Statistical physics and epidemic inference: methods and applications

Original
Statistical physics and epidemic inference: methods and applications / Mazza, Fabio. - (2023 Jul 17), pp. 1-153.

Publisher:
Politecnico di Torino

Published
DOI:

Politecnico
di Torino
1859

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Physics (35th cycle)

Statistical physics and epidemic inference: methods and applications

Fabio Mazza

* * * * *

Supervisor
Prof. Luca Dall'Asta

Doctoral Examination Committee:
E. Valdano, *Referee*, Chargé de Recherche, INSERM, Paris
A. Lage Castellanos, *Referee*, Professor, University of Havana
A. M. Chiarello, *External member*, Professor, University of Napoli
C. Bosia, *Internal member*, Researcher, Politecnico di Torino
A. Gamba, *Internal member*, Professor, Politecnico di Torino
L. Dall'Asta, *Internal member*, Professor, Politecnico di Torino

Politecnico di Torino
17th July 2023

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .
Fabio Mazza
Turin, 17th July 2023

Summary

Performing epidemic inference at the individual scale is a difficult task because of the complex interactions that are present. As the size of the considered system grows, its inherent complexity makes the task practically infeasible, even with the computational capabilities available today. This thesis is then devoted to the study of inference methods grounded in statistical physics, that can perform this task efficiently and provide a good approximate solution to the problem when an exact one is out of reach from the practical point of view.

The structure of the thesis follows the order in which the PhD program unfolded. In the first chapter, after briefly discussing some of the theory of network science, the models for epidemic spreading used in the thesis are introduced. In particular, the probabilistic individual-based version of the Susceptible-Infected-Recovered model is defined and analysed, and the epidemic inference problem is introduced on this particular model.

Then, the application of Autoregressive Neural Networks (ANNs) to this problem is discussed. This technique involves training artificial neural networks in order to generate epidemic cascades that are compatible with the observations made on a single outbreak. This approach is applied to three problems of epidemic inference, the patient zero detection problem, the epidemic risk evaluation and the inference of epidemic parameters.

The following chapters deal with one of the challenges put forward by the COVID-19 pandemic. During the first and second epidemic waves of the disease (both in Europe and the world), digital contact tracing has been developed as a mechanism to track the transmission of the disease and isolate infectious individuals. However, in order to perform these task efficiently, the evaluation of the epidemic risk of every individual in the population is required. A large part of the work of the PhD, then, has been dedicated to the application of epidemic inference techniques derived from statistical physics, in order to show how they can improve the detection of infected individuals during an epidemic outbreak.

In the third chapter, the methods are first applied to one particular agent-based model for COVID-19. The focus of this investigation is on the containment capability of the methods, that is their ability to identify the infectious individuals, who are then isolated preventing new infections from occurring. In order to make the containment experiments more realistic, several features are introduced, such as the isolation of households, the test error rate and the impartial collection of the individual contact network. The methods derived from statistical physics are compared here to contact tracing, along with a control method.

In the next chapter, the analysis is expanded to other two agent-based models for COVID-19. The containment performance is investigated in these two other models, with the same features introduced previously. Then, the statistical physics methods are applied to two other tasks, the identification of super-spreaders (the individuals that are able to infect many others) and to

the reconstruction of the propagations from the observations, both forward and backward in epidemic time.

Finally, another method for epidemic inference is described, based on a small coupling approximation of the dynamic cavity approach to Susceptible-Infected epidemic process. After its derivation, the method is tested for the prediction of the dynamics without observations, and in the risk inference setting, in both synthetic and real contact graphs.

*To my family and my
friends, without whom this
thesis would not have seen
the light of day*

Contents

Chapter 1

Introduction

1.1 A brief introduction to graph theory

A graph is a mathematical structure that describes the relationships between pairs of objects: these are called *vertices* or *nodes*, and their set is denoted by V. While in principle the indexing of nodes can be arbitrary, we will adopt the convention that each one is associated to an index i, so that $V = \{1,2,3,\dots,N\}$, where N is the total number of objects. The relationships between nodes are expressed as connections, or *edges*, which are represented as the pair of nodes that are connected, (i,j) for example. The set of all the edges of a graph is denoted by E. The graph itself is then defined with the corresponding sets of vertices and edges, $\mathcal{G} = (V,E)$.

The graph formalism is very general, and can be applied in many different areas: from representing physical connections between computers (computer networks, such as the Internet [7]), electrical systems [60], or the road network [17], to more abstract connections, like social networks [40], or in biology (gene regulation is a famous example [19]). In statistical physics graphs usually represent the local connections between individual components of a system.

A given graph can be represented in different ways: the easiest is giving a list of all vertices, and for each one a list of the other ones that it is connected to. This is called *adjacency list*. Instead, one can write it in the form of matrix A_{ij}, where

$$A_{ij} = \begin{cases} 1 & (i,j) \in E \\ 0 & \text{else} \end{cases}$$

that is called the *adjacency matrix*. In the context of this thesis, a graph is forbidden to have *self-edges*, which are edges that connect the same node, so by definition $A_{ii} = 0\ \forall i \in V$.

If the direction of the edges of a graph is not important, then the graph is called *undirected*, and the adjacency matrix is symmetric, while in the opposite case the graph is *directed*: each edge is then defined as coming *from* a node i and going *to* another node j, and is denoted as $i \to j$, or, when using simply a pair, the origin of the edge is put first, as in (i,j). Two simple examples of the two kinds of graphs are, in social network science, the Facebook friendships and Instagram followers. In the former case, friendships are bilateral, so that if individual i is friend of j, the reverse is true, giving an undirected network for the friendships. In the latter case, if profile i follows j, the reverse is not true, therefore the followers networks is directed. In the rest of the thesis, when a graph is not specified to be undirected or directed, it will be implied to be of the former kind.

The neighbours of a node i in a graph $\mathscr{G}$ are the nodes that share an edge with i; the set of the neighbours will be called, in the following, ∂i, with the formal definition

$$\partial i = \{j \in V \ : \ (i,j) \in E\} \tag{1.1}$$

The degree of a node is the number of neighbours it has, $k_i = |\partial i|$. In the case of a directed network, the set of the neighbours is split into incoming neighbours, and outgoing neighbours, based on whether the considered edge is pointing to i or out of i. The same distinction is applied to the degree, so that in the former case we have the in-degree, and in the latter the out-degree. A graph is said to be *complete* when every node in the graph is connected to every other node. Such graphs are usually indicated as K_n, where n indicates the number of vertices, as no other information is needed to characterize the graph. A graph *clique* is a subset of the nodes of the graph, $V' \subseteq V$, in which any node is connected to every other one. This means that the induced subgraph $\mathscr{G}'$, that is made by all the edges between the nodes in V', is complete.

A graph is called *bipartite* when its vertices can be divided in two distinct sets, so that the edges only connect nodes from one set to the other. In mathematical terms, we can write $E \subseteq \{(i,k) \ : \ i \in U, k \in W\}$,where $V = U \cup W$ and $U \cap W = \emptyset$.

A *walk* in a graph, is a sequence of nodes such that every consecutive pair of nodes is connected by an edge (in a directed graph, with the appropriate directionality). A *path* in a graph is a walk that never intersects itself, so that all edges and vertices traversed in a path are unique. In an undirected graph, two nodes u, v are said to be *connected* if there is a path starting from u going to v (or vice-versa). The whole graph is then said to be connected if every pair of vertices is connected, otherwise, it is disconnected. Then, *connected components* can be defined, that are the maximal connected sub-graphs of the original graph. From this definition, the connected component of a node can be found taking the sub-graph of all the nodes and edges that can be reached or traversed from that node. For directed graphs, the definition of connected components is trickier, as it involves first deciding whether to ignore the directionality of the edges (giving *weakly connected* components in doing so), and then how to take the directionality into account, giving rise to several different definitions [62].

A path in a graph where the initial and final vertices are the same, is called a *cycle*. In the thesis, the term *loop* will be used with the same meaning. If a graph is connected and does not contain any cycle, it is called a *tree*, and it has special properties: first, it can be shown that a tree of N vertices has exactly $N-1$ edges (which can be used as a definition of a tree, including connectivity). Secondly, it can be shown that in a tree graph there is only one path between any pair of nodes. This has important implications, as removing one of the edges effectively disconnects two nodes, and the graph becomes disconnected.

1.1.1 Graph metrics

It is necessary to introduce a few metrics that are used when analysing a given graph, as they will be used in the rest of the thesis. These metrics are often used to characterize a graph, and distinguish between different kinds of graphs.

The *density* of a graph is a measure of how many edges it contains. In an undirected graph, the maximum number of edges is given by the binomial coefficient $\binom{N}{2}$, where $N = |V|$ is the number of nodes. Then, the density d is defined as

$$d = \frac{|E|}{\binom{N}{2}} = \frac{2\,|E|}{N(N-1)} \tag{1.2}$$

In a directed graph, the edges are distinguished by direction, so it is possible to have double the amount of edges, and the density thus reads

$$d = \frac{|E|}{N(N-1)}.$$

The *clustering coefficient* quantifies the propensity of the nodes of a graph to cluster together. Locally, it can be defined by looking at the neighbours of each node, and counting the number of edges there are between them. In mathematical terms, if we consider a node i and its neighbourhood ∂i, the number of possible edges that can exists between the neighbours are $k_i(k_i-1)/2$, $k_i = |\partial i|$ being the number of neighbours. Then, the local clustering coefficient is

$$C_i = \frac{2\left|\left\{(u,v) \,:\, u,v \in \partial i, (u,v) \in E\right\}\right|}{k_i(k_i-1)} \tag{1.3}$$

as has been first proposed by Watts and Strogatz [11]. This formula is valid only when $k_i \geq 2$, when this is not true the local clustering is $c_i = 0$. From this definition, it is possible to give a clustering measure for the entire graph by averaging on the nodes:

$$C_{\text{avg}} = \frac{1}{N}\sum_{i \in V} C_i \tag{1.4}$$

Clustering coefficient measures have been found to be significantly important in the analysis of real world networks, as they are often quite high in social networks [62].

1.2 Models for epidemic spreading

Mathematical models of epidemic spreading are largely used to forecast the evolution of outbreaks at different spatial and temporal scales, to evaluate the effects of public-health interventions and to guide governments decisions [86, 102, 147, 22, 124, 15, 146]. The study of the spreading of diseases started more than two centuries ago: one of the earliest models was, in fact, developed by Bernoulli in 1760 for the diffusion of smallpox [1]. With the increasing availability of computational power, and motivated by the emergence of global epidemic outbreaks (H1N1, SARS, and most recently COVID-19) and other severe transmissible diseases, like AIDS or Ebola, the field of mathematical modelling of epidemic outbreaks has grown significantly [33, 42, 56, 164, 162]. In this section, the models used in the thesis will be described, along with the epidemic processes underpinning them.

1.2.1 The homogeneous SIR Model

One of the first mathematical models for the spreading of an infectious disease in a population was formulated by Kermack and McKendrick in 1927 [2]. It was originally called the *xyz* model, the *x*,*y*,*z* variables standing respectively for the number of susceptible (S), infected (I) and recovered (or removed, R) individuals in the population. This model has been a cornerstone for mathematical epidemiology, and has also been applied outside of this field (in rumour spreading [34], or cryptocurrency [58], for example).

The model divides the population into three different *compartments*: susceptible (S) individuals are naive to the disease, so they can be infected and become ill when they come in contact with it. Infectious (or infected, *I*) individuals are the one carrying the disease, and they can infect others (the susceptible ones, *S*). At the same time, they can heal from the disease, gaining life-long immunity from the disease: in this new state, they can't be infected any more, and they are in the so-called Recovered (*R*) compartment. This compartment is also called *Removed* because *R* individuals also can't infect anyone, so they do not participate anymore in the propagation process. Schematically, these transitions can be written like this:

$$\begin{aligned} S + I &\rightarrow 2I \\ I &\rightarrow R \end{aligned} \tag{1.5}$$

Defining as β the rate at which one I individual is able to infect others, and γ the rate at which he/she can recover, it is possible to write the following set of deterministic equations, which define the SIR model by describing the time evolution of the number of individuals in each compartment:

$$\frac{dS}{dt} = -\frac{\beta}{N} S(t)\, I(t) \tag{1.6a}$$

$$\frac{dI}{dt} = \frac{\beta}{N} S(t)\, I(t) - \gamma I(t) \tag{1.6b}$$

$$\frac{dR}{dt} = \gamma I(t) \tag{1.6c}$$

These equations also assume that the total number of individuals is constant over time:

$$S(t) + I(t) + R(t) = N$$

where we define N as the number of individuals.

