Probabilistic inference framework for estimating risk of infection from proximity tracking data

Antoine Baker, Florent Krzakala, Marc Mézard, Stefano Sarao,
Maria Refinetti, Lenka Zdeborová, ...

17th April 2020

It became beyond any doubts that strategies to mitigate impact of pandemic are a key priority for current technological development. Wide-spread use of smart mobile phones enables automated, real time, and privacy-preserving proximity tracking between individuals. Distance and duration of a contact between two individuals can be measured as well as contextual information such as an individual wearing a mask or not. Every individual phone can store the history of contacts and is able to communicate with them in a privacy-preserving manner via a server. Availability of such data allows for simplified tracking and notification of individual that have been in close contact with an individual tested positive. But it allows for much more than just that.

Modern probabilistic inference techniques are capable to estimate the probability that any given individual is at high risk of being infected. They do it using the measured history of contacts, combine it with established epidemic spreading models and small amount of communication between the phone trough a server. Such estimation is able to account for increased risk due to presence of syndromes associated to the illness, as well as for additional information from individual who have been tested (whether positive or negative). Probabilistic inference is able to concatenate all such available information in a way that can lead to much more efficient epidemic mitigation strategies. For instance, in terms on confinement strategies, smart probabilistic inference can be able to mitigate the propagation of the epidemic by the same amount by confining only a small fraction of the population, compared to the whole non-essential workforce. Thus being able to mitigate an epidemic while preserving the economy and societal well-being.

If our project we develop such a probabilistic inference framework for estimating risk of infection from proximity tracking data. Our strategy is readily implementable on mobile phone proximity tracking applications.

1 The Model

We study an SIR model with N individuals. An individual i is in either one of the three states: Susceptible, Infected, Recovered$\equiv \{S, I, R\}$.
The dynamics of the infection process is Markovian in nature. When going from time t to time $t+1$, the following events can take place:

- If $q_i(t) = R$: After having contacted the disease a person either dies or recovers and develops an immunity to it: $q_i(t+1) = R$.

- If $q_i(t) = I$: we call μ_i the probability that an infected individual i recovers:
$\begin{cases} q_i(t+1) & = I \quad \text{with probability } 1-\mu_i \\ q_i(t+1) & = R \quad \text{with probability } \mu_i \end{cases}$

- If $q_i(t) = S$: One looks at all the individuals j which are infected and have been in contact with i at time t. Define this set of individuals as $\partial i(t)$. Each individual j in $\partial i(t)$ infects i with a probability $\lambda_{j\to i}(t)$. More precisely:
$\begin{cases} q_i(t+1) & = I \quad \text{with probability } 1-\prod_{j\in\partial i(t)}\left(1-\lambda_{j\to i}(t)\right) \\ q_i(t+1) & = S \quad \text{with probability } \prod_{j\in\partial i(t)}\left(1-\lambda_{j\to i}(t)\right) \end{cases}$

To resume, the parameters of the model are:

- μ_i: The probability of recovery of an infectious patient. This should be considered individual-dependent since an i-dependent μ_i can allow to have different categories of population, with different probabilities to recover (or pass away).

- $\lambda_{i\to j}(t)$: The transmission probability given there was a contact between an infected i and susceptible j at time t. This should be kept time dependent as the transmission probability changes with contact time. It should also be kept individual dependent since the transmission probability changes if some sanitary measures, such as wearing masks, are taken. The app would record all the contacts a given person had, including the duration and the distance of this contact thereby leading to estimate of $\{\lambda_{i\to j}(t)\}$.

2 A Mean Field Approach

2.1 The dynamical message passing

Here we present the dynamical message passing equations from [1]. These equations are derived from the full dynamical process where variables are the full trajectories and are exact in the limit of large random graphs [2]. In tests they worked well on fixed graphs taken from applications. We are validating them now on the dynamical contacts coming from the modelling of the epidemic spread.

We define an auxiliary dynamics D_j where node j receives infection signals, but ignores them and thus is fixed to the S state at all times. Note that D_j is identical to the original dynamics D for all times for which $q_j(t) = S$. Then define the message $\theta^{k\to i}(t)$ as the probability that the infection signal has not been passed from person k to i up to time t in the dynamics D_i:

$$\theta^{k\to i}(t) = \text{Prob}^{D_i}\left(\sum_{t'=0}^{t} d^{k\to i}\left(t'\right) = 0\right). \tag{1}$$

Here the infection signal $d^{i\to j}(t)$ is defined as a random variable which is equal to one with probability $\delta_{q_i(t-1),I}\lambda_{i\to j}(t)$, and equal to zero otherwise.

