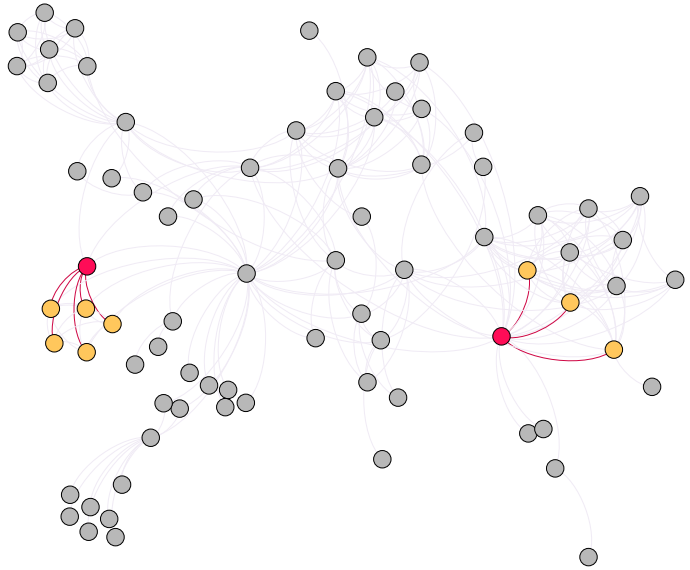
Two individuals are in contact at time t





...



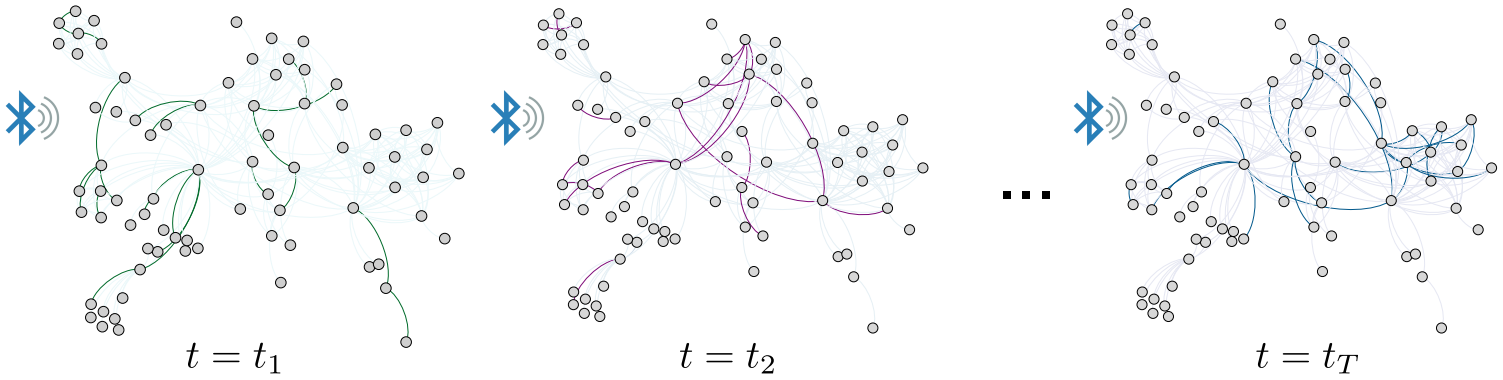


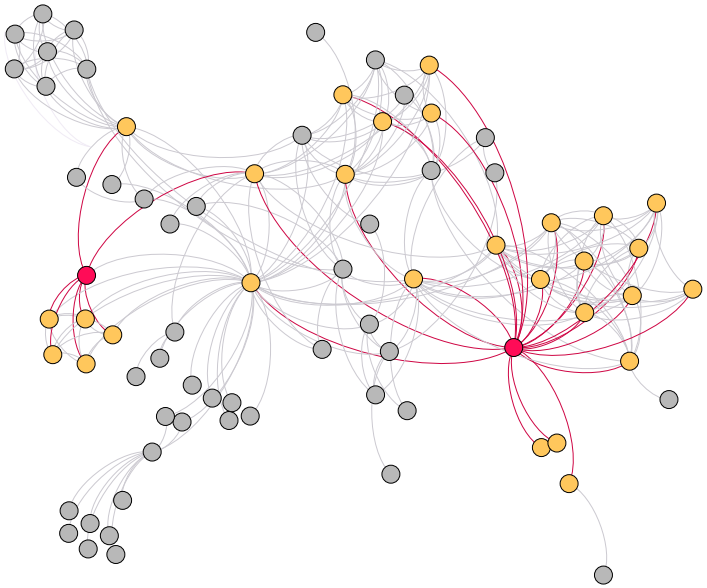
Whenever an infected individual is detected, only a fraction of his/her recent contacts is reached and tested.

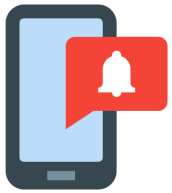
- Tested-positive node
- Individuals to be tested

Digital contact tracing

Trace the contacts using the Bluetooth







Whenever an infected individual is detected, all his/her contacts receive a notification

- Tested-positive node
- Individuals to be tested

Digital contact tracing

A combination of the two strategies has appeared to be beneficial in some context [1][2]; however, it has some drawbacks:

- The notification does not quantify how likely an individual is infected;
- As the epidemic spreads, the number of individuals to be tested increases dramatically (leading to a misuse of medical resources).