While the time evolution depends on the initial conditions $S(0)$, $I(0)$, $R(0)$, which in principle can be set to any value (providing their sum is N), the usual study case for this model is when a new disease is introduced into a population, meaning that almost everybody is susceptible, and only a few people are I, with no Recovered individuals. With this initial condition, that implies $S(0) \simeq N$, equation (1.6b) can be approximated as

$$\frac{dI}{dt} \simeq (\beta - \gamma)\, I(0)$$

This means that at the beginning, the number of infected will grow if $\beta > \gamma$, and it will decline if the inverse is true ($\gamma > \beta$). This leads to the define a fundamental ratio, the basic reproduction number

$$R_0 = \frac{\beta}{\gamma} \tag{1.7}$$

which, in general, corresponds to the expected number of secondary infections generated by a primary infection, in a non-immunized population.

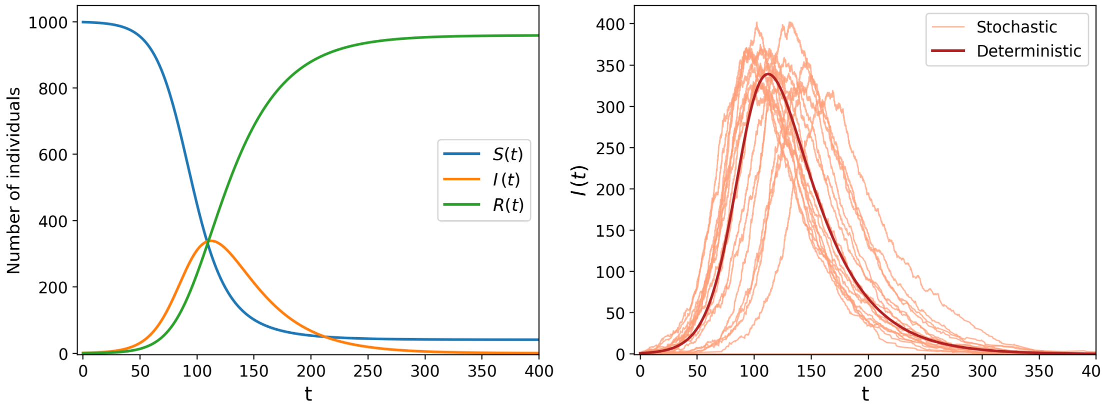

Figure 1.1: **Examples of the homogeneous SIR Model**: on the left, the solution of the SIR deterministic system of equations (1.6a) - (1.6c) over a population of size $N = 1000$. On the right, only the number of infected (I) is shown as a thick line. Also, the solution of the stochastic version of the same system is shown on the right pane, for the same parameters. In order to appreciate the effect of random chance, 20 different possible trajectories of the stochastic system are shown: in three of them, the number of infected drops to 0 shortly after the beginning of the epidemic. The parameters used are: $\beta = 0.1$, $\gamma = 0.03$, with initial values $S(0) = N - 1$, $I(0) = 1$ and $R(0) = 0$.

SI model

In the case of $\gamma = 0$, the number of recovered individuals is fixed to $R(0)$. In this case the R state can discarded altogether, and equations (1.6a)-(1.6b) become

$$\frac{dS}{dt} = -\frac{\beta}{N} S(t) I(t) \tag{1.8}$$

$$\frac{dI}{dt} = \frac{\beta}{N} S(t) I(t) \tag{1.9}$$

with the conservation equation that now reads $S(t)+I(t) = N-R(0)$. So, in this case the number of infected individual will grow, given that $I(0) > 0$, until there are no susceptible individuals left.

1.2.2 Stochastic SIR Model

The SIR model presented so far has been written as a set of deterministic dynamic equations. However, the transitions in eq 1.5 can also be interpreted in a stochastic way, meaning that the two processes, infection of a new individual, $S + I \to I + I$, or recovery, $I \to R$, can happen at random with a certain probability rate each, which will be respectively called r_{SI} and r_{IR}. In a well-mixed population, $r_{SI} = \beta/N$, and $r_{IR} = \gamma$. The difference with the previous setting is that now one process may happen before the other. Also, the number of individuals in each state is not real-valued, but a discrete integer taking values in the interval $[0, N]$.

This kind of model can be simulated using methods derived from computational chemistry. In particular, the Gillespie algorithm [4] allows to generate statically correct trajectories of the system. In Figure 1.1 the solution of the SIR deterministic equations is shown, along with simulations of the stochastic system, starting from a single infected individual in a population with size $N = 1000$. In the figure it can be seen that, while the growth of the number of infected individuals is immediate in the deterministic system, in the stochastic version this number can grow sooner or much later. This is because of the possibility of a recovery transition interrupting the growth: when the number of individuals are very few, this might even bring the epidemic to an early stop.

1.3 Probabilistic SIR Model on a contact network

The general SIR model doesn't consider explicitly every individual in a population, as it simply describes the number or fraction of them in each compartment for the dynamics, thus giving a coarse-grained vision of the epidemic spreading process. In this sense, it also works on the assumption of homogeneous mixing, meaning that any individual has the same chance of meeting any other individual at random, and this approximation is quite limiting, for example when considering a population spanning several towns, where the assumption breaks down.

In order to model the interactions between individuals, a network of contacts, or graph, $\mathcal{G} = (V, E)$ is defined [18], where the individuals are the vertices $i \in V$ and there is an edge between individuals i and j, $(i, j) \in E$, if and only if they are in contact.

Now, we can represent the state of each individual at any time: in the following, we will use the variable $x_i^t \in \{S, I, R\}$ for the state of individual i at time t, and the state of the whole population at time t will be written as $\mathbf{x}^t = \left(x_1^t, x_2^t, \dots x_N^t\right)$ (with $N = |V|$ the number of individuals).

In the following we will restrict ourselves to consider the epidemic process only at specific uniform time intervals, taking snapshots of the epidemic process in between each interval, so that t will be an index representing the number of elapsed time steps (that could correspond to, for example, the number of days elapsed since the beginning of the epidemic). We will consider $t = T$ as the last time instant, starting from $t = 0$ as the initial time.

The spreading process now under consideration is a stochastic one: defining λ as the (instantaneous) probability of infection, and μ as the probability of recovery at every time instant, the probabilistic SIR model on a contact graph follows Markov dynamics (i.e. the state of the individuals as the next time instant depends on the current states, and not the states at previous times). This is also valid when the parameters depend on the individual (μ_i) or the network edge ($\lambda_{i\to j} \doteq \lambda_{ij}$, of i infecting j). Then, it's possible to write the probability of an individual i being found in state x_i^{t+1}, given the states of everyone at the previous time t:

$$p\left(x_i^{t+1} = S \mid \mathbf{x}^t\right) = \mathbb{I}\left[x_i^t = S\right] \prod_{j\in\partial i} \left(1 - \lambda_{ji}\mathbb{I}\left[x_j^t = I\right]\right) \tag{1.10a}$$

$$p\left(x_i^{t+1} = I \mid \mathbf{x}^t\right) = \mathbb{I}\left[x_i^t = S\right] \left[1 - \prod_{j\in\partial i} \left(1 - \lambda_{ji}\mathbb{I}\left[x_j^t = I\right]\right)\right] + \mathbb{I}\left[x_i^t = I\right](1 - \mu_i) \tag{1.10b}$$

$$p\left(x_i^{t+1} = R \mid \mathbf{x}^t\right) = \mathbb{I}\left[x_i^t = R\right] + \mathbb{I}\left[x_i^t = I\right]\mu_i \tag{1.10c}$$

where the product in the first and second equation runs over the neighbours of i in $\mathcal{G}$, $\partial i = \{j \in V : (i, j) \in E\}$, and $\lambda_{ji} = \lambda_{ij}$ (written in the former way to emphasize the direction of the spreading on the network). The $\mathbb{I}$ is the indicator function, so that $\mathbb{I}[b] = 1$ only when its argument b is true, otherwise $\mathbb{I}[b] = 0$.

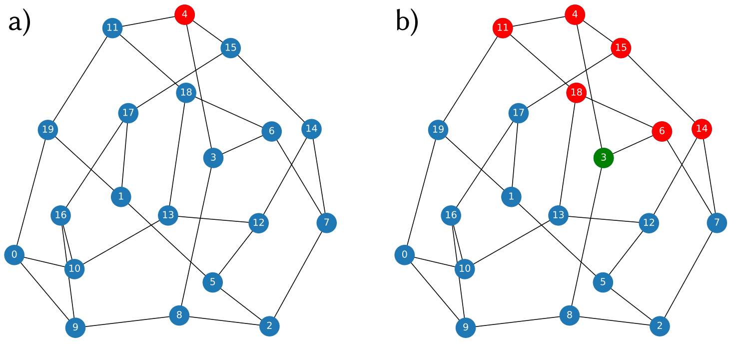

Figure 1.2: **Example of a SIR epidemic spreading on a contact network:** In this case, the contact network is a Random Regular Graph, with 20 nodes and fixed degree $k = 3$. Blue nodes are S, red nodes are I and green nodes are R. Panel a) shows the initial state of the epidemic, panel b) an intermediate snapshot (at $t = 3$) and panel c) shows the final state ($t = 6$).

The probability of obtaining a given epidemic cascade $\underline{\mathbf{x}} = \left(\mathbf{x}^{t=0}, \mathbf{x}^{t=1}, \dots \mathbf{x}^{t=T}\right)$, running from time 0 to time T, can thus be written as

$$p\left(\underline{\mathbf{x}}\right) = p(\mathbf{x}^0) \prod_{t=1}^{T} \prod_{i=1}^{N} p\left(x_i^t \mid \mathbf{x}^{t-1}\right) \tag{1.11}$$

where $\mathbf{x}^0$ is the initial state of the system. The difference with respect to the stochastic model of section 1.2.2 is that now an individual cannot be infected by any other, but only by its neighbours in the graph $\mathcal{G}$. In order to obtain a situation similar to the stochastic homogeneous model, a complete contact graph would be required, where everybody is in contact with everyone. Also in this case, it's possible to run simulations of this model using equations (1.10a) to (1.10c), with a given contact graph and initial state. One example of such a simulation is given in figure 1.2.

It's important to note that this epidemic process is irreversible: once an individual is infected, they cannot go back to being susceptible, and the same is true for the recovered and infected state, respectively. This characteristic is of essential importance in the inference of epidemics.

The probabilistic network-based SIR model, in contrast with the previous ones, also shows a much higher degree of complexity. For example, if one would like to determine the probability of an individual i being in state $X \in \{S, I, R\}$ at some time instant τ, it would be necessary to sum on all the possible configurations available at time $\tau - 1$:

$$p\left(x_i^\tau = X\right) = \sum_{\mathbf{x}^{\tau-1}} p\left(x_i^\tau = X \mid \mathbf{x}^{\tau-1}\right) p\left(\mathbf{x}^{\tau-1}\right) \tag{1.12}$$

and possible number of such configurations is 3^N.

Then, one might want to find such probability from the information about the probability of the initial state $p\left(\mathbf{x}^0\right)$. For example, it might be that each individual is either susceptible or infected, but not recovered, and the probability of each node being infected is η. This probability is called *patient zero* probability, and gives the following probability distribution:

$$p\left(\mathbf{x}^0\right) = \prod_{i=1}^{N} \gamma_i(x_i^0) \tag{1.13}$$

with

$$\gamma_i(x) = \begin{cases} 1-\eta & x = S \\ \eta & x = I \\ 0 & x = R \end{cases} \tag{1.14}$$

In order to compute $p\left(x_i^\tau\right)$, then, it's necessary to iterate equation (1.12), going backward in time until $t = 0$, summing on all possible configurations:

$$\begin{aligned} p\left(x_i^\tau = X\right) &= \sum_{\mathbf{x}^{\tau-1}} p\left(x_i^\tau = X \mid \mathbf{x}^{t-1}\right) \sum_{\mathbf{x}^{\tau-2}} p\left(\mathbf{x}^{\tau-1} \mid \mathbf{x}^{\tau-2}\right) \sum_{\mathbf{x}^{\tau-3}} p\left(\mathbf{x}^{\tau-2} \mid \mathbf{x}^{\tau-3}\right) p\left(\mathbf{x}^{\tau-3}\right) \\ &= \sum_{\{\mathbf{x}^{\tau-1},\mathbf{x}^{\tau-2},\dots,\mathbf{x}^{1},\mathbf{x}^{0}\}} p\left(x_i^\tau = X \mid \mathbf{x}^{\tau-1}\right) \prod_{t=1}^{\tau-1} p\left(\mathbf{x}^{t} \mid \mathbf{x}^{t-1}\right) p\left(\mathbf{x}^{0}\right) \\ &= \sum_{\{\mathbf{x}^{\tau-1},\mathbf{x}^{\tau-2},\dots,\mathbf{x}^{1},\mathbf{x}^{0}\}} p\left(x_i^\tau = X \mid \mathbf{x}^{\tau-1}\right) \prod_{t=1}^{\tau-1} p\left(\mathbf{x}^{t} \mid \mathbf{x}^{t-1}\right) \prod_{j=1}^{N} \gamma\left(x_j^0\right) \end{aligned} \tag{1.15}$$

It can be seen that the number of configurations taken under consideration to perform the sum grows exponentially with τ (more precisely as $O\left(3^{\tau N}\right)$), making this a very costly computation that becomes practically infeasible even with medium-sized systems. In the following, we will discuss the simplest approximation that can be made to the model in order to avoid the incumbency of summing on every state at every time.