Define the estimated probability of individual i to be in either three of the cases at time t as $P_R^i(t)$, $P_S^i(t)$ and $P_I^i(t)$.
Our starting point will be equations for the probabilities $P_S^i(t), P_R^i(t), P_I^i(t)$ that site i is in state S, R, I at time t are:

$$\begin{aligned}
P_S^i(t+1) &= P_S^i(0) \prod_{k\in\partial i} \theta^{k\to i}(t+1) = P_S^i(t) \prod_{k\in\partial i(t)} \frac{\theta^{k\to i}(t+1)}{\theta^{k\to i}(t)}, \\
P_R^i(t+1) &= P_R^i(t) + \mu_i P_I^i(t), \\
P_I^i(t+1) &= 1 - P_S^i(t+1) - P_R^i(t+1),
\end{aligned} \tag{2}$$

where $\theta^{k\to i}(t)$ is the probability that individual k *has not* passed the virus to i up to time t, and ∂i are here all the contacts i had up to time t, and $\partial i(t)$ contacts at time t. Note that $\theta^{k\to i}(t) \neq 1$ only if k, i have already been in contact in the past, before time t.

The dynamical message passing equations from [1] for a pair i, j that has been in contact at time t read:

$$\theta^{j\to i}(t+1) = \theta^{j\to i}(t) - \lambda_{j\to i}(t)\phi^{j\to i}(t) \tag{3}$$

$$P_S^{j\to i}(t+1) = P_S^j(0) \prod_{k\in\partial j\setminus i} \theta^{k\to j}(t+1) = P_S^{j\to i}(t) \prod_{k\in\partial j(t)\setminus i} \frac{\theta^{k\to j}(t+1)}{\theta^{k\to j}(t)} \tag{4}$$

$$\phi^{j\to i}(t+1) = \phi^{j\to i}(t)\left[(1-\lambda_{j\to i}(t))(1-\mu_j)\right] - \left[P_S^{j\to i}(t+1) - P_S^{j\to i}(t)\right] \tag{5}$$

Here $\phi^{j\to i}(t)$ is the probability that the infection signal has not been passed from node j to node i up to time t in the dynamics D_i and that node j is in the state I at time t. We stress that here ∂j means contacts j had up to time t and $\partial j(t)$ contacts j has at time t.

The initial conditions, starting from a configurations where some sites are I, others are S, is

$$\theta^{j\to i}(0) = 1 \tag{6}$$

$$\phi^{j\to i}(0) = 0 \text{ if } q_j(0) = S \tag{7}$$

$$= 1 \text{ if } q_j(0) = I\,. \tag{8}$$

For general initial condition we need

$$\theta^{j\to i}(0) = 1 \tag{9}$$

$$\phi^{j\to i}(0) = P_I^j(t=0)\ . \tag{10}$$

2.2 Mean Field Dynamical Equations

Here we present a simplification of the forward equations derived in [1]. The meanfield (MF) evolution evolution equation for $\theta^{k\to i}(t+1)$ is given by:

$$\theta^{k\to i}(t+1) = \theta^{k\to i}(t)\left[1 - P_I^k(t)\lambda_{k\to i}(t)\right]. \tag{11}$$

Furthermore, we shall work with the approximation that $\lambda_{ik}(t-1)$ is small so we can keep only temrs linear in λ and neglect the higher order terms:

$$\begin{aligned}
\prod_{k\in\partial i(t)} \theta^{k\to i}(t+1) &= \prod_{k\in\partial i(t)} \left\{\theta^{k\to i}(t)\left[1 - P_I^k(t)\lambda_{k\to i}(t)\right]\right\} \\
&\approx \left[\prod_{k\in\partial i(t)} \theta^{k\to i}(t)\right]\left(1 - \sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t)\right).
\end{aligned} \tag{12}$$

Under the assumptions of small λ_{ij} (2) gives us the final set of very intuitive equations:

$$\begin{aligned} P_S^i(t+1) &= P_S^i(t)\left(1 - \sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t)\right), \\ P_R^i(t+1) &= P_R^i(t) + \mu_i P_I^i(t), \\ P_I^i(t+1) &= P_I^i(t) + P_S^i(t)\sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t) - \mu_i P_I^i(t). \end{aligned} \tag{13}$$

Note that the last equation was obtained by writing $P_I^i(t+1) - P_I^i(t) = -\left[P_S^i(t+1) - P_S^i(t)\right] - \mu_i P_I^i(t)$ and making use of the first of the three equations.