Do we obtain a better tracing if we assume a **model** for the *propagation*? This allows for the **reconstruction** of the process and the assessment of **individual risk**.

---

[1] C. Wymantet al., The epidemiological impact of the NHS COVID-19 app. Nature (2021)

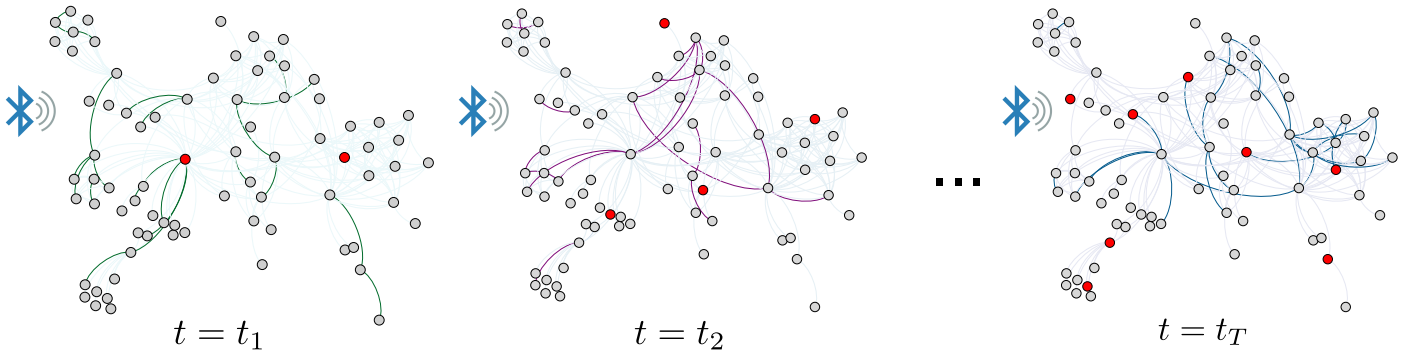
[2] L Ferretti, et al., Quantifying Sars-Cov-2 transmission suggests epidemic control with digital contact tracing. *Science* **368** (2020)

Probabilistic risk assessment

Let us define the (unknown) collective trajectory of N individuals at time t

$$\boldsymbol{x}(t) = \{x_1(t), \cdots, x_N(t)\}$$

and a set of observations (results of medical tests) $\mathcal{O}$



The probability of a trajectory $\boldsymbol{x}$ given the evidence $\mathcal{O}$ is

$$P[\boldsymbol{x}|\mathcal{O}] = P[\mathcal{O}|\boldsymbol{x}]\, P[\boldsymbol{x}]\, P[\mathcal{O}]^{-1}$$

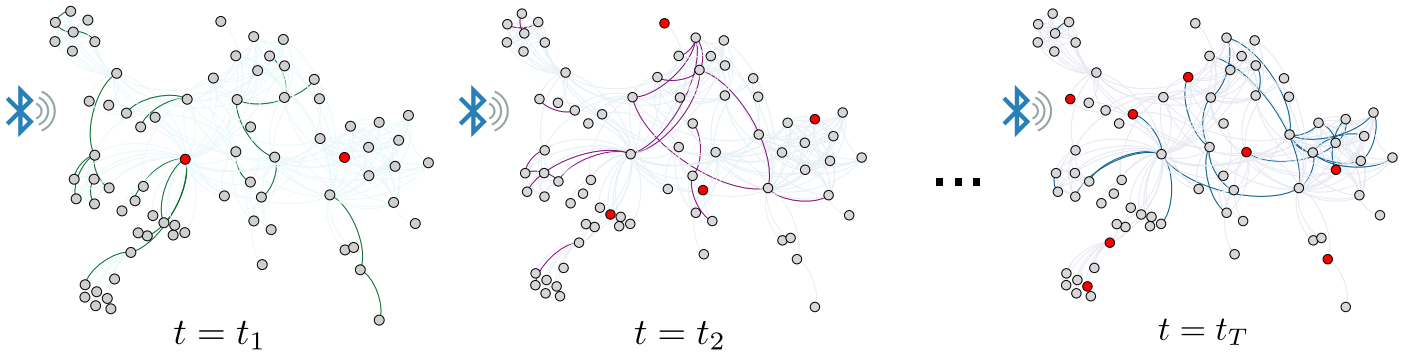
The *a posteriori* probability of each trajectory, given the contact network and the observations

Probabilistic risk assessment

Let us define the (unknown) collective trajectory of N individuals at time t

$$\boldsymbol{x}(t) = \{x_1(t), \cdots, x_N(t)\}$$

and a set of observations (results of medical tests) $\mathcal{O}$



The probability of a trajectory $\boldsymbol{x}$ given the evidence $\mathcal{O}$ is

$$P[\boldsymbol{x}|\mathcal{O}] = P[\mathcal{O}|\boldsymbol{x}] P[\boldsymbol{x}] P[\mathcal{O}]^{-1}$$