1.3.1 Mean-field approximation

Computing the probability of a certain individual at a certain time is not computationally light itself. From equation (1.12), the probability of being in state S can be written as:

$$\begin{aligned} p\left(x_i^\tau = S\right) &= \sum_{\mathbf{x}^{\tau-1}} p\left(x_i^\tau = S \mid \mathbf{x}^{\tau-1}\right) p\left(\mathbf{x}^{\tau-1}\right) \\ &= \sum_{\mathbf{x}_{\partial i}^{t-1}} p\left(x_i^{\tau-1} = S, \mathbf{x}_{\partial i}^{\tau-1}\right) \prod_{j\in\partial i} \left(1 - \lambda_{ji} \mathbb{I}\left[x_j^{\tau-1} = I\right]\right) \end{aligned} \tag{1.16}$$

where we have used the transition probability of the process (1.10a) and the fact that, since it only depends on the neighbours of i, the probability of the state of the other individuals $k \notin \partial i$ at time $\tau - 1$ sums up to 1 (the variable $\mathbf{x}_{\partial i}^{t-1}$ thus represents the state of all the neighbours of i at time τ). Now, in order to compute this probability exactly, one could perform the sum: the number of states to consider then would still be growing exponentially in the number of neighbours. To avoid the sum, then, one can employ a common approximation in statistical physics, called Mean Field approximation. This amounts to treat all the variables in the system as independent, thus neglecting all the correlations that are present. This in mathematical terms means $p(\mathbf{x}) \approx \prod_i p(x_i)$. Applying this approximation to the probabilities in the above equation, one gets that

$$p\left(x_i^{\tau-1}, \mathbf{x}_{\partial i}^{\tau-1}\right) \approx p\left(x_i^{\tau-1}\right) \prod_{j\in\partial i} p\left(x_j^{\tau-1}\right),$$

which effectively decouples the probabilities between neighbours. Then, the summation in (1.16) can be applied directly to each state $x_j^{\tau-1}$, giving

$$p\left(x_i^\tau = S\right) \approx p\left(x_i^{\tau-1} = S\right) \prod_{j\in\partial i} \left(1 - \lambda_{ji} p\left(x_j^{\tau-1} = I\right)\right) \tag{1.17}$$

Applying the same treatment to $p\left(x_i^T\right)$ for the other states gives the equations for what, in the following, will be called Individual-based Mean Field (IMF):

$$p_{\mathrm{MF}}\left(x_i^{t+1}=S\right)=p_{\mathrm{MF}}\left(x_i^t=S\right)\prod_{j\in\partial i}\left[1-\lambda_{ji}p_{\mathrm{MF}}\left(x_j^t=I\right)\right] \tag{1.18a}$$

$$p_{\mathrm{MF}}\left(x_i^{t+1}=I\right)=p_{\mathrm{MF}}\left(x_i^t=S\right)\left[1-\prod_{j\in\partial i}\left(1-\lambda_{ji}p_{\mathrm{MF}}\left(x_j^t=I\right)\right)\right]+p_{\mathrm{MF}}\left(x_i^t=I\right)\left(1-\mu_i\right) \tag{1.18b}$$

$$p_{\mathrm{MF}}\left(x_i^{t+1}=R\right)=p_{\mathrm{MF}}\left(x_i^t=R\right)+p_{\mathrm{MF}}\left(x_i^t=I\right)\mu_i \tag{1.18c}$$

These equations have the advantage of presenting an intuitive content which is easy to understand, as they reflect the equations of the transition probabilities. Iterating these equations forward in time from $p\left(\mathbf{x}^{t=0}\right)$, one can obtain a set of probabilities the whole epidemic propagation. However, neglecting the correlations between neighbouring nodes induces a lot of errors, leading to an overestimation of the number of infected individuals in the system (see section 6.3). To tackle this issue, some methods have been developed that also avoid the costly sum on the number of states and give much better approximations [39, 44].

1.3.2 Extension to the SEIR model

While the SIR model is detailed enough to well describe the spreading of diseases with lasting immunity, in its Markovian formulation it might be insufficient to describe certain diseases, which are subject to a non-negligible incubation period: in these cases, following infection, an individual is not immediately able to immediately infect others [137]. To model this effect, another compartment/state is introduced for *exposed* individuals: in the Susceptible-Exposed-Infectious-Recovered (SEIR) model, exposed individuals count as infected, but they cannot infect others. Therefore, the state transitions are

$$\begin{aligned}S+I&\to E+I\\E&\to I\\I&\to R\end{aligned}$$

Then, for the probabilistic SEIR model on a contact network we need to introduce the probability η_i for individual i to pass from the exposed to infectious state in one time step. The full equations read:

$$p\left(x_i^{t+1}=S\mid\mathbf{x}^t\right)=\mathbb{I}\left[x_i^t=S\right]\prod_{j\in\partial i}\left(1-\lambda_{ji}\mathbb{I}\left[x_j^t=I\right]\right) \tag{1.19a}$$

$$p\left(x_i^{t+1}=E\mid\mathbf{x}^t\right)=\mathbb{I}\left[x_i^t=S\right]\left[1-\prod_{j\in\partial i}\left(1-\lambda_{ji}\mathbb{I}\left[x_j^t=I\right]\right)\right]+\mathbb{I}\left[x_i^t=E\right]\left(1-\eta_i\right) \tag{1.19b}$$

$$p\left(x_i^{t+1}=I\mid\mathbf{x}^t\right)=\mathbb{I}\left[x_i^t=I\right]\left(1-\mu_i\right)+\mathbb{I}\left[x_i^t=E\right]\eta_i \tag{1.19c}$$

$$p\left(x_i^{t+1}=R\mid\mathbf{x}^t\right)=\mathbb{I}\left[x_i^t=R\right]+\mathbb{I}\left[x_i^t=I\right]\mu_i \tag{1.19d}$$

1.4 Epidemic spreading in temporal networks

In many realistic cases, contacts between individuals change over time, with a time scale comparable to that of the epidemic spreading process. In these cases, we have a *temporal network*,

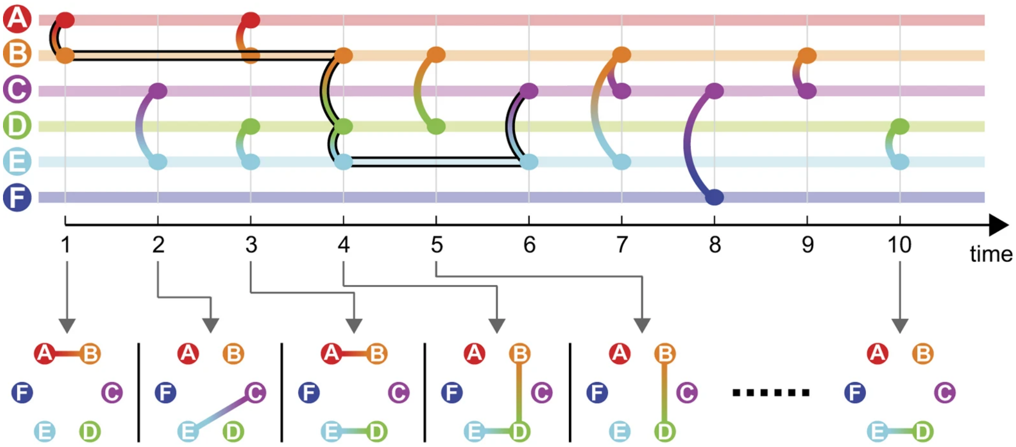

Figure 1.3: A sample temporal network ($N = 6, T = 10$) represented as a sequence of contacts. At the top, the network is shown in a timeline fashion, with each edge drawn at its corresponding time, while in the bottom several snapshots of the network are shown at different time instants. In this case nodes are marked with letters, and the time indices start from 1. Figure taken from [49], with Creative Commons 4.0 Attribution License, and unmodified.

that is generally defined as a network that has different edges at different time instants [30]. This means that for every pair of individuals i and j, we have a list of the time instants they are in contact: while these could well be real-valued, we will restrict ourselves to integer times, to be consistent with our SIR spreading model. This means that the temporal networks considered will be well described as a series of static graphs with different time indices, that will be the temporal snapshots of the network.

Contrary to the static network case discussed so far, the additional dimension of time in the network here contributes significantly to the spreading of an epidemic. In fact, it becomes critical whether a node i becomes infected *before* coming in contact with j, rather than after. This sparsity of contacts further restricts the number of possible epidemic cascades: if one creates a temporal network from a static one and tehn randomly drawing the times at which each edge is present, it is easy to see that the number of possible spreading events is actually reduced. At the same time, a temporal network collected in a real-world setting might see moments of sharp increase in the number of contacts, like in the case of an event where large crowds are present. A lot of effort has been put into the collection of proximity data in real-world scenarios in recent years, in both human and animal populations [26, 61, 87, 143]. While the scale of the contact datasets is somewhat limited, there are some studies that collected large-scale high-accuracy data [40, 67].

With an underlying temporal network, the formulation of the SIR model on a graph changes slightly. Assuming that from the proximity data a temporal network can be defined, where each edge represents a contact at each time instant and has an associated probability of infection λ_{ij}^{t}, we can rewrite equations (1.10a) and (1.10b) as:

$$p\left(x_i^{t+1} = S \mid \mathbf{x}^t\right) = \mathbb{I}\left[x_i^t = S\right] \prod_{j \in \partial i_t} \left(1 - \lambda_{ji}^t \mathbb{I}\left[x_j^t = I\right]\right)$$

$$p\left(x_i^{t+1} = I \mid \mathbf{x}^t\right) = \mathbb{I}\left[x_i^t = S\right] \left[1 - \prod_{j \in \partial i_t} \left(1 - \lambda_{ji}^t \mathbb{I}\left[x_j^t = I\right]\right)\right] + \mathbb{I}\left[x_i^t = I\right] (1 - \mu_i) \tag{1.20}$$

Where now the products run over the set ∂i_t of the individuals in contact with i at the time instant t.

1.5 Bayesian inference on the SIR model

Whenever there is a model for a certain process or event, along with a distribution that gives the probabilities for the outcomes of this process, there is the issue of updating a the model when new observations (or evidence) become available. The model should then be updated to account for this new evidence, for example, if one wanted to predict the most probable outcomes, given the previous observations, or to understand which factors were the most probable causes of the events observed. When employing the Bayesian approach to probability theory, this is in principle straightforward. Starting from a certain probability distribution describing the process (in our case, equation (1.11)), it's possible to update this distribution so that it reflects the new evidence that we have on the process realization.

This is done by the use of Bayes' rule, that says that given two events A and B, and that $p(B) > 0$, if we have observed event B has happened, we can compute the probability of A happening as well:

$$p(A \mid B) = \frac{p(B \mid A)\, p(A)}{p(B)}$$

The first term in the numerator on the right-hand side is then called the *likelihood*, while the $p(A)$ is called the *prior* probability. This is because this equation can be interpreted as updating the probability $p(A)$ with the information that event B has happened, conveyed by $p(B \mid A)$. The left-hand side is called the *posterior* probability, and is the updated one with the information on what happened.

In the case of the SIR probabilistic model, our prior probability distribution is given by equation (1.11). So, calling $\mathcal{O}$ the set containing the observations, the update rule with Bayes' theorem reads

$$p(\underline{\mathbf{x}} \mid \mathcal{O}) = \frac{p(\mathcal{O} \mid \underline{\mathbf{x}})\, p(\underline{\mathbf{x}})}{p(\mathcal{O})} \tag{1.21}$$

Therfore, in order to apply this rule, one needs to compute both the likelihood $p(\mathcal{O} \mid \underline{\mathbf{x}})$ and the denominator $p(\mathcal{O})$. The former can be written in a simple way: since it is assumed that observations come in the form of medical tests performed on the individuals, it is acceptable to assume that each one is independent from the others. Then the probability of a set of tests $\mathcal{O}$ can be written as

$$p(\mathcal{O} \mid \underline{\mathbf{x}}) = \prod_{(i,t,z) \in \mathcal{O}} p\left(\tilde{x}_i^t = z \mid x_i^t\right) \tag{1.22}$$

Where we introduced the variables $\tilde{x}_i^t \in \{S, I, R\}$, each one associated to an observation, which is mathematically defined as the triplet of the time of observation t, the index i and the observed state z of the individual being tested. This way of writing the likelihood accounts for the fact that the observations could be imprecise, like testing for a viral infection with a tool-kit that provides a fast result but with a low accuracy [25, 160].