Equations (13) determines a Markovian evolution of $\{P_S^i(t), P_R^i(t), P_I^i(t)\}_{i=1,\dots,N}$. Given an initial condition at time t_0 of $\{P_S^i(t_0), P_R^i(t_0), P_I^i(t_0)\}_{\{i=1,\dots,N\}}$, in order to obtain $\{P_S^i(t), P_R^i(t), P_I^i(t)\}$, one needs to know all the individuals i has been in contact with in the period $[t_0, t[$ (i.e. one must know $\partial i(\tau)$ for all $\tau \in [t_0, t[$). Moreover for each for each time $\tau \in [t_0, t[$, and for each $j \in \partial i(\tau)$, one must know the estimated probability of j being infected at that time: $P_I^j(\tau)$.

2.3 Law of mass action

A more rough approximation, very much used in epidemiology, is given by the global SIR equations which are the law of mass action. They involve three numbers, f_S, f_I, f_R which are the fractions of the global population which are in the three states. Assuming that the typical number of contacts at time t is $c(t)$, one writes these equations as

$$f_S(t+1) = f_S(t)\left[1 - f_I(t)\lambda(t)c(t)\right] \tag{14}$$

$$f_R(t+1) = f_R(t) + \mu f_I(t) \tag{15}$$

$$f_I(t+1) = f_I(t) + f_S(t)f_I(t)\lambda(t)c(t) - \mu f_I(t) \tag{16}$$

where $\lambda(t)$ and $\mu(t)$ are some averages of the parameters $\lambda_{i\to j}(t)$ and $\mu_i(t)$.

If λ and μ are time independent, and the population is large, at early times one has $f_S \simeq 1$, and the fraction of infected people grows like

$$f_I(t) \simeq_I (0)[1 + \lambda c - \mu]^t \tag{17}$$

This allows to estimate the "basic reproduction ratio" R_0 of our model, defined as the expected number of secondary cases produced, in a completely S population, by a "typical" infected individual during his entire period of infectiousness. A typical individual is infectious for a typical time $1/|\log(1-\mu)|$, therefore

$$R_0 = \exp\left[\log(1 + \lambda c - \mu)/|\log(1-\mu)|\right] \tag{18}$$

When λ and μ are small, one gets

$$R_0 = \exp\left[(\lambda c - \mu)/\mu\right] \tag{19}$$

3 Using symptoms and tests to update the equations

So far we have described the estimation of probabilities for each individual based on an initial condition and probabilistic infections based on the contacts. Now we describe how to take into account and propagate back in time the information about tests and syndroms.

Suppose at time t^* an individual i is tested or presents illness associated syndromes. Concerning tests, ideally, this is a test of PCR + serology, then after one test, the state of i would be known:

$q_i^*(t) \in \{S, I, R\}$ and $P_q^i(t^*) = \delta_{q,q_i^*(t)}$. We include R since serology will tell us whether somebody is immune to the virus indicating that at some point in the past this person was infectious and now cannot transmit. In case of syndromes at time t^* the probability $P_q^i(t^*)$ is update based on external medical data, specifically the probability to have illness of interest among all people presenting the same set of syndromes.

The information about tests and syndromes now need to propagate back in time and be used to update the risk levels of the contacts of person i in recent times. There are two types of information that we need to propagate back in time:

- (a) Individual i could have infected susceptible individuals that were in contact with him in the period between his infection and the time t^* of his test
- (b) Individual i must have been infected by one of his contacts.

We treat these two types separately in the present approach.

3.1 (a) Infecting others

Based on external medical data/models, the probabilities $P_q^i(t)$ are updated for all $t^* - \tau \geq t \geq t^*$ where τ is the medically estimated time from start of being infectious to the typical time of test/syndromes. Based on the work of Di Domenico et al., we can estimate τ as the sum of the prodromic time (1.5 days, and the typical time between the appearance of the first syndromes and the test, say 1.5 to2.5 days. This gives an estimate of $\tau = 3 - 4$ days.