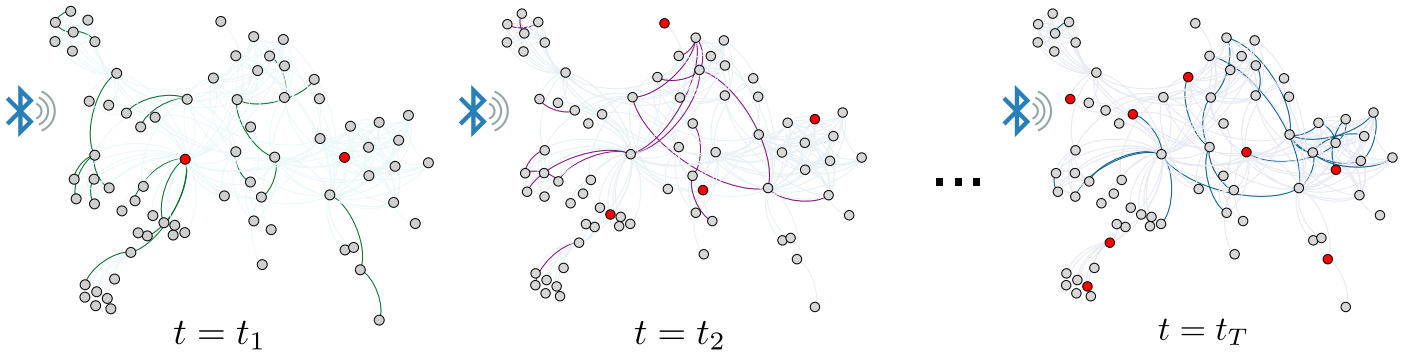
The *likelihood* of the observations, given the trajectories

Probabilistic risk assessment

Let us define the (unknown) collective trajectory of N individuals at time t

$$\boldsymbol{x}(t) = \{x_1(t), \cdots, x_N(t)\}$$

and a set of observations (results of medical tests) $\mathcal{O}$



The probability of a trajectory $\boldsymbol{x}$ given the evidence $\mathcal{O}$ is

$$P[\boldsymbol{x}|\mathcal{O}] = P[\mathcal{O}|\boldsymbol{x}]\, P[\boldsymbol{x}]\, P[\mathcal{O}]^{-1}$$

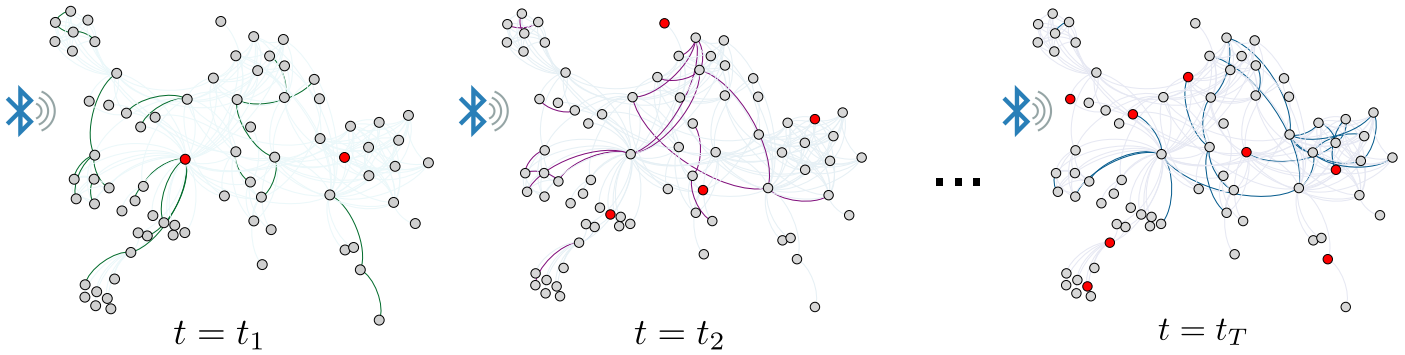
The *evidence* of the observations, independent of the trajectories and therefore neglected

Probabilistic risk assessment

Let us define the (unknown) collective trajectory of N individuals at time t

$$\boldsymbol{x}(t) = \{x_1(t), \cdots, x_N(t)\}$$

and a set of observations (results of medical tests) $\mathcal{O}$



The probability of a trajectory $\boldsymbol{x}$ given the evidence $\mathcal{O}$ is

$$P[\boldsymbol{x}|\mathcal{O}] = P[\mathcal{O}|\boldsymbol{x}] P[\boldsymbol{x}] P[\mathcal{O}]^{-1}$$

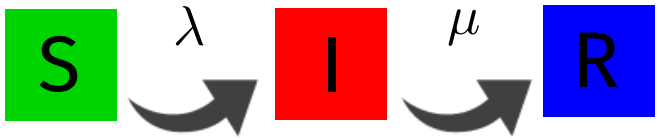
The *prior* probability of the trajectories, i.e. the assumed epidemic model.