The prior probability of the observations, $p(\mathcal{O})$, should account all the ways in which these observations can be derived, without considering the specific epidemic realization. This in practice amounts to consider every possible epidemic cascade:

$$p(\mathcal{O}) = \sum_{\underline{\mathbf{x}}} p(\mathcal{O}, \underline{\mathbf{x}}) = \sum_{\underline{\mathbf{x}}} p(\mathcal{O} \mid \underline{\mathbf{x}})\, p(\underline{\mathbf{x}}) \tag{1.23}$$

It is plain to see that the term in the sum is just the numerator of Bayes' rule, so $p(\mathcal{O})$ can be substituted with a normalization term. For this reason, from now on we will write equation (1.21) as:

$$p(\underline{\mathbf{x}} \mid \mathcal{O}) = \frac{1}{Z}\, p(\mathbf{x}^0) \prod_{t=1}^{T} \prod_{i=1}^{N} p\left(x_i^t \mid \mathbf{x}^{t-1}\right) \prod_{(i,\tau,z)\in\mathcal{O}} p\left(\tilde{x}_i^\tau = z \mid x_i^\tau\right) \tag{1.24}$$

where the normalization constant is called Z in analogy with statistical physics, and $Z = Z(\mathcal{O})$.

This means that in order to do exact Bayesian inference on the SIR Model, it is necessary, in principle, to sum over all epidemic cascades that is possible to obtain on a given contact graph, and then account for compatibility with the observations. This has a computational cost of order $O\left(3^{NT}\right)$, so it rapidly becomes practically infeasible when the network or the number of time intervals become large. Doing exact Bayesian inference at the individual level in epidemics is thus an open problem, for which no reasonably fast techniques have been discovered yet. Therefore, in order to find an efficient technique for inference, it is necessary to make some approximations [76].

1.5.1 The source inference problem

It has been shown above that the inference of the whole epidemic cascade from a set of observations can be a difficult problem. While being able to know the posterior distribution (1.24) would effectively solve any problem in epidemic inference, it should be noted that in the cases where we do not need the entire epidemic cascade, but just the states at a specific time instant. In these cases, a different solution can be formulated.

Let us now consider the case when the inference is focused on the initial state of the epidemic. In this case, it is important to find the individuals who were infected at the beginning, called *patient zeros.* The posterior probability of the initial state can be written as:

$$\begin{aligned} p\left(\mathbf{x}^0 \mid \mathcal{O}\right) &= \sum_{\mathbf{x}^1,\mathbf{x}^2,\dots,\mathbf{x}^T} p(\underline{\mathbf{x}} \mid \mathcal{O}) \\ &= \frac{1}{Z}\left(\sum_{\mathbf{x}^1,\mathbf{x}^2,\dots,\mathbf{x}^T} \prod_{t=1}^{T} \prod_{i=1}^{N} p\left(x_i^t \mid \mathbf{x}^{t-1}\right) p(\mathcal{O} \mid \underline{\mathbf{x}})\right) p(\mathbf{x}^0) \\ &= \frac{1}{Z} p\left(\mathcal{O} \mid \mathbf{x}^0\right) p(\mathbf{x}^0) \end{aligned} \tag{1.25}$$

which corresponds to applying Bayes' theorem to the initial state of the epidemic only. Analogously, one can write Bayes' rule for the case in which we consider only one individual infected at the beginning of the epidemic and want to infer marginal source probabilities $p(s_0 = i)$:

$$p(s_0 = i \mid \mathcal{O}) \propto p(\mathcal{O} \mid s_0 = i)\, p(s_0 = i) \tag{1.26}$$

This formulation does not overly simplify the problem, as one still has to consider a large number of possible realizations in order to compute the likelihood term, but it can lead to approximate methods for the calculation of the posterior which may be simpler and less computationally intensive then the one used for the inference of the full epidemic cascade.

Chapter 2

Graphical models and Belief Propagation

2.1 Factor graphs

In statistical physics, systems of many interacting variables are often found to exhibit interactions that are local, and often mutually dependent. In many of these instances it is possible to 'factorize' the probability distribution of the system, that is writing it as the product of several factors that depend on different sets of variables. In this case, one natural and efficient way to represent the probability distribution is through the use of *factor graphs* [23].

Factor graphs are bipartite undirected graphs made of two types of vertices, which are the variables and factors. Assume that there are N variables $x_1, x_2, \dots, x_N$, and their probability distribution can be written as

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{a=1}^{M} \psi_a(\mathbf{x}_{\partial a}) \tag{2.1}$$

where M factors ψ_a can be identified, each one having a set of dependencies variables $\mathbf{x}_{\partial a}$. For example, Let us consider a simple case in which there are $N = 7$ variables:

$$P(\mathbf{x}) = \frac{1}{Z}\, \psi_a(x_1, x_2, x_3, x_6)\, \psi_b(x_3, x_4, x_5)\, \psi_c(x_5, x_6, x_7)\, \psi_d(x_5) \tag{2.2}$$

Then there are four factors, and the corresponding factor graph is depicted in Figure 2.1, where each factor is connected to every variable it is dependent on. As a convention, in the rest of the thesis squares will be used to represent factor nodes in a factor graph. Factor graphs are the basis for the study of *graphical models*, which are often used not only in physics, but also in combinatorial optimization, machine learning and Bayesian inference in general [8, 23, 12, 21, 31]. In this sense, the factors ψ can also be called *compatibility functions*, and they only need to give non-negative values, while Z has to be a positive constant.

A factor graph then is usually redefined as $\mathcal{G} = (V, F, E)$, splitting the nodes into the variables V and the factors F. When all the factor nodes have degree 2, a factor graph can be reduced to a simple graph, as in this case each factor ψ_a can actually be renamed $\psi_{i,j}$ from the indices of the variables to which it is connected, and so in the simplified graph it will be represented by an edge (i, j).

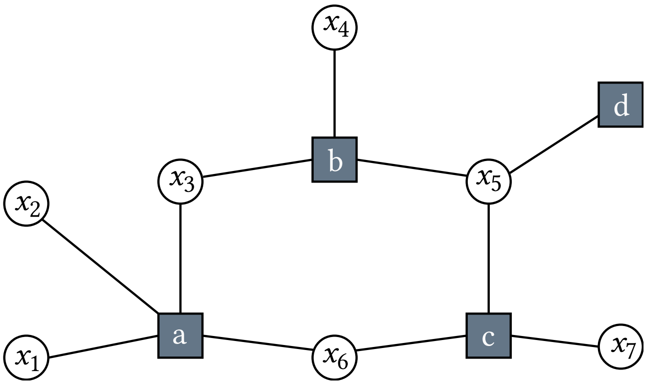

Figure 2.1: Example of a factor graph with 7 variables and 4 factors

2.2 Belief Propagation

When dealing with a graphical model, it is often the goal to calculate the marginal distribution of some of the variables in the factor graph. The most naive way to do this is to sum on all the possible configurations, which has a time cost that goes exponentially with the number of variables involved. However, when the factor graph of the complete distribution is a tree, it is possible to compute marginal distributions of the variable, or alternatively the partition function Z, in a way that involves much less computational cost. The *Belief Propagation* (BP) algorithm belongs to the class of *message passing* methods, that exploit the local natures of the interactions of a model by using distributed computation. Its equations have been derived several times in different fields: in physics, the idea was first put forward by Bethe [9]; in coding theory, it has been used for error correction in low-density parity-check codes [3], while in Bayesian inference the algorithm was proposed by Pearl [5].

Again, let us assume that there is a distribution of N variables $x_1, x_2, \dots, x_N$, with the important additional assumption now that the corresponding factor graph is a tree:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{a=1}^{M} \psi_a(\mathbf{x}_{\partial a}) \tag{2.3}$$

This is not the case for the previous example, as the graph in figure 2.1 has a loop, however it can become a tree by removing one edge, thus changing the functional form of the corresponding factor, like, for example, eliminating the dependency of ψ_a on x_3.

In the following, in order to use a light notation, the set of the variables that are connected to a factor a will be denoted as ∂a, and analogously the set of the factors connected to a particular variable i will be called ∂i. Also, indices i, j, k will be used for node variables, and a, b, c for factor nodes.

The BP algorithm defines two types of messages, and one message of each kind for each edge (i, a) of the factor graph: one is $m_{a\to i}(x_i)$, that goes from factor node to variable node, and the other is $\nu_{i\to a}(x_i)$, that goes from variable to factor node. Each message takes values in the space of probability distributions of the variable of the relative edge, so that $m_{a\to i}(x_i) \geq 0$ (and $\nu_{i\to a}(x_i) \geq 0$), and is normalized, so that $\sum_{x_i} m_{a\to i}(x_i) = 1$ (and $\sum_{x_i} \nu_{i\to a}(x_i) = 1$).

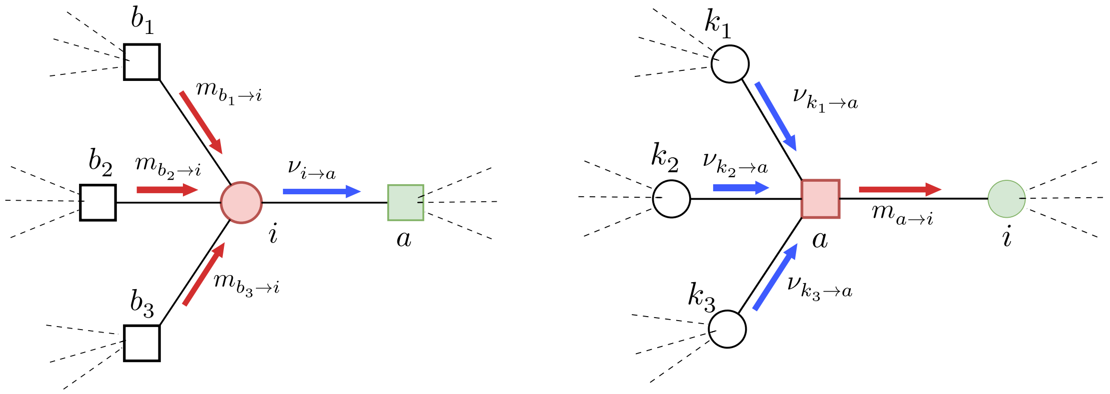

Figure 2.2: Graphical representation of the update of messages in belief propagation: on the left, the update for the messages from variable to factor, on the right from factor to variable. To update one outgoing message along an edge, the incoming messages on the other edges are multiplied together (which are of the opposite type). In the examples, each node has 4 neighbours for convenience.