Subsequently we update the probabilities of the individuals i had contact with (their contacts, and the contacts of their contacts, etc.) in the period of time $t \in (t^* - \tau, t^*)$ by re-running the forward update equations (13) (or (3)-(5)) using the new $P_q^i(t)$.

3.2 (b) Being infected by somebody

If an individual i was tested at time t^* then he/she must have been in contact with somebody infected in the past. We estimate from medical data and procedures that this infection must have happened in time $(t^* - \tau_1, t^* - \tau_2)$ (where $\tau_1 > \tau_2 > \tau$). We then consider a set $V(i)$ of all the individuals j that i have met in that period and adjust their probabilities of infection to

$$P_{I,new}^j(t) = \frac{P_I^j(t)}{\sum_{t' \in (t^* - \tau_1, t^* - \tau_2)} \sum_{k \in V(i)} P_I^k(t') \lambda_{k \to i}(t')} \tag{20}$$

where $t \in (t^* - \tau_1, t^* - \tau_2)$. Where in the spirit of the mean field equations we neglected terms quadratic in λ and we also neglect the possibility of recovery in the period $t \in (t^* - \tau_1, t^* - \tau_2)$. This is an approximation and needs to be validated by tests.

3.2.1 Associated inference problem

We can formulate an inference problem as follows. We start from a rough, high-level formulation. Imagine that individual number 0 is tested "I" at time T. You know his contacts at previous times, and for each of them you have a prior estimation that they are infected at the time of contacts. This "prior" estimation can be the one estimated by message passing, or in a simpler approach it could be just given by the average prevalence of "I" in the population. Given the new information that individual number 0 is tested "I" at time T, among his contacts at times before T, what is the optimal subset that should be tested (assuming that we have not enough testing facilities to test all of them?

We can formulate the full inference question more precisely as follows.

A crucial input is the probability that an individual who has been infected at time t_I is still infected at time $t_I + \tau$. We denote this function by $\psi_I(\tau)$. For instance in the Markovian setup that we use in these notes we have $\psi_I(\tau) = (1 - \mu)^\tau$, but we could use other functions.

We look at individual "0" at time T. The history of his contacts -say going back in time up to a time t_0 such that $\psi_I(T-\tau) > \varepsilon$ (we could take for instance $\varepsilon = .1$ in order to have a cutoff)- is as follows: he has been in contact with a number of individuals $a, b, c, ...$ and each of these he has met at times τ_r^a, τ_r^b, τ_r^c etc... We suppose that each individual a had a prior probability ρ_r^a of being infected at time τ_r^a, and that the probability of transmission from a to "0" during the contact at time τ_r^a was λ_r^a. For instance if we have no exchange of individual information we could take for ρ_r^a the estimated fraction of I in the full population at time τ_r^a

Now individual "0" is tested I at time T. What is the probability that each of the individuals $a, b, ...$ with whom he has been in contact is I at time T?

We need to solve actually two inference problems.

1. Knowing that "0" is tested I at time T, what is the probability that he was infected by a at time τ_r^a?

2. Assuming that "0" has been infected at time τ_r^a, what is the probability that he infected other contacts b at times $\tau_s^b > \tau_r^a$?

We shall first address the first question and for this we proceed by steps.

Another important function is the probability that, if an individual has been detected "I" at time τ, he is still "I" at time $T > \tau$. Let us denote this by $\hat{P}(T-\tau)$. Denoting by t_I the infection time, we can write:

$$\hat{P}(T-\tau) = P(q(T) = I | q(\tau) = I) = \int dt_I P(q(T) = I | t_I, q(\tau) = I) P(t_I | q(\tau) = I) \tag{21}$$

and one gets after a short computation

$$\hat{P}(T-\tau) = \frac{\int_{T-\tau}^{\infty} \rho_I(x) dx}{\left[\int_0^{\infty} \rho_I(x) dx\right]} \tag{22}$$

In the special case of Markovian recovery, one gets

$$\hat{P}(T-\tau) = (1-\mu)^{T-\tau} = \psi_I(T-\tau) \tag{23}$$

We turn to the next exercise. Imagine that two individuals, a and b, have been in contact with individual "0" at times τ^a and τ^b respectively. We have a prior information saying that the probability that a is I at time τ^a is ρ^a, and similarly for b. Now at a time $T > \tau^a, \tau^b$, individual "0" is tested I. What are the posterior probabilities that each of the two contacts a and b have been I and have infected "0"? The probability that "0" is detected I, in the case where $q^a(\tau^a) = I$ and $q^b(\tau^b) \neq I$, is $\lambda^a \hat{P}(T-\tau^a)$, where λ^a is the probability of transmission from a to 0 at time τ^a. Similarly, the probability that "0" is detected I, in the case where $q^a(\tau^a) \neq I$ and $q^b(\tau^b) = I$, is $\lambda^b \hat{P}(T-\tau^b)$. What happens when $q^a(\tau^a) = I$ and $q^b(\tau^b) = I$?