SIR model

A simple but effective model is the Susceptible-Infected-Recovered (SIR) model

i

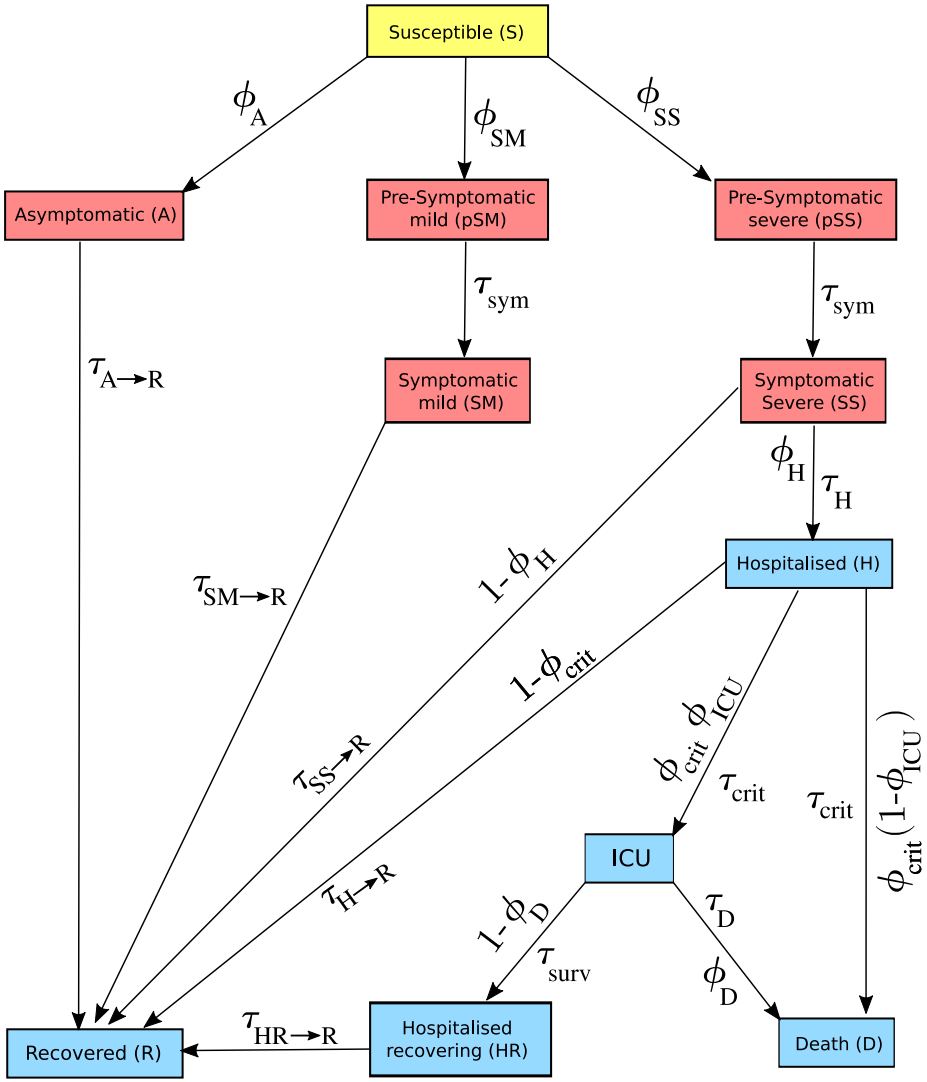
$x_i(t) \in \{S, I, R\}$



Transition rules:
- If $x_i(t) = I$, he/she can recover/die with probability μ, changing state as $x_i(t+1) = R$
- If (i, j) are in contact at time t and $x_i(t) = I, x_j(t) = S$, j can get infected with probability λ
- For each node i the probability of infection at time t is $\Lambda_i = 1 - \prod_{j \neq i} (1 - \lambda_{i,j} \delta_{x_j(t), I})$

Unfortunately, it does not catch the details of the COVID-19 epidemics...

Realistic epidemic model: OpenABM[1]



Mapping:

Infection and recovery probabilities depend on the time elapsed since infection

$$\lambda_{ij}^{ABM}(\Delta t) = \Gamma(\Delta t | \mu_I, \sigma_I)$$

$$\mu_I = 6\,days,\ \sigma_I = 2.5\,days$$

$$\tau_R \sim \Gamma\left(\mu_R^{ABM}, \sigma_R^{ABM}\right)$$

$$\mu_R^{ABM} = 16\,days,\ \sigma_R^{ABM} = 5.6\,days$$

[1] L Ferretti, et al., Quantifying Sars-Cov-2 transmission suggests epidemic control with digital contact tracing. *Science* **368** (2020)

Prior and likelihood

The *prior* reads:

$$P\left[\boldsymbol{x}\left(t_1\right), \cdots, \boldsymbol{x}\left(t_T\right)\right] = P\left[\boldsymbol{x}\left(t_1\right)\right] \prod_{t'=t_2}^{t_T} P\left[\boldsymbol{x}\left(t'\right) | \boldsymbol{x}\left(t'-1\right), \cdots, \boldsymbol{x}\left(t_1\right)\right]$$

Assumed initial condition (number of patient zeros)

Prior and likelihood

The *prior* reads:

$$P\left[\boldsymbol{x}\left(t_1\right), \cdots, \boldsymbol{x}\left(t_T\right)\right] = P\left[\boldsymbol{x}\left(t_1\right)\right] \prod_{t'=t_2}^{t_T} P\left[\boldsymbol{x}\left(t'\right) | \boldsymbol{x}\left(t'-1\right), \cdots, \boldsymbol{x}\left(t_1\right)\right]$$

Non-Markovian process.
For COVID-19, a memory term of two or three weeks is sufficient.