In the Belief Propagation scheme, messages $m_{a\to i}$ and $\nu_{i\to a}$ are updated, edge by edge, from the values of the messages from the other edges. The update rules are the following:

$$\nu^{(t+1)}_{i\to a}(x_i) \propto \prod_{b\in\partial i\setminus a} m^{(t)}_{b\to i}(x_i) \tag{2.4}$$

$$m^{(t+1)}_{a\to i}(x_i) \propto \sum_{\mathbf{x}_{\partial a\setminus i}} \psi_a(\mathbf{x}_{\partial a}) \prod_{k\in\partial a\setminus i} \nu^{(t)}_{k\to a}(x_k) \tag{2.5}$$

where the superscript (t) indicates the update round, and the shorthand $\partial a\setminus j$ stands for all the neighbour indices of factor a, excluding j (analogously $\partial j\setminus a$ stands for the neighbour factors of variable j, excluding a). After doing many iterations of the above equations, it is possible to give an estimate of the of a single variable marginals of the original distribution (2.3), by the product of all the incoming messages in the variable node of the factor graph:

$$b^{(t+1)}_i(x_i) = \frac{1}{Z_i}\prod_{a\in\partial i} m^{(t)}_{a\to i}(x_i) \tag{2.6}$$

These *beliefs* for the probability can be updated every time the messages are. The distribution of the variables of a single factor can also be estimated as:

$$b^{(t+1)}_a(\mathbf{x}_{\partial a}) = \frac{1}{Z_a}\prod_{i\in\partial a} \nu^{(t)}_{i\to a}(x_i) \tag{2.7}$$

Running (2.4) and (2.5) many times for all variables and factor nodes on a tree factor graph ensures that the messages converge in a finite number of iterations t^*, and that at convergence the beliefs obtained correspond exactly to the marginals, $b^*_i = \mu_i(x_i)$ and (see [23] for the proof). It can also be shown that, when the factor graph is a tree, the probability distribution if all the variables can be written as a product of the marginals:

$$p(\mathbf{x}) = \prod_{a\in F}\mu_a(\mathbf{x}_{\partial a})\prod_{i\in V}\mu_i(x_i)^{1-|\partial i|} \tag{2.8}$$

where $\mu_a(\mathbf{x}_{\partial a})$ is the marginal probability distribution of all the variables that are connected to factor a. In fact, these marginals correspond to the beliefs at convergence of the message passing, $\mu_a = b_a^*(\mathbf{x}_{\partial a})$. This result can be proven by induction on the factor graph [23], and has important consequences. For instance, the entropy $S[p] = -\sum_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x})$ can be expressed as:

$$S[p] = -\sum_{a \in F} \sum_{\mathbf{x}_{\partial a}} \mu_a(\mathbf{x}_{\partial a}) \log \mu_a(\mathbf{x}_{\partial a}) - \sum_{i \in V} (1 - |\partial i|) \sum_{x_i} \mu_i(x_i) \log \mu_i(x_i) \tag{2.9}$$

If the graphical model in (2.3) is treated as a physical system, each factor ψ_a then can be written as $\psi_a = \exp\{-\beta E_a(\mathbf{x}_{\partial a})\}$, where E_a is the energy contribution of the factor, and β is the inverse temperature. This is done with probabilistic models by putting $\beta = 1$, and this convention will be used in the rest of the thesis, unless stated otherwise. However, if a graphical model contains factors that could become 0, as is the case with hard constraints, special care must be taken with the energy representation, as the energy E_a corresponding to $\psi_a = 0$ is infinite.

Excluding systems with this kind of issues, the average energy U can then be written

$$U[p] = -\sum_{a \in F} \sum_{\mathbf{x}_{\partial a}} \mu_a(\mathbf{x}_{\partial a}) \log \psi_a(\mathbf{x}_{\partial a}) \tag{2.10}$$

which leads to the free energy $\mathbb{F} = U - S$ being calculated as:

$$\mathbb{F}[p] = -\sum_{a \in F} \sum_{\mathbf{x}_{\partial a}} \mu_a(\mathbf{x}_{\partial a}) \log \frac{\psi_a(\mathbf{x}_{\partial a})}{\mu_a(\mathbf{x}_{\partial a})} - \sum_{i \in V} (|\partial i| - 1) \sum_{x_i} \mu_i(x_i) \log \mu_i(x_i) \tag{2.11}$$

When the factor graph is a tree, it is also possible to write $\mathbb{F}$ as an expression of the BP messages at convergence (2.4):

$$\mathbb{F} = \sum_{a \in F} \mathbb{F}_a - \sum_{(a,i) \in E} \mathbb{F}_{ai} + \sum_{i \in V} \mathbb{F}_i \tag{2.12}$$

where:

$$\mathbb{F}_a = \log \left[\sum_{\mathbf{x}_{\partial a}} \psi_a(\mathbf{x}_{\partial a}) \prod_{i \in \partial a} \nu^*_{i \to a}(x_i) \right] \tag{2.13}$$

$$\mathbb{F}_{ai} = \log \left[\sum_{x_i} m^*_{a \to i}(x_i)\, \nu^*_{i \to a}(x_i) \right] \tag{2.14}$$

$$\mathbb{F}_i = \log \left[\sum_{x_i} \prod_{b \in \partial i} m^*_{b \to i}(x_i) \right] \tag{2.15}$$

2.2.1 Bethe approximation and loopy belief propagation

In the case the factor graph is not a tree, but contains loops, it is not possible to use the messages at convergence to compute the exact marginals or the free energy. However, the same expressions can still be used to give an approximation, assuming that, locally, the graph is tree-like. In fact, rewriting equation (2.11) substituting the marginals with the beliefs gives:

$$\mathbb{F}[\mathbf{b}] = -\sum_{a \in F} \sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) \log \frac{\psi_a(\mathbf{x}_{\partial a})}{b_a(\mathbf{x}_{\partial a})} - \sum_{i \in V} (|\partial i| - 1) \sum_{x_i} b_i(x_i) \log b_i(x_i) \tag{2.16}$$

which defines the *Bethe free energy*. The beliefs b do not need to be defined as the fixed points of the BP equations, but simply as non-negative distributions ($b_i(x_i) \geq 0$, $b_a(\mathbf{x}_{\partial a}) \geq 0$) that are normalized ($\sum_{x_i} b_i(x_i) = 1$, and $\sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) = 1$), that respect the condition of being *locally consistent*, that is the following marginalization:

$$\sum_{\mathbf{x}_{\partial a \setminus i}} b_a(\mathbf{x}_{\partial a}) = b_i(x_i) \tag{2.17}$$

Then, from this definition it can be shown that the $\mathbf{b} = \{b_i, b_a\}$ corresponding to the stationary points of the Bethe free energy (2.16) do in fact also have one-to-one correspondence to fixed points of the BP equations (2.4) - (2.5) [23]. This case is then called *Loopy Belief Propagation* (sometimes abbreviated as LBP), even if the equations used for the fixed point calculation are the same, because the exactness of the fixed points is lost, and even reaching the convergence of the BP equation is not guaranteed.

However, loopy BP has found wide adoption in practice, as the result of the approximation is generally better then simpler Mean Field methods [32, 36, 59].

2.3 Belief propagation for epidemic inference

In this section, a derivation of the belief propagation method for epidemic processes will be shown [38, 37]. This method, which can perform approximate Bayesian inference of epidemics with the SIR model on a graph, has been used extensively during the thesis' work, either as a comparison or for application on different epidemic models.

Let us recall now the probabilistic inference setting. We will consider epidemic cascades spreading on a contact network according to the SIR model (see section 1.3). The probability distribution $p(\underline{\mathbf{x}})$ of one such cascade can be written as:

$$p(\underline{\mathbf{x}}) = \gamma(\mathbf{x}^0) \prod_{t=1}^{T} \prod_{i=1}^{N} p\left(x_i^t \,\middle|\, \mathbf{x}^{t-1}\right) \tag{2.18}$$

where the γ is the probability of the initial state, and $p\left(x_i^t \,\middle|\, \mathbf{x}^{t-1}\right)$ is the transition probability given by the SIR model (1.10a) - (1.10c).

2.3.1 Infection times representation and factor graph

The irreversible nature of the SIR model makes it possible to explicitly map an epidemic cascade $\underline{\mathbf{x}}$ to a set of infection times $t_i = \min\{t : x_i^t = I\}$ and recovery times $r_i = \min\{t : x_i^t = R\}$ for each individual i. In this definition, if an individual has never been infected (or, respectively, recovered), her $t_i = T^\infty$ (respectively, $r_i = T^\infty$). This T^∞ should in principle be infinite to represent an event that never happened, however, since the final time of observation of the epidemic is T, it is equally possible to use any larger value, like $T^\infty = T + 1$. Then, the mapping between the times and the state is very simple:

$$X(t, t_i, r_i) = \begin{cases} S & \text{if } t < t_i \\ I & \text{if } t_i \leq t < r_i \\ R & \text{if } t \geq r_i \end{cases} \tag{2.19}$$

It is however more intuitive to introduce the recovery delay for every individual $g_i = r_i - t_i$, that is the time it takes for an infected individual to become recovered. In the SIR Markov model (1.10a) - (1.10c), this follows a geometric distribution:

$$\mathbb{G}_i\left(g_i\right) = \mu_i\left(1 - \mu_i\right)^{g_i - 1} \tag{2.20}$$

where μ_i is the recovery probability for each time interval. Now, in order to account for the propagation of the disease, it is easy to see that, after a person i has been infected, the disease propagates to j with a *transmission delay* s_{ij}. This delay follows a probability distribution $\omega_{ij}\left(s_{ij} \,\middle|\, g_i\right)$, which is depending on g_i because, once the individual has recovered, he/she cannot infect any more. In the Markov SIR Model this distribution reads:

$$\omega_{ij}\left(s_{ij} \,\middle|\, g_i\right) = \begin{cases} \lambda_{ij}\left(1 - \lambda_{ij}\right)^{s_{ij}} & s_{ij} < g_i \\ \sum_{s \geq g_i} \lambda_{ij}\left(1 - \lambda_{ij}\right)^{s} & s_{ij} = \infty \end{cases} \tag{2.21}$$

where the probability of all the other cases is concentrated on the infinite time case. Then, infection times are now deterministic, that is, they have to satisfy the relation

$$t_j = 1 + \min_{k \in \partial j}\left\{t_k + s_{kj}\right\} \tag{2.22}$$

With this framework, it is now possible, when given an initial state $\mathbf{x}^0$, to generate g_i and s_{ij} and that correspond to an SIR epidemic cascade. Formally, the probability of such a cascade can now be written as

$$\begin{aligned} p\left(\mathbf{t}, \mathbf{g} \,\middle|\, \mathbf{x}^0\right) &= \sum_{\mathbf{s}} p\left(\mathbf{t} \,\middle|\, \mathbf{s}, \mathbf{g}, \mathbf{x}^0\right) p\left(\mathbf{s} \,\middle|\, \mathbf{g}\right) p\left(\mathbf{g}\right) \\ &= \sum_{\mathbf{s}} \prod_{i,j} \omega_{ij}\left(s_{ij} \,\middle|\, g_i\right) \prod_{i} \phi_i\left(t_i, \{t_k, s_{ki}\}_{k \in \partial i}\right) \mathbb{G}_i\left(g_i\right) \end{aligned} \tag{2.23}$$

where

$$\phi_i\left(t_i, \{t_k, s_{ki}\}_{k \in \partial i}\right) = \delta\left(t_i, \mathbb{I}\left[x_i^0 \neq I\right]\left(1 + \min_{k \in \partial i}\{t_k + s_{ki}\}\right)\right) \tag{2.24}$$

enforces the dynamical constraint on each node i.

The probability distribution written so far (2.23), is problematic for applying Belief Propagation, as the factor graph describing the distribution contains a lot of short range loops (see Figure 2.3). In order to develop a BP method for the computation of the marginals of this distribution, it would be ideal to reach a factor graph representation that retains the same topological properties as the original contact graph. This solution then would also retain the convergence guarantees when the contact graph is a tree.

In order to ”disentangle” the factor graph, it is necessary to introduce *copy* variables for the infection times $t_i^{(j)}$ for each edge (i, j) of the contact graph. These variables are set equal to t_i with an extra constraint $\prod_{j \in \partial i} \delta\left(t_i^{(j)}, t_i\right)$ (where ∂i is the set of neighbours of i in the contact graph).

Then, since the factors ϕ_i depend on the infection times and delays only trough their sum, it is more convenient to introduce $t_{ij} = t_i^{(j)} + s_{ij}$. Now the factors ω_{ij} and ω_{ji} can be put together in the same factor ϕ_{ij}:

$$\phi_{ij} = \omega_{ij}\left(t_{ij} - t_i^{(j)} \,\middle|\, g_i^{(j)}\right) \omega_{ji}\left(t_{ji} - t_j^{(i)} \,\middle|\, g_j^{(i)}\right) \tag{2.25}$$

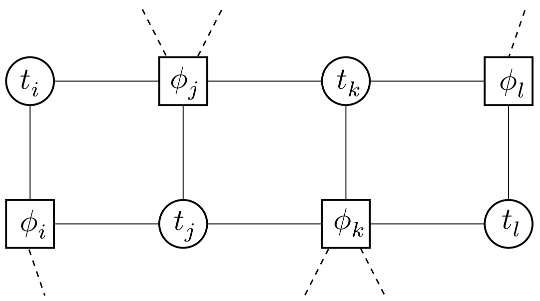

Figure 2.3: The factor graph given by the probability distribution (2.23), with a fixed contact graph that is a linear chain $i - j - k - l$. Only a part of the factor graph is drawn, omitting the s_{ij}, g_i variables and related factors, but it can already be seen that this situation is far from ideal for BP, as there are many short range loops.

where the copy variables (on the edges) for g_i and g_j have been introduced. The constraints for $g_i^{(j)}$ and $g_j^{(i)}$ can be collected with the others in another factor ψ_i:

$$\psi_i = \delta\left(t_i, \mathbb{I}\left[x_i^0 \neq I\right]\right) \prod_{j\in\partial i} \delta\left(t_i^{(j)}, t_i\right) \delta\left(g_i^{(j)}, g_i\right) \tag{2.26}$$

Also, the probability of the initial state, $p\left(\mathbf{x}^0\right)$, can be assumed to be independent for each individual i:

$$\begin{aligned} p\left(\mathbf{x}^0\right) &= \prod_i \gamma_i\left(x_i^0\right) \\ \gamma_i\left(x_i^0\right) &= \gamma\delta\left(x_i^0, I\right) + (1-\gamma)\,\delta\left(x_i^0, S\right) \end{aligned} \tag{2.27}$$

The conversion function between states and times has to be introduced as another constraint:

$$\zeta_i^t = \mathbb{I}\left[x_i^t = S, t < t_i\right] + \mathbb{I}\left[x_i^t = I, t_i \leq t < t_i + g_i\right] + \mathbb{I}\left[x_i^t = R, t \geq t_i + g_i\right] \tag{2.28}$$

so that is possible to write the full probability distribution

$$Q\left(\mathbf{g}, \mathbf{t}, \left\{t_{ij}\right\}, \mathbf{x}^0\right) = \frac{1}{Z} \prod_{(i,j)} \phi_{ij} \prod_i \psi_i \mathbb{G}_i \gamma_i \zeta_i^0 \tag{2.29}$$

which has a factor graph representation (shown in Figure 2.4) that has the same topology of the original contact graph.