Assume that $\tau_a < \tau_b$. If the virus is transmitted by a at time τ_a, then what happens at τ_b is irrelevant. Therefore, the probability that "0" is detected I, in the case where $q^a(\tau^a) = I$ and $q^b(\tau^b) = I$, is $\lambda^a \hat{P}(T-\tau^a) + (1 - \lambda^a \hat{P}(T-\tau^a))\lambda^b \hat{P}(T-\tau^b)$. Let us call $\alpha^a = \lambda^a \ \hat{P}(T-\tau^a)$, and same for b. We have:

$$\begin{aligned}
P(\text{"0" is I at time } T) =& P(\text{"0" is I at time } T | q^a(\tau^a) = I \text{ and } q^b(\tau^b) \neq I)\rho_a(1-\rho_b) && (24)\\
&+P(\text{"0" is I at time } T | q^a(\tau^a) \neq I \text{ and } q^b(\tau^b) = I)(1-\rho_a)\rho_b && (25)\\
&+P(\text{"0" is I at time } T | q^a(\tau^a) = I \text{ and } q^b(\tau^b) = I)\rho_a\rho_b && (26)\\
=& \alpha^a\rho_a(1-\rho_b) + \alpha^b\rho_b(1-\rho_a) + [\alpha^a + \alpha^b - \alpha^a\alpha^b]\rho_a\rho_b && (27)
\end{aligned}$$

The probability $P(\text{"0" is I at time } T | q^a(\tau^a) = I \text{ and } q^b(\tau^b) \neq I)$is equal to α^a. The probability $P(\text{"0" is I at time } T | q^a(\tau^a) \neq I \text{ and } q^b(\tau^b) = I)$ is equal to α^b. As for the probability $P(\text{"0" is I at time } T | q^a(\tau^a) = I \text{ and } q^b(\tau^b) = I)$, we must be more careful, and consider the ordering of times. Suppose that $\tau_a < \tau_b$. We have three events of infection of "0" to consider.

- a infects "0", b does not. Probability $\alpha^a(1-\alpha^b)$
- b infects "0", a does not. Probability $\alpha^b(1-\alpha^a)$
- a and b infect "0". In this case the responsible for the infection is a who transmitted it first. Probability $\alpha^a\alpha^b$

From this, using Bayes' rule, we can answer the question.

The probability γ_a that a was I at time τ_a and has transmitted the infection to "0" at τ^a, given that "0" has been found I at time T is

$$\gamma_a = \frac{\alpha^a\rho_a}{\alpha^a\rho_a + \alpha^b\rho_b - \rho_a\rho_b\alpha_a\alpha_b} \tag{28}$$

The probability γ_b that b was I at time τ_b and has transmitted the infection to "0" at τ^b, given that "0" has been found I at time T is

$$\gamma_b = \frac{\alpha^b\rho_b(1-\alpha^a\rho_a)}{\alpha^a\rho_a + \alpha^b\rho_b - \rho_a\rho_b\alpha_a\alpha_b} \tag{29}$$

Note that $\gamma_a + \gamma_b = 1$ as it should.

We should now evaluate the probabilities that a and b were infected at times respectively τ_a and τ_b, given that "0" has been found I at time T. These are correlated events : the probability of both being not infected is zero. We call respectively $(1,1)$, $(1,0)$, $(0,1)$,$(0,0)$, the events where both, only a, only b and none of them was infected and we define the probability of these events as $p_{ab}(.,.)$. The easiest approach is to condition on the event of infection transmission.