Prior and likelihood

The *prior* reads:

$$P\left[\boldsymbol{x}\left(t_1\right), \cdots, \boldsymbol{x}\left(t_T\right)\right] = P\left[\boldsymbol{x}\left(t_1\right)\right] \prod_{t'=t_2}^{t_T} P\left[\boldsymbol{x}\left(t'\right) | \boldsymbol{x}\left(t'-1\right), \cdots, \boldsymbol{x}\left(t_1\right)\right]$$

The *likelihood* can cope with the diagnostic errors

$$P\left[\mathcal{O}_r | \, \boldsymbol{x}\left(t\right)\right] = \left(1 - p_{FNR}\right) \delta_{x_r(t),I} + p_{FPR}\left(1 - \delta_{x_r(t),I}\right)$$

p_{FNR}: the probability that a test resulted negative is instead positive

p_{FPR}: the probability that a test resulted positive is instead negative

Compute the posterior

In the work

- Baker, A. et al. Epidemic mitigation by statistical inference from contact tracing data. arXiv:2009.09422 (Accepted in PNAS)
  `https://github.com/sibyl-team/epidemic_mitigation`

we show how to apply a Simple Mean Field [3](SMF) and the Belief Propagation[4] [5](BP) approximations to the unknown posterior.

SMF

- It provides an approximation to the marginal probabilities $P_{\mathrm{MF}}\left[x_i\left(t\right)=S\right]$, $P_{\mathrm{MF}}\left[x_i\left(t\right)=I\right]$, $P_{\mathrm{MF}}\left[x_i\left(t\right)=R\right]$
- Faster than BP but less accurate:
  - It heuristically introduces the observations $\mathcal{O}$ and the propagation model is the Markovian SIR

[3] AY Lokhov et al. *Phys. Rev. E* **90** (2014)

[4] F Altarelli et al. *Phys. Rev. Lett.* **112, 118701** (2014)

[5] F Altarelli et al. *Phys. Rev. X* **4, 021024** (2014)

Belief Propagation

Let us define two sets of auxiliary variables,
the infection times $\{\boldsymbol{z} : z_i = \min\{t : x_i(t) = I\}\}$
and the recovery times $\{\boldsymbol{r} : r_i = \min\{t : x_i(t) = R\}\}$

The trajectories $\{\boldsymbol{x}(t_1), \cdots, \boldsymbol{x}(T)\}$ are fully determined by the infection and recovery times

$$P[\boldsymbol{x}(t) | \boldsymbol{z}, \boldsymbol{r}] = \prod_i \psi_i(x_i(t), z_i, r_i)$$

$$\begin{aligned}\psi_i(x_i(t), z_i, r_i) = {} & \delta_{x_i(t),S} \mathbb{I}[t < z_i] + \\ & \delta_{x_i(t),I} \mathbb{I}[z_i \leq t < r_i] + \\ & \delta_{x_i(t),R} \mathbb{I}[t \geq r_i]\end{aligned}$$

Belief Propagation

.

Modelling infective contacts

s_{ij} : time at which i (I) infects j (S)

$$S_{ij}\left(s_{ij}|z_i,r_i\right)=\mathbb{I}\left[z_i<s_{ij}<r_i\right]\lambda_{ij}^{\mathrm{ABM}}\left(s_{ij}-z_i\right)\prod_{z_i<s<s_{ij}}\left[1-\lambda_{ij}^{\mathrm{ABM}}\left(s-z_i\right)\right]$$
$$+\delta_{s_{ij},\infty}\prod_{s\geq r_i}\left[1-\lambda_{ij}^{\mathrm{ABM}}\left(s-z_i\right)\right]$$

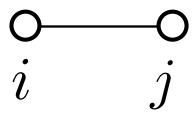
Recovery delay $R_i\left(r_i-z_i\right)=\Gamma\left(r_i-z_i|\mu_R^{ABM},\,\sigma_R^{\mathrm{ABM}}\right)$

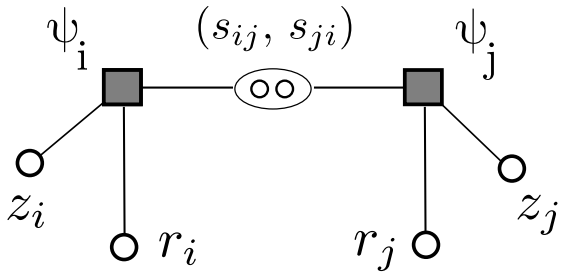
Target *a posteriori* probability

$$P\left(\boldsymbol{z},\boldsymbol{r},\boldsymbol{s}|\mathcal{O}\right)\propto\prod_i\delta_{z_i,\min_{k\in\partial i}s_{ki}}R_i\left(r_i-z_i\right)P\left(\mathcal{O}_i|z_i,r_i\right)\prod_{(ij)}S_{ij}\left(s_{ij}|z_i,r_i\right)$$