2.3.2 Bayesian inference from observations

In order to account for observations, it is necessary to define a probability of making a certain observation given the full state of the epidemic $\underline{\mathbf{x}}$. It's natural to assume observations are independent, so for each observation at time t of node i, there will be a probability factor

$$p\left(z_i^t \,|\, \underline{\mathbf{x}}\right) = \eta_i^t\left(z_i^t \,|\, x_i^t\right)$$

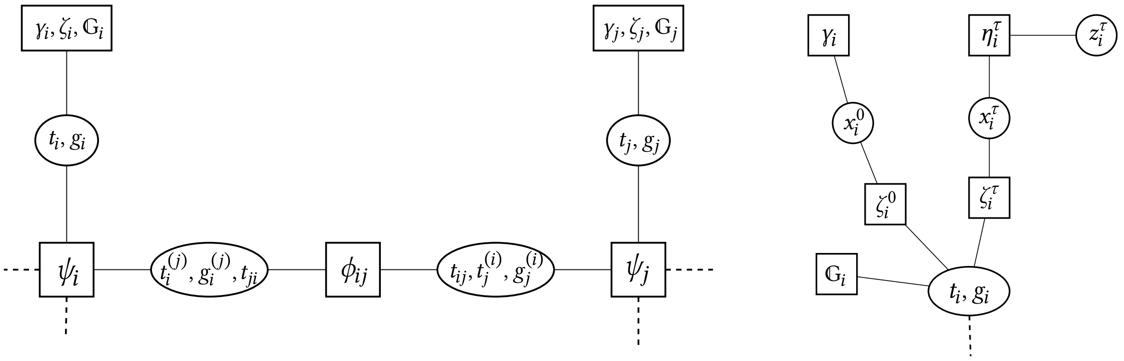

Figure 2.4: The new factor graph for the application of Belief Propagation. On the left, the part of the graph corresponding to contacts between i and j, related to factors ψ_i, ψ_j and ϕ_{ij}. The factors $\gamma_i, \zeta_i, \mathbb{G}_i$ are grouped on the left for simplicity: their structure is shown on the right pane. In this example, the state of individual i is observed at time τ, which corresponds to factors ζ_i^τ, for the conversion to the state x_i^τ, and η_i^τ, which accounts for the probability of observing z_i^τ.

that needs to be attached to the factor graph described so far. This is because if the inference of the state of individuals at a particular time (say, $t = 0$) is needed, the posterior probability given by Bayes' rule will be:

$$p\left(\mathbf{x}^0 \,\middle|\, \mathcal{O}\right) \propto \sum_{\mathbf{t},\mathbf{g},\{t_{ij}\}} Q\left(\mathbf{g}, \mathbf{t}, \{t_{ij}\}, \mathbf{x}^0\right) \prod_{(j,\tau,\sigma)\in\mathcal{O}} \zeta_j^\tau\left(t_j, g_j, x_j^\tau\right) \eta_k^\tau\left(z_j^\tau = \sigma \,\middle|\, x_j^\tau\right) \tag{2.30}$$

This marginal can be calculated with belief propagation, which will converge to exact distribution if the contact graph of the epidemic cascade is a tree. This approach can be used for inference at any time instant, since it is possible to obtain marginals for t_i and g_i. Moreover, it is a very efficient approach, requiring a number of operations per update round which is polynomial in the number of time instants and the number of edges in the original contact graph. In particular, in this approach an optimized calculation of the messages results in a complexity of $O\left(TG^2|E|\right)$ operations per BP iteration on the whole factor graph, where G is the number of time instants that are considered for the recovery delay, and $|E|$ is the original number of edges in the original contact graph [38].

While writing the BP equations for this case is not the objective of this chapter, we remark here that, while the method shown used the parameters of the Markovian SIR model, it is also possible to use different parameters, redefining $\mathbb{G}_i$ and ω_{ij} to account for non-Markovian spreading. Moreover, the whole approach can be derived in the case of non-discrete time epidemics, as has been done in [47].

Chapter 3

Autoregressive Artificial Neural Networks

While the Belief Propagation approximation of the posterior distribution (1.24) is efficient in the calculation of the node marginals, the presence of loops in the contact graph structure can be detrimental to the convergence of the algorithm, and the estimations of BP can thus become quite inaccurate. This phenomenon could happen, for example, in dense and very structured contact networks, and calls for the search of alternative inference methods which could overcome such a limitation while maintaining comparable performances on sparse networks.

In this chapter we discuss a new method for inference that uses generative neural networks, specifically autoregressive neural networks (ANNs), to learn the posterior probability of an epidemic process and efficiently sample from it. In practice, the autoregressive neural network can generate realizations of the epidemic process according to the stochastic dynamical rules of the prior model but compatible with the evidence.

Deep autoregressive neural networks are used to generate samples according to a probability distribution learned from data, for instance for images [55], audio [51], text [13, 64] and protein sequences [153] generation tasks and, more generally, as a probability density estimator [28, 43, 54]. Autoregressive neural networks have recently been used to approximate the joint probability distributions of many (discrete) variables in statistical physics models [68], and applied in different physical contexts [106, 144, 98, 130]. Neural networks have already been applied to epidemic forecasting [101, 66, 78] but rarely to epidemic inference and reconstruction problems.

In this chapter, it will be shown how we used a deep autoregressive neural network architecture to efficiently sample from a posterior distribution composed of a prior, given by the epidemic propagation model (even though the parameters of such model can be contextually inferred), and from an evidence given by (time-scattered) observations of the state of a subset of individuals. The work and the results presented in this chapter have been published in [158].

3.1 Learning the posterior probability using autoregressive neural networks

Considering the posterior distribution of an epidemic cascade $p(\underline{\mathbf{x}} \mid \mathcal{O})$ defined first in eq. (1.24), under the hypothesis of independent observations it can be rewritten as

$$\begin{aligned} p(\underline{\mathbf{x}} \mid \mathcal{O}) &= \frac{1}{p(\mathcal{O})} p(\mathcal{O} \mid \underline{\mathbf{x}}) p(\underline{\mathbf{x}}) \\ &= \frac{1}{Z} \prod_{i=1}^{N} p(x_i^0) \prod_{t=1}^{T} p\left(x_i^t \mid \mathbf{x}_{\partial i}^{t-1}, x_i^{t-1}\right) \prod_{r \in \mathcal{O}} p_r(O_r \mid x_{i_r}^{t_r}) \\ &= \frac{1}{Z} \prod_i \Psi_i \left(\mathbf{x}_i, \underline{\mathbf{x}}_{\partial i}\right) \end{aligned} \tag{3.1}$$

highlighting the dependency of one individual's trajectory $\mathbf{x}_i$ on the ones of its neighbours $\underline{\mathbf{x}}_{\partial i}$. This suggests that, given a permutation $\boldsymbol{\pi} = \{\pi_1, \pi_2, \dots, \pi_N\}$ of the individuals of the system, which imposes a specific ordering to the variables $\{\mathbf{x}_i\}$, the probability of the realization $\underline{\mathbf{x}}$ can be written as the product of conditional probabilities (chain rule) in the form

$$p(\underline{\mathbf{x}} \mid \mathcal{O}) = \prod_{i=1}^{N} p(\mathbf{x}_i \mid \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-1}, \mathcal{O}) \tag{3.2}$$

where $\mathbf{x}_i = \{x_i^0, \dots, x_i^T\}$ and $\underline{\mathbf{x}}_{<i} = \{\mathbf{x}_j \mid \pi_j < \pi_i\}$ is the set of epidemic-state trajectories of individuals with label lower than i according to the given permutation π. The distribution (3.2) can be approximated by a trial distribution $q_{\boldsymbol{\theta}}(\underline{\mathbf{x}})$ with the same conditional structure

$$q_{\boldsymbol{\theta}}(\underline{\mathbf{x}}) = \prod_i q_i^{\theta_i}(\underline{x}_i \mid \underline{\mathbf{x}}_{<i}), \tag{3.3}$$

which can be interpreted as an autoregressive neural network depending on a set of parameters $\boldsymbol{\theta} = \{\theta_i(\mathcal{O})\}$. From the analytical expression of the posterior distribution $p(\underline{\mathbf{x}} \mid \mathcal{O})$ defined in Eq. (3.1), the operation of parameters learning can be performed using a variational approach proposed in [68], in which the (reversed) Kullback-Liebler (KL) divergence

$$D_{KL}(q_{\boldsymbol{\theta}} \parallel p) = \sum_{\underline{\mathbf{x}}} q_{\boldsymbol{\theta}}(\underline{\mathbf{x}}) \log \frac{q_{\boldsymbol{\theta}}(\underline{\mathbf{x}})}{p(\underline{\mathbf{x}} \mid \mathcal{O})} \tag{3.4}$$

is minimized with respect to the parameters $\boldsymbol{\theta}$ of the trial distribution $q_{\boldsymbol{\theta}}(\underline{\mathbf{x}})$. The minimization of the KL divergence can be performed using standard gradient descent algorithms. This thus defines the loss function used for the training, F_q

$$F_q \doteq D_{KL}(q_{\boldsymbol{\theta}} \parallel p) = \sum_{\underline{\mathbf{x}}} q_{\boldsymbol{\theta}}(\mathbf{x}) \left[-\log p(\underline{\mathbf{x}} \mid \mathcal{O}) + \log q_{\boldsymbol{\theta}}(\underline{\mathbf{x}})\right] \tag{3.5}$$

The computational bottleneck of these calculation in these equations is that the sum runs over all possible epidemic realizations, a set that grows exponentially large with the size of the system. This issue is avoided by exploiting the generative power of autoregressive neural networks by generating a large sample dataset of epidemic cascades with them and using this data for

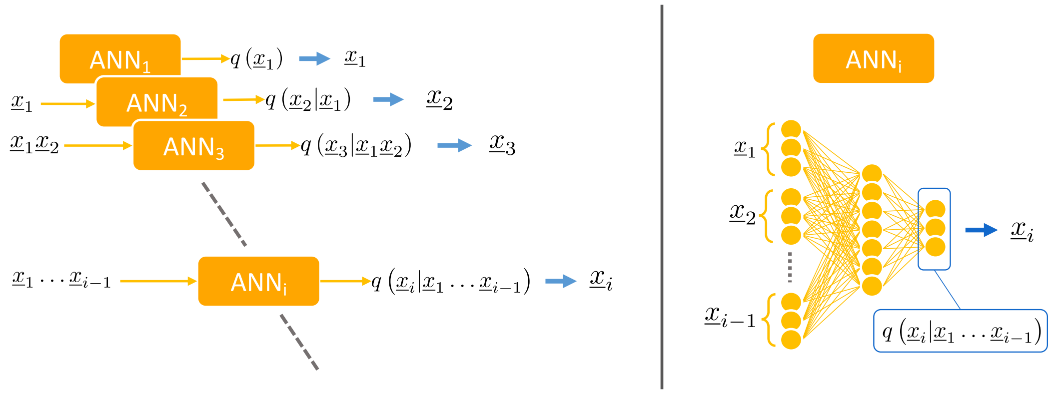

Figure 3.1: **Ancestral sampling of epidemic cascades. Left.** Ancestral sampling of epidemic cascades using artificial neural networks. For each individual i there is a neural network ANN_i that computes the probability $q\left(\underline{x}_i \mid \underline{x}_{i-1} \dots \underline{x}_1\right)$ of its time trajectory $\underline{x}_i$ given the time trajectory of previous individuals. The time trajectory $\underline{x}_i$ is extracted from the conditional probability $q\left(\underline{x}_i \mid \underline{x}_{i-1} \dots \underline{x}_1\right)$ and passed to the following neural networks. **Right.** Each neural network is composed of several fully connected layers (see section 3.3 for details).

training. This allows the averages over the autoregressive probability distribution to be approximated as a sum over a large number of independent samples extracted from the autoregressive probability distribution q_θ. During the sampling, the conditional structure of the distribution q_θ allows to use ancestral sampling [50] to gather the samples for each individual, as shown in figure 3.1. In section 3.4 it is described in detail how the learning is effectively performed by minimizing the Kullback-Leibler divergence, and how an annealing procedure needs to be introduced in order to avoid convergence issues due to very large values of the KL divergence (3.4).