If the infection was transmitted by a at time τ_a (probability γ_a), we have clearly

$$p_{ab}(0,0) = 0 \tag{30}$$
$$p_{ab}(1,0) = (1-\rho_b) \tag{31}$$
$$p_{ab}(0,1) = 0 \tag{32}$$
$$p_{ab}(1,1) = \rho_b \tag{33}$$

If the infection was transmitted by b at time τ_b (probability γ_b), we have probably (**to be checked**):

$$p_{ab}(0,0) = 0 \tag{34}$$
$$p_{ab}(1,0) = 0 \tag{35}$$
$$p_{ab}(0,1) = 1-\rho_a(1-\alpha^a) \tag{36}$$
$$p_{ab}(1,1) = \rho_a(1-\alpha^a) \tag{37}$$

In total, we have

$$p_{ab}(0,0) = 0 \tag{38}$$
$$p_{ab}(1,0) = (1-\rho_b)\gamma_a \tag{39}$$
$$p_{ab}(0,1) = [1-\rho_a(1-\alpha^a)]\gamma_b \tag{40}$$
$$p_{ab}(1,1) = \rho_b\gamma_a + \rho_a(1-\alpha^a)\gamma_b \tag{41}$$

We can now solve the full inference problem with two persons having each one contact. We first neglect the case where "0" has infected one of them. We define q_a and q_b the states of a and b ($q_a = 1$ for a infected, 0 otherwise) at τ_a and τ_b. The probabilities $p_{ab}(q_a, q_b)$ are given in the table above. Then we define f_a, f_b as the states of a, b at time T. we can compute the joint probability $P(q_a, q_b; f_a, f_b)$. Clearly:

$$P(q_a, q_b; f_a, f_b) = p_{ab}(q_a, q_b)g(q_a, f_a, \tau_a)g(q_b, f_b, \tau_b) \tag{42}$$

where

$$g(q,f,\tau) = q\left[\hat{P}(T-\tau)f + (1-\hat{P}(T-\tau)(1-f)\right] + (1-q)(1-f) \tag{43}$$

We get finally the final probabilities of the states f_a and f_b at time T using $P^T(f_a,f_b) = \sum_{q_a,q_b} P(q_a,q_b;f_a,f_b)$:

$$\begin{aligned}
P^T(0,0) =& (1-\rho_b)\gamma_a(1-\hat{u}_a) + [1-\rho_a(1-\alpha a)]\gamma_b(1-\hat{u}_b) && (44)\\
&+ [\rho_b\gamma_a + \rho_a(1-\alpha^a)\gamma_b](1-\hat{u}_a)(1-\hat{u}_b) && (45)\\
P^T(1,0) =& (1-\rho_b)\gamma_a(1-\hat{u}_a) + [\rho_b\gamma_a + \rho_a(1-\alpha^a)\gamma_b]\hat{u}_a(1-\hat{u}_b) && (46)\\
P^T(0,1) =& [(1-\rho_a(1-\gamma_a)]\gamma_b\hat{u}_b + [\rho_b\gamma_a + \rho_a(1-\alpha^a)\gamma_b](1-\hat{u}_a)\hat{u}_b && (47)\\
P^T(1,1) =& [\rho_b\gamma_a + \rho_a(1-\alpha^a)\gamma_b]\hat{u}_a\hat{u}_b && (48)
\end{aligned}$$

where $\hat{u}_a = \hat{P}(T-\tau_a)$ and $\hat{u}_b = \hat{P}(T-\tau_b)$

Note that the small ρ limit of these expressions is simple and intuitive:

$$\begin{aligned}
P^T(0,0) =& 1-\gamma_a\hat{u}_a - \gamma_b\hat{u}_b && (49)\\
P^T(1,0) =& \gamma_a\hat{u}_a && (50)\\
P^T(0,1) =& \gamma_b\hat{u}_b && (51)\\
P^T(1,1) =& 0 && (52)
\end{aligned}$$

To this we should add the probability that "0" has infected one of the neighbours. It can only be b, and it is conditioned to a having infected "0" at time $\tau_a < \tau_b$. The probability that b is infected at time T due to this process is $\gamma_a\lambda_0\hat{u}_b$

In the end, in the small ρ limit, conditioned to "0" having been detected I at time T, the probability that a is infected at time T is $\gamma_a\hat{u}_a$ and the probability that b is infected at time T is $\gamma_b\hat{u}_b + \gamma_a\lambda_0\hat{u}_b$.