Belief Propagation

From the contacts, one has to build the corresponding factor graph





$$m_{i\to j}\left(s_{ij}, s_{ji}\right)$$

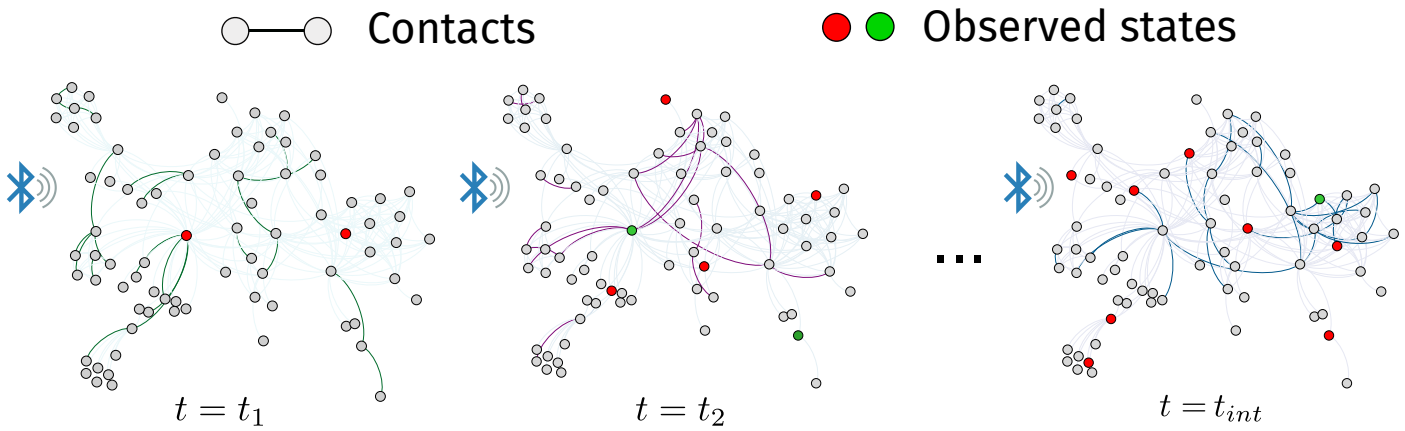
By iteratively solving the fixed point equations of the messages, one can compute the *beliefs* associated with the infection times

$$b_i\left(z_i\right) \propto \quad \text{Probability that node } i \text{ get infected at time } z_i$$

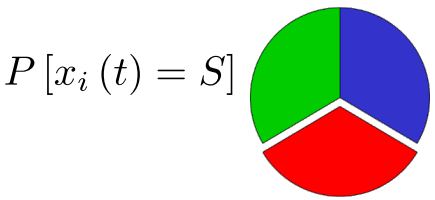
and the risk at time t_{int}

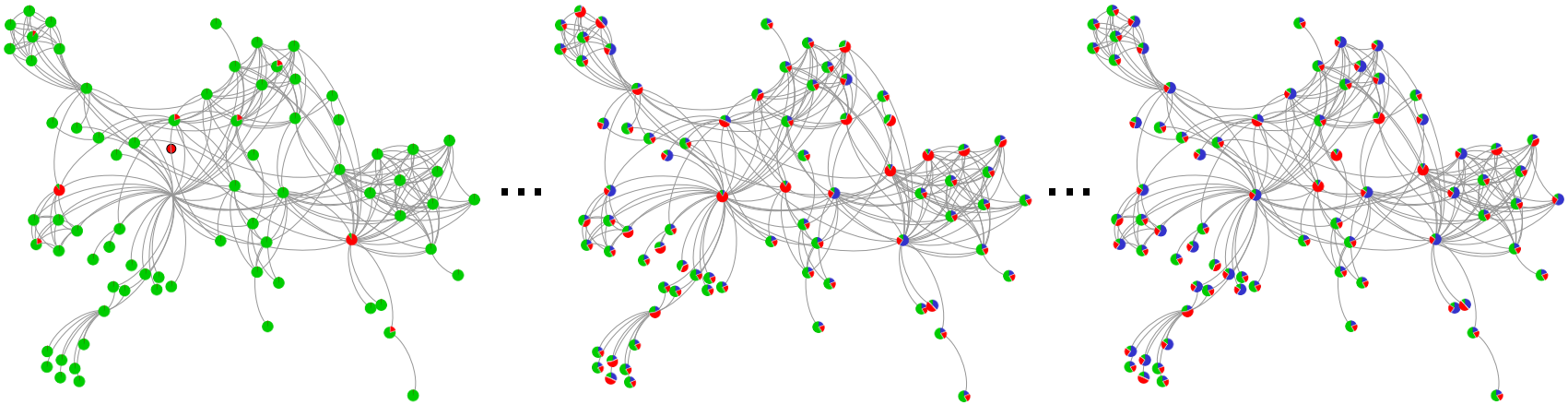
$$\text{risk}\left(i\right) = \sum_{t=t_{int}-\tau}^{t_{int}} b_i\left(t\right)$$