A common way to represent the conditional probabilities in Eq.(3.3) is by means of feed-forward deep neural networks with sharing schemes architectures [54, 52] to reduce the number of parameters. Due to the possible high variability in the dependence of $p(\underline{x}_i \mid \underline{\mathbf{x}}_{<i})$ on $\underline{\mathbf{x}}_{<i}$ [130], instead of adopting a sharing parameters scheme we reduce the number of parameters by limiting the dependency of the conditional probability to a subset of $\underline{\mathbf{x}}_{<i}$. This subset that we consider is formed by all $\mathbf{x}_j \in \underline{\mathbf{x}}_{<i}$ such that j is at most a second-order neighbour of i in the contact network, that is the one in which there is an edge between two individuals if and only if they are in contact at least once during the epidemic cascade. This is because of the following reasoning: while the permutation order $\boldsymbol{\pi}$ of the variables generally influences the approximation, in the case of acyclic graphs, it is possible to define an order by which the aforementioned second-order neighbours' approximation is exact: the variables are ordered according to a spanning tree computed starting from a random node chosen as a root (the proof is given in appendix A). We thus imagine that the same procedure yields good approximations for sparse interacting networks, but for general interaction graphs, we are unaware of arguments

for choosing an order with respect to another. Therefore, also for the non-acyclic contact networks that are used throughout for the results, the same *next nearest neighbours* approximation is used, using random permutations of the nodes. This introduces some losses in the accuracy of the ANN technique, which are investigated in section 3.9.

3.2 Inferring the parameters of the propagation model

In a real case scenario, the epidemic parameters governing the propagation model are usually unknown and they should be inferred from the available data. Calling Λ the set of these parameters (e.g. for uniform SIR models $\Lambda = (\lambda, \mu)$), the goal is to estimate them by computing the values Λ^* that maximize the likelihood function given the set of observations $\mathcal{O}$, i.e.

$$\begin{aligned} p(\mathcal{O} \mid \Lambda) &= \sum_{\underline{\mathbf{x}}} p(\mathcal{O} \mid \underline{\mathbf{x}}) p(\underline{\mathbf{x}} \mid \Lambda) && (3.6) \\ &= \sum_{\underline{\mathbf{x}}} \prod_i \Psi_i(\mathbf{x}_i) && (3.7) \\ &= Z(\Lambda)\,. && (3.8) \end{aligned}$$

The quantity Z is the same normalization constant introduced in Eq. (3.1), where the dependence on the parameters was dropped. Formally,

$$\Lambda^* = \arg\max_{\Lambda} Z(\Lambda) = \arg\max_{\Lambda} \log Z(\Lambda)\,. \tag{3.9}$$

Recalling that $P(\underline{\mathbf{x}} \mid \mathcal{O}) = Z^{-1} \prod_i \Psi_i(\mathbf{x}_i, \Lambda)$ and thanks to Gibbs' inequality we have that

$$\begin{aligned} \log Z(\Lambda) &= \sum_{\underline{\mathbf{x}}} p(\underline{\mathbf{x}} \mid \mathcal{O}) \log \prod_i \Psi_i(\mathbf{x}_i, \Lambda) - \sum_{\underline{\mathbf{x}}} p(\underline{\mathbf{x}} \mid \mathcal{O}) \log p(\underline{\mathbf{x}} \mid \mathcal{O}) && (3.10) \\ &\geq \sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \log \prod_i \Psi_i(\mathbf{x}_i, \Lambda) - \sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \log q_\theta(\underline{\mathbf{x}}) && (3.11) \\ &= -\langle H \rangle_q + S_q && (3.12) \end{aligned}$$

where we first replaced the probability function $P(\underline{\mathbf{x}} \mid \mathcal{O})$ with the variational probability distribution $q_\theta(\underline{\mathbf{x}})$ and defined the energetic and entropic terms

$$\begin{aligned} \langle H \rangle_q &= -\sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \log \prod_i \Psi_i\left(\mathbf{x}_i, \Lambda\right) && (3.13) \\ S_q &= -\sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \log q_\theta(\underline{\mathbf{x}}). && (3.14) \end{aligned}$$

Since S_q does not depend on Λ, minimizing $\langle H \rangle_q$ with respect to parameters Λ corresponds to maximizing $\log Z(\Lambda)$. The quantity $\langle H \rangle_q$ and its derivatives w.r.t. Λ can be computed efficiently, in an approximate way, by replacing the sum over all configurations with the average on the samples extracted by ancestral sampling from the autoregressive probability distribution q_θ. Therefore, we use the following heuristic procedure, inspired by the Expectation-Maximization (EM) algorithm, to infer the parameters, while minimizing the KL divergence between q_θ and the posterior probability $p(\underline{\mathbf{x}} \mid \mathcal{O})$. During the learning process, two sequential steps are performed:

1. Update the parameters $\{\theta_i\}$ of the autoregressive neural network to minimize the KL divergence in Eq. (3.4).
2. Update the parameters Λ to maximize the quantity $\langle H \rangle_q$.

These steps are repeated until the end of the learning process. This approach has been inspired by [38], where a similar method is applied to the Belief Propagation based inference algorithm. As shown in that work, using the above heuristic of performing one step in the minimization of the KL divergence and one step in the maximization of (3.13) yields good results in the inference of parameters (shown in section 3.7.4).

3.3 Neural network architecture

Here, the structure used in the proposed method is discussed in more detail. In the autoregressive approach, each conditional probability function $p_i(\mathbf{x}_i \mid \underline{\mathbf{x}}_{<i})$ has to be approximated with a neural network (NN) $q_i^{\theta_i}\left(\mathbf{x}_i | \underline{\mathbf{x}}_{<i}\right)$. However, for monotonic models such as the SIR epidemic model, the state-space representation of temporal trajectories as a sequence $\underline{x}_i \in \mathcal{X}^{T+1}$ of $T+1$ individual states ($\mathcal{X} = \{S, I, R\}$) is redundant, and it turns out to be more efficient to represent them by the time instant $t_i^I \in \mathcal{T}$ at which individual i becomes infected and time instant $t_i^R \in \mathcal{T}$ at which it recovers, where $\mathcal{T} = \{0, 1, \dots, T\} \cup T_\infty$, the time T_∞ corresponding to the case of no infection or recovery occurred for i in the interval $[0, T]$. More precisely, the case $\left(t_i^I = T_\infty,\, t_i^R = T_\infty\right)$ corresponds to individual i being susceptible at every time of the epidemic process, and $\left(t_i^I \neq T_\infty,\, t_i^R = T_\infty\right)$ corresponds to individual i being infected but not recovering in the time interval $[0, T]$. In practice, we can set $T_\infty = T + 1$ without loss of generality. This parametrization reduces the state space that is explored during the training, decreasing the number of incorrect time trajectories generated (for instances, time trajectory with transitions from state I to S).

In order to compute the probability of a trajectory of a single individual, there are now two variables, t_i^I and t_i^R. For this work, two neural networks are employed to encode the probabilities, one for each variable. This is chosen to reduce the number of parameters needed in total for the ANN method, and for ease of construction. These neural networks receive as input the (already sampled) infection and recover times of previous individuals k with $\pi_k < \pi_i$, while for the recovery probability NN, there are also the samples of the infection time for the same individual. This way, the probability distribution of the trajectory of a single individual is split into two conditional ones:

$$q_i^{\theta_i}\left(t_i^I, t_i^R \,\middle|\, \left\{t_j^I, t_j^R\right\}_{\pi_j<\pi_i}\right) = q_i^{(I)}\left(t_i^I \,\middle|\, \left\{t_k^I, t_k^R\right\}_{\pi_k<\pi_i}\right) q_i^{(R)}\left(t_i^R \,\middle|\, t_i^I, \left\{t_k^I, t_k^R\right\}_{\pi_k<\pi_i}\right)$$

and $\theta_{i,I}$ and $\theta_{i,R}$ are the respective weights of each network.

In our implementation, each neural network is a multi-layer perceptron (MLP) composed of three hidden layers plus one output layer; each layer is fully connected, and can be written as $\mathbf{L}_{k+1} = \sigma\left(\overline{\mathbf{W}}_k \mathbf{L}_k + \mathbf{b}_k\right)$ where $\mathbf{L}_k$ is the input vector (output of layer k), $\overline{\mathbf{W}}_k \in \theta$ is the matrix of the weights, $\mathbf{b}_k \in \theta$ is the bias vector, and σ is the activation function, which is *ReLU* for the hidden internal layers and *Softmax* for the last layer, which is needed to get a normalized probability distribution. The width of each layer (the number of neurons) varies linearly from the input size of each network to the output size. The input trajectories for each neural network are

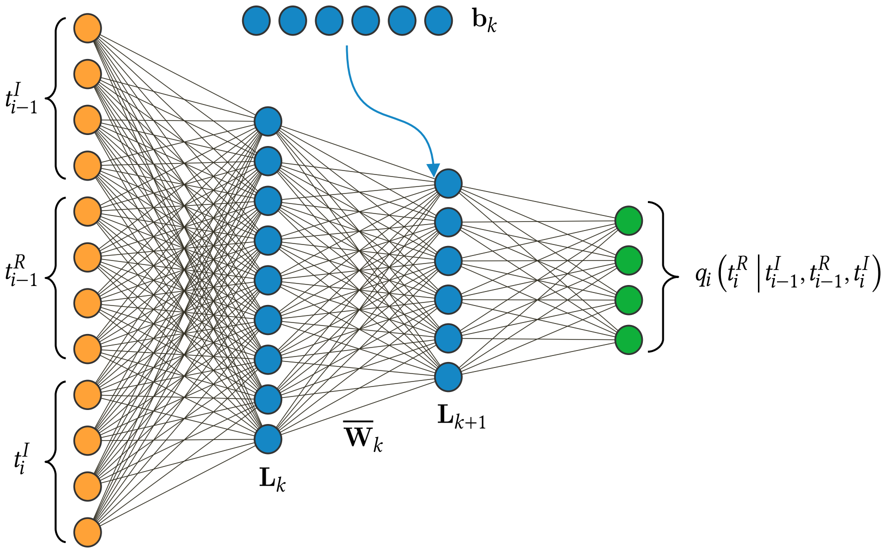

Figure 3.2: **Actual implementation of a neural network.** An example showing the layers of a neural network used for computing the conditional probability. The hidden layers, represented in blue, are two, and the output layer, in green, is used to output the probability distribution for t_i^R. Hidden layers and output layer have different nonlinear functions. In this example, $T = 2$, so that the time instants are $0, 1, 2$ and $T_\infty = 3$.

converted into the *one hot* encoding representation: in this scheme, each infection or recovery time of a sample is transformed into a vector of all zeros, except for that index which is 1. For example, to represent $t_i^I = 6$, a vector of length $T + 2$ is created where only the 7th element is 1 and the rest are zeros.

The size of the input and outputs of the neural networks used for each individual depends on the observations and the contact graph in the following way: a) the input size depends on the number of individuals with lower index that are considered (depending on the approximation made, see section 3.9) and b) the observations made on an individual restrict the phase space of the possible instants of infection (for example, if we observe individual i is in the infected state at time t_O, this means that his/her infection time $t_i^I \leq t_O$ and the recovery time $t_i^R > t_O$). This last condition affects the size of the output of each network, but also the input size of the networks for the individuals with higher index, whose conditional neural networks depends on it: for example, if individual i is observed in state S at the final time T, when the observations are noiseless, this implies that the infection and recovery times are fixed, $t_i^I = T + 1$ and $t_i^R = T + 1$, and the network will always give the same value for them. Thus, in this case, for all individuals j with $\pi_j > \pi_i$, the dependence of $q_j^{(I)}$ and $q_j^{(R)}$ on i will effectively disappear.