The next exercise is with two persons, a and b, who have had multiple contacts with "0" at previous time. We call n^a and n^b the number of these contacts, and τ_r^a the times of contacts with a, with $r \in \{1,\dots,n^a\}$, and we call τ_r^b the times of contacts with b, with $r \in \{1,\dots,n^b\}$. Denoting

$$\beta_r^a = \lambda_r^a\hat{P}(T-\tau_r^a)\ , \tag{53}$$

we see that the previous equations apply if we substitute

$$\alpha^a = 1 - \prod_{r=1}^{n_a}(1-\beta_r^a) \tag{54}$$

Well, probably, not quite so. This should be fixed first

Finally we can solve the full first inference question We look at individual "0" at time T. Knowing that individual "0" is tested I at time T. What is the probability that each of the individuals $a, b, ...$ with whom he has been in contact has transmitted the infection to "0" and at what time?

TBD

4 Additional remarks

There equations are not loopy Belief Propagation, thus existing implementation of there off will not be on much use. At the same time it takes just a couple of lines of code to implement them into any system that stored the needed data.

We are working on a code to test this approach. This code should be easy to feed with the parameters $\mu, \lambda_{ij}(t), \partial i(t), P_q^i(t=0)$ as well as syndromes and test results with times at which they were present. And comparison with ground truth coming from generative modelling will be output.

What is presented above is the simplest set of equations achieving the desired goal on a measured dynamically changing network of contacts. Two directions need to be explores - going beyond linear expansion in λ, and treating the inference due to tests and syndromes in a way joint to the forward propagation. Both these are an interesting research program, but will lead to more costly equations and hence trade-offs between precision and computational cost will need to be carefully evaluated.

The algorithm can be implemented fully locally, for privacy preserving considerations.

Having a grasp of $P_I^i(t)$ opens the way to many interesting applications. Possible outcomes of the procedure include:

- Whenever the probability of being infected of an individual i is beyond a threshold β: $P_I^i(t) > \beta$, then i gets a recommendation to be tested or quaranteened.

- If an individual i calls emergencies, its value of $P_I^i(t)$ could be transmitted thus helping doctors in triage procedure.

5 Pseudo-Code

class InferenceModel()

Inference model is a class that represents a population of N individuals, each having probability P_S^i to be in either one of the three state $\{S, I, R\}$. It contains the method time_evolution which simulates the time evolution of the probabilities as dictated by equations (13).

Initialisation

An inference model is initialized to a population size, positions of the individuals which affects the contact between them as well as initial probabilities to be in the three different states.

- N: Number of individuals

- (x_pos,y_pos): spatial position of all the individuals

- initial_probas: for each of the N individuals defines the probability to be in either states (Sane=0,Infected=1,Recovered=2)

time_evolution_no_feedback

Time_evolution is the main method of the class. It makes the probabilities of each individual to be in a given state evolve according to (13).

- parameters:
  - recover_probas: μ_i: the probability of an individual to recover or die

  - transmissions: a list containing all the contacts an individual had over the period of time T. For all individuals and all times it contains the transition rates: $\lambda_{i\to j}(t)$ for an infected i to pass the infection to a susceptible j.

- for $t < T$:
  - Infection_Probability: for each individual what is the probability that it was infected at time t given the contacts it had and the probability of all the persons he had contacts with to be infected.
  - Time step: Implements a time step according to (13).

$$\begin{aligned} P_S^i(t+1) &= P_S^i(t)\left(1 - \sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t)\right) \\ P_R^i(t+1) &= P_R^i(t) + \mu_i P_I^i(t) \\ P_I^i(t+1) &= P_I^i(t) + P_S^i(t)\sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t) - \mu_i P_I^i(t) \end{aligned} \tag{55}$$

time_evolution_with_feedback

A crucial contribution of this works considers a modification of the time_evolution method in which, at each time t, probabilities of the individuals at all times are modified by using results from tests as well as symptoms at t.

- parameters:
  - recover_probas: μ_i: the probability of an individual to recover or die
  - transmissions: a list containing all the contacts an individual had over the period of time T. For all individuals and all times it contains the transition rates: $\lambda_{i\to j}(t)$
  - Δt : the typical number of days between the infection and the appearance of symptoms or testing.
  - M: The number of individuals that are tested each time step.
- $t < T$:
  - M individuals in the populations are tested. The test result is either $\{S, I, R\}$. Let $r_j(t)$ be the result of the test for individual j at time t.
  - Update the probabilities of the tested individuals according to:

$$\begin{cases} \text{if } r_j(t) = S & P_S^j(t') = 1 \text{ for all } t' \le t \\ \text{if } r_j(t) = I & P_I^j(t') = 1 \text{ for all } t - \Delta t \le t' \le t \\ \text{if } r_j(t) = R & P_R^j(t') = 1 \text{ for all } t' \ge t \end{cases}$$