Risk tracing

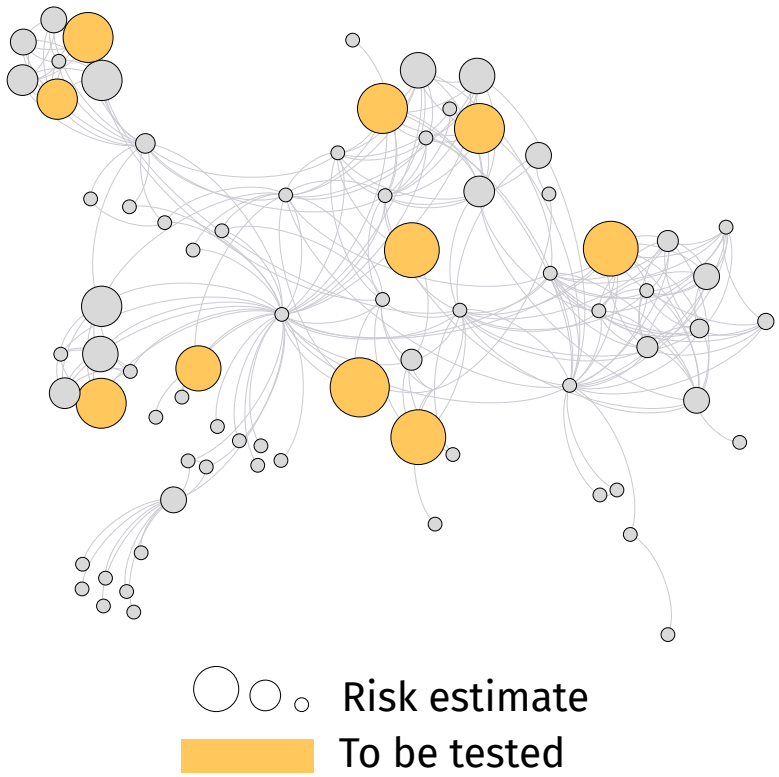


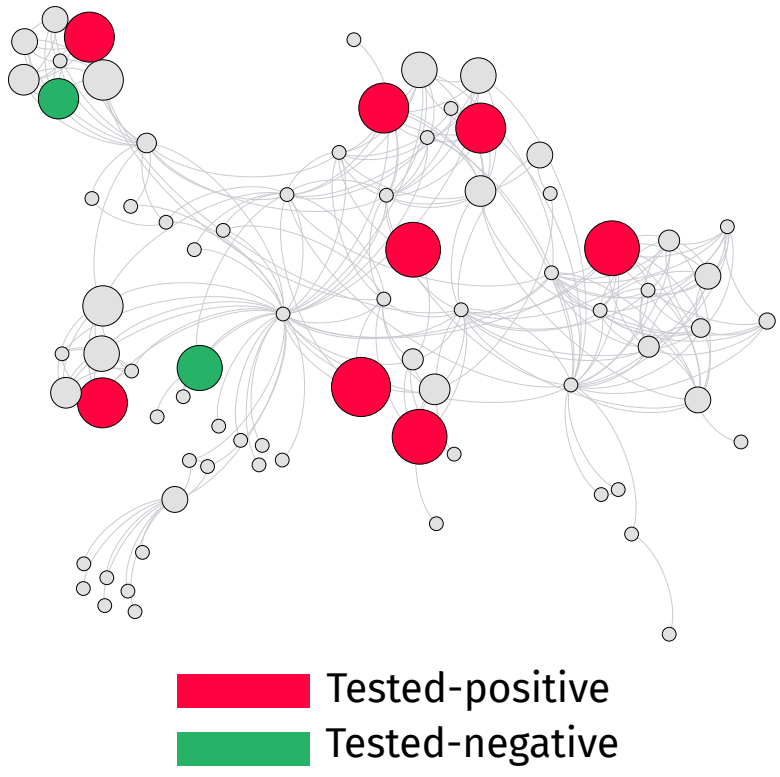
Reconstruction step





Risk tracing





Results

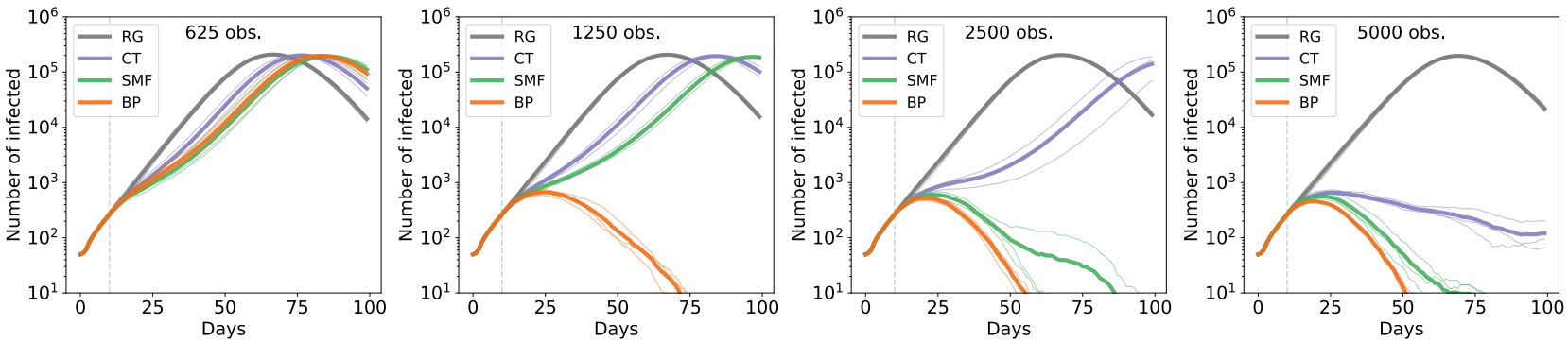
We simulate realistic epidemic outbreaks using OpenABM while the model used to perform the reconstruction is the SIR.