  - Evolve the system from $t - \Delta t$ to t using the updated probabilities. This gives at time t that take into account the test results.
  - Time step: Perform the time step according to (13) with the updated probabilities.

$$\begin{aligned} P_S^i(t+1) &= P_S^i(t)\left(1 - \sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t)\right) \\ P_R^i(t+1) &= P_R^i(t) + \mu_i P_I^i(t) \\ P_I^i(t+1) &= P_I^i(t) + P_S^i(t)\sum_{j\in\partial i(t)} P_I^j(t)\lambda_{j\to i}(t) - \mu_i P_I^i(t) \end{aligned} \tag{56}$$

Possible modifications of this method include considering removing the individuals that have been tested positive from the system in order to account for hospitalisation or quarantine. One could also take into account the syndromes of the individuals and modify the probabilities accordingly.

class EpidemicModel()

Epidemic Model is a class that simulates the propagation of the virus in a population. It is contains the "true" state of all individuals at all times agains which we will compare our inference model. EpidemicModel further stores the list of all contacts that occured in the time of simulation.
Here we consider a propagation scenario given by a ProximityModel class.

Initialisation

An epidemic model is initialized to a population size, positions of the individuals which will affect the contact between them as well as initial probabilities to be in the three different states.

- N: Number of individuals
- (x_pos,y_pos): spatial position of all the individuals
- initial_states: the initial states of all the individuals in the population

time_evolution

Time_evolution is the main method of the class. It will evolve the state of all individuals in the population until time T.

- parameters:
  - recover_probas: μ_i: the probability of an individual to recover or die
  - transmissions: a list containing all the contacts an individual had over the period of time T for all individuals and their associated transmission rates: $\lambda_{i\to j}(t)$.
- for $t < T$:
  - Infection_Probability: for each individual what is the probability that it was infected at time t given the contacts it had and the probability of all the persons he had contacts with to be infected.
  - Time step: Updates the state of each individual in the community. The infected individuals recover with a probability μ_i and healthy individuals get infected with a probability $\sum_j \lambda_{j\to i}(t)$.

class ProximityModel()

Proximity Model is a subclass of EpidemicModel. It implements an actual Epidemic model by defning the initial states, the recory probability μ_i as well as the transmission rates $\lambda_{j\to i}$. Here we choose a model in which the probability of having transmission given a contact decays with the distance at which the contact occur. The decay rate is given by the parameter *scale*.

Initialisation

An ProximityModel model is initialized to a population size, a probability to recover for each individual μ_i as well as a transmission rate $\tilde{\lambda}$ which is here chosen not to depend on the individuals.

- N: Number of individuals

- lamb: the transmission rate given there was contact between two individuals. In this model it is contant amoung the individuals $\tilde{\lambda}$.
- mu: the recovery rate μ_i
- scale: defines the rate at which the contact probability decays with distance.
- (x_pos,y_pos) *(optional)*: initial spatial positions of all the individuals in the population
- proba_contact *(optional)*: probability of having a contact between two individuals at a given distance. If not specified is computed as $p^{ij}_{\text{contact}}(t) = \mathrm{e}^{\frac{-\text{dist}(i,j)(t)}{\text{scale}}}$

sample_transmissions()

Sample_transmissions computes the values of $\lambda_{i\to j}(t)$ for every contact that occurred in the population and stores them. These values are then used in order to propagate the inference model as well as the Epidemic model. Here $\lambda_{i\to j}(t)$ is computed as:

$$\lambda_{i\to j}(t) = \begin{cases} \tilde{\lambda} & \text{with probability } p^{ij}_{\text{contact}}(t) \\ 0 & \text{with probability} 1 - p^{ij}_{\text{contact}}(t) \end{cases}$$

References

1. Lokhov, A. Y., Mézard, M., Ohta, H. & Zdeborová, L. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E* **90.** `http://dx.doi.org/10.1103/PhysRevE.90.012801` (July 2014).
2. Lokhov, A. Y., Mézard, M. & Zdeborová, L. Dynamic message-passing equations for models with unidirectional dynamics. *Physical Review E* **91,** 012811 (2015).