The contact network is a superposition of household, workplace and random networks.

We let the epidemics evolve up to 10 days; then we perform the reconstruction and the testing/quarantine step each day. Severe symptomatic individuals are automatically confined.

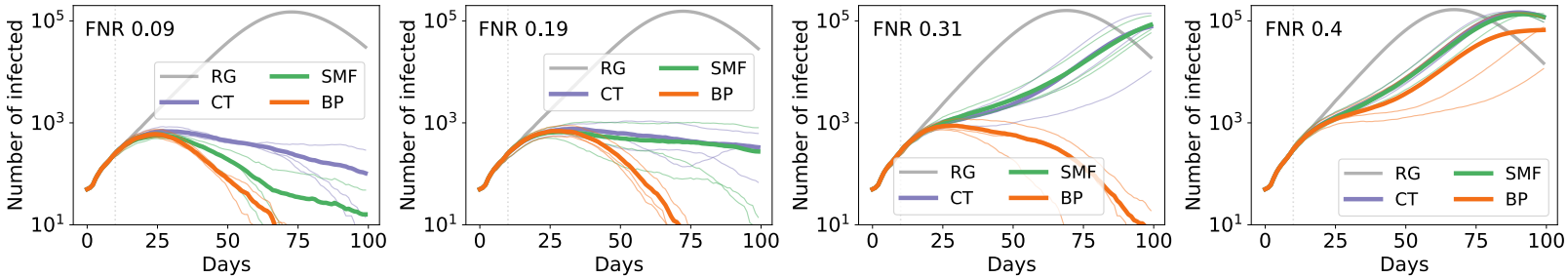
RG : Random guessing

CT: Contact tracing

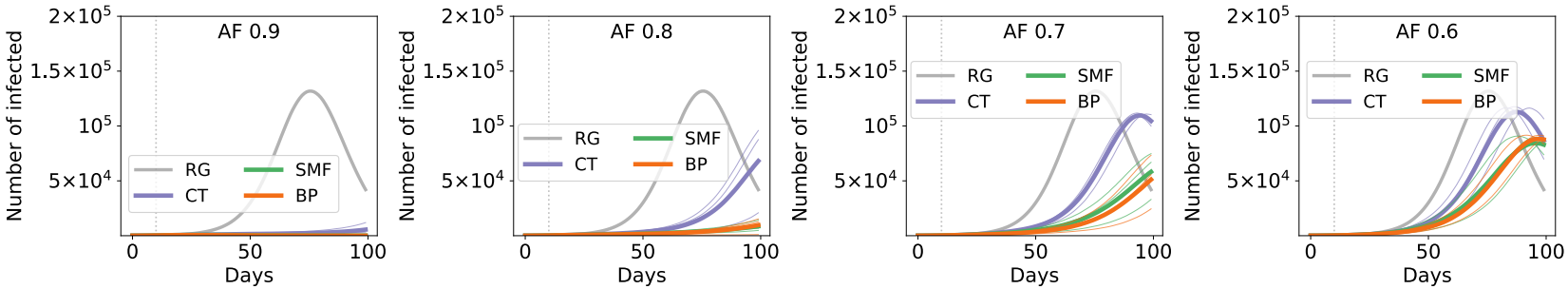


500.000 individuals, exact medical tests and app adoption fraction of 100 %

Results



500.000 individuals, 5000 observations, inaccurate medical tests and app adoption fraction of 100%



500.000 individuals, 5000 observations, accurate medical tests and app adoption fraction < 100%

Conclusions

We have seen:

- How to use methods borrowed from statistical mechanics (Simple Mean Field and Belief Propagation) to reconstruct epidemic trajectories and quantify individual risk to be infected;
- A trace/test/confine strategy based on risk assessment and on digital contact tracing is very effective in mitigating realistic synthetic outbreaks

Some future development:

- Introduce the learning of the unknown epidemic parameters through the minimization of the Bethe free energy;
- Test the strategy on other COVID-19 models (consider new settings, i.e. include vaccinated individuals) and perhaps in real apps.

Conclusions

We have seen:

- How to use methods borrowed from statistical mechanics (Simple Mean Field and Belief Propagation) to reconstruct epidemic trajectories and quantify individual risk to be infected;
- A trace/test/confine strategy based on risk assessment and on digital contact tracing is very effective in mitigating realistic synthetic outbreaks

Some future developments:

- Introduce the learning of the unknown epidemic parameters through the minimization of the Bethe free energy;
- Test the strategy on other COVID-19 models (consider new settings, i.e. include vaccinated individuals) and perhaps in real apps.

Thank you for your attention!