3.4 Learning procedures and regularization

The learning procedure involves the minimization of the Kullback-Leibler divergence, eq. (3.4), between the posterior probability and the variational autoregressive neural networks. Starting from equation (3.1), the expression of the posterior can be also factorized as a product over the time instants as well as the individuals,

$$p\left(\underline{\mathbf{x}} \mid \mathcal{O}\right) \propto \prod_{i} \Psi_i\left(\mathbf{x}_i, \underline{\mathbf{x}}_{\partial i}\right) = \prod_{i,t} \psi_i^t\left(x_i^t, \mathbf{x}_{\partial i}^{t-1}\right)$$

where

$$\psi_i^t\left(x_i^t, \mathbf{x}_{\partial i}^{t-1}\right) = \begin{cases} p_i^t\left(x_i^{(t)} | \mathbf{x}_{\partial i}^{t-1}\right) \prod_{(i_r,t_r)=(i,t)} p_r(O_r^{t_r} \mid x_i^t) & t > 0 \\ p_i(x_i^0) \prod_{(i_r,t_r)=(i,t)} p_r(O_r^{t_r} \mid x_i^t) & t = 0 \end{cases} \tag{3.15}$$

Since the factor ψ_i^t can be zero for some epidemic configuration $\underline{\mathbf{x}}$, for example when they are forbidden by dynamical constraints, the gradient of the training loss (3.5) can diverge preventing the application of gradient descent algorithms. In these cases, the diverging term $\log \psi_i^t = -\infty$ is replaced with a regularization term $\log \epsilon$ with $\epsilon \ll 1$. The gradient of the KL divergence with respect to the parameters of the trial distribution reads

$$\nabla_{\theta} F_q = \nabla_{\theta} \sum_{\underline{x}} q_{\theta}\left(\underline{\mathbf{x}}\right)\left[-\log p\left(\underline{\mathbf{x}} \mid \mathcal{O}\right) + \log q_{\theta}\left(\underline{\mathbf{x}}\right)\right] \tag{3.16}$$

$$= \sum_{\underline{x}} \nabla_{\theta} q_{\theta}\left(\underline{\mathbf{x}}\right)\left[-\log\left(\prod_{i} \Psi_i\left(\underline{x}_i, \underline{\mathbf{x}}_{\partial i}\right)\right) + \log q_{\theta}\left(\underline{\mathbf{x}}\right)\right] + \sum_{\underline{x}} q_{\theta}\left(\underline{\mathbf{x}}\right) \nabla_{\theta} \log q_{\theta}\left(\underline{\mathbf{x}}\right) \tag{3.17}$$

$$= \sum_{\underline{x}} q_{\theta}\left(\underline{\mathbf{x}}\right)\left[-\log\left(\prod_{i} \Psi_i\left(\underline{x}_i, \underline{\mathbf{x}}_{\partial i}\right)\right) + \log q_{\theta}\left(\underline{\mathbf{x}}\right)\right] \nabla_{\theta} \log q_{\theta}\left(\underline{\mathbf{x}}\right), \tag{3.18}$$

where we used that

$$\sum_{\mathbf{x}} q_{\theta}\left(\underline{\mathbf{x}}\right) \nabla_{\theta} \log q_{\theta}\left(\underline{\mathbf{x}}\right) = \sum_{\underline{x}} \nabla_{\theta} q_{\theta}(\underline{\mathbf{x}}) = 0$$

and $\sum_{\mathbf{x}} \nabla_{\theta} q_{\theta}(\underline{\mathbf{x}}) = 0$ due to the normalization $\sum_{\underline{x}} q_{\theta}(\underline{\mathbf{x}}) = 1$. The presence of large negatives values in the derivatives (e.g. due to $\log \epsilon$ terms) reduces the ability of gradient descent algorithms to explore all configurations compatible with the constraints. To overcome this issue, an annealing procedure is adopted, in which a fictitious inverse temperature β is introduced in the computation the gradient of the KL divergence. Also, as already mentioned above, we replace the sum over all configuration with the average on a large number of epidemic cascades sampled from the ANN:

$$\nabla_{\theta} F_q^{\beta} \approx \mathbb{E}_{\underline{\mathbf{x}} \sim q_{\theta}}\left[\left(-\beta \log\left(\prod_{i} \Psi_i\left(\underline{x}_i, \underline{\mathbf{x}}_{\partial i}\right)\right) + \log q_{\theta}\left(\underline{\mathbf{x}}\right)\right) \nabla_{\theta} \log q_{\theta}\left(\underline{\mathbf{x}}\right)\right]. \tag{3.19}$$

The minimization procedure starts with $\beta = 0$, where all the configurations are allowed with uniform probability, then the parameter β is slowly increased until it reaches $\beta = 1$, at which the original expression of the loss function is recovered.

For all the results shown in the following, at each step of the training we generate 10 000 epidemic cascades from the ANN before updating the weights using the ADAM optimizer [57] with learning rate $lr = 0.001$. At each step β increases linearly, and we perform 10 000 training steps(except for the **hospital** case (see section 3.7) where we do 20 000 steps).

In the case of the risk inference problem, we also include a prior probability in the model, with strength decreasing linearly with β. This prior is computed from the probabilities of each individual being sampled as S, I or R at the last time instant $t = T$, and it is designed in such a way that, when $\beta = 0$, the probabilities for the three states at the final time are equal.

The running time of the ANN algorithm on a single instance ranges from one hour to two days on a single GPU (Nvidia TITAN RTX), depending on the structure of the observations $\mathcal{O}$ and that of the underlying contact network. Our implementation using the PyTorch framework [65] is publicly available in the repository [156]. The results presented in this chapter can be reproduced using the code published in the repository [157].

3.5 Soft margin estimator

For the results we will also use the Soft Margin estimator from [41] for finding the patient zero. This is a Monte Carlo based estimator, which applies Bayes formula to estimate the source posterior probability $P(s = k \mid \mathbf{x}_O)$ given a set of observations $\mathbf{x}_O$ on an epidemic cascade:

$$P(s = k \mid \mathbf{x}_O) \propto P(\mathbf{x}_O \mid s = k)\, P(s = k) \tag{3.20}$$

In order to evaluate the likelihood $P(\mathbf{x}_O \mid s = k)$, the method uses M Monte Carlo simulations in which the source of the epidemic is k and compares the resulting epidemic cascade $\mathbf{x}_i$ with the observations.

Since the original method in [41] was developed for the SI epidemic model, we devised an extension for the likelihood calculation to the SIR model. In order to compute $P(\mathbf{x}_O \mid s = k)$, a measure of the similarity to the observation is needed. For this we use the Jaccard similarity function $\phi_X(\mathbf{x}_i, \mathbf{x}_O)$, relating which individuals are in state X (either infected, $X = I$ or recovered, $X = R$) in the generated cascade and the observations:

$$\phi_X(\mathbf{x}_i, \mathbf{x}_O) = \frac{|v_X(\mathbf{x}_i) \cap v_X(\mathbf{x}_O)|}{|v_X(\mathbf{x}_i) \cup v_X(\mathbf{x}_O)|} \tag{3.21}$$

Defining $v_X(\mathbf{x}) = \left\{(i, t) \in \mathcal{O} \,:\, x_i^t = X\right\}$.

Clearly, if no individuals are observed in state X, then $\phi_X = 1$ regardless of the realization. Then, we can define an estimator L of the likelihood of a given source:

$$L(\mathbf{x}_O \mid s = k) = \frac{1}{M} \sum_{i=1}^{M} \exp\left\{-\frac{(\phi_I(\mathbf{x}_i, \mathbf{x}_O) - 1)^2 + (\phi_R(\mathbf{x}_i, \mathbf{x}_O) - 1)^2}{a^2}\right\} \tag{3.22}$$

where the coefficient a regulates the sharpness of the peak around the case of perfect matching of observations ($\phi_I = 1$ and $\phi_R = 1$), where the corresponding value in the sum is 1. In particular, when $a \to 0$ the exponential term is either 1 in the case that is just described, or 0. Thus, the estimator L reduces to the exact Monte Carlo likelihood probability:

$$L(\mathbf{x}_O \mid k) \underset{a \to 0}{\longrightarrow} P(\mathbf{x}_O \mid s = k) = \frac{1}{M} \sum_{i=1}^{M} \delta(\phi_I(\mathbf{x}_i, \mathbf{x}_O) - 1)\, \delta(\phi_R(\mathbf{x}_i, \mathbf{x}_O) - 1)$$

which is much more difficult to estimate because a very large number of generated epidemics give a zero contribution to the sum, even if the result is very close to the observations, and getting the same exact observations has vanishing probability. For the patient zero problems, we select the a values that give better results (retrospectively) because they seem to be strongly dependent on the contact graphs and the number of samples considered.

3.6 Contact graphs employed

Before showing the results of the ANN technique, we briefly discuss the contact graphs to which this method is applied. We use both real-world and synthetic contact networks for the generation of epidemic cascades to test our method on. For the former type, we consider one network, taken from the dataset *InVS13* [61], that has been collected in a work environment (which we call **work**), and a second one that has been collected in a hospital (**hospital**) [45]. In both cases, the dataset used is the temporal list of contacts, respectively between 95 and 330 individuals, for a period of two weeks. Since the real duration $\delta_{i,j}^t$ of each contact is known, the probability of infections between individuals i, j at time t is computed as

$$\lambda_{i,j}^t = 1 - e^{-\gamma \delta_{i,j}^t}$$

where γ is the rate of infection. The obtained networks have been analyzed, and in Figure 3.3 the obtained metrics of the graphs are shown, like the fraction of active nodes, the density and the average clustering coefficient [16].

We also consider two synthetic contact networks: a random regular graph (**RRG**) with $N = 100$ individuals and degree equal to 10, and a random geometric graph (**proximity**), in which $N = 100$ individuals are randomly placed on a square of linear size $\sqrt{N}$. In the latter, the probability that individuals i and j are in contact is $e^{-d_{ij}/l}$, where d_{ij} is the distance between i and j and l is a parameter (set to $l = 10$) that controls the density of contacts.

3.7 Results

As a preliminary illustration of the ability of the proposed Autoregressive Neural Network (ANN) to sample epidemic realizations from a given posterior distribution, we consider epidemic cascades generated on the **hospital** contact graph. These are depicted in figure 3.4, where it can be seen that, even when starting from the same individual, different epidemic cascades can have very different outcomes. When we train the ANN to learn the posterior probability composed by the prior, i.e. the epidemic model that generates the cascade, and the final configuration (at day 12) of one such epidemic (the blue one, in figure 3.4), we can then sample more cascades from the ANN and compare the difference between them and the original one. This is shown in Figure 3.5: the epidemic cascades generated by the ANN have Hamming distances from the reference one that reduce to zero at day 12 (central-bottom plot) and a fraction of them have prior probabilities larger than the probability of the (blue) epidemic cascade taken as reference (right-bottom plot in Fig. 3.5) .This example suggests that the ANN approach can generate epidemic cascades compatible with the observations and sampled according the prior epidemic model.

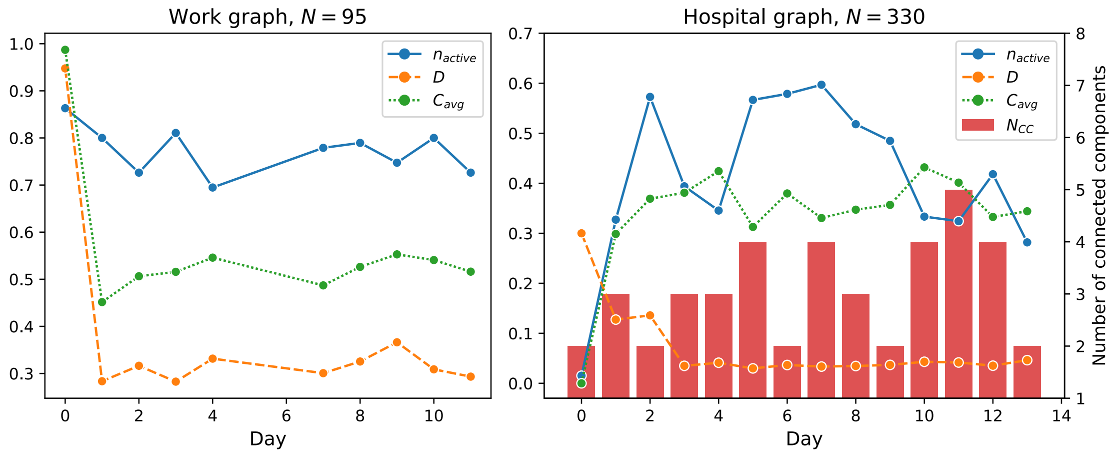

Figure 3.3: **Characteristics of the real contact graphs:** here some metrics of the two contact graphs employed in the results, **work** and **hospital**, are shown. For each day, the metrics are: the fraction of the nodes that are active n_{active} (inactive nodes have no contacts), the density D and the average clustering coefficient C_{avg} for the subgraphs made by the active nodes. For the **hospital** graph, the number of connected components N_{CC} is also shown, while for the **work** graph, there is always one connected component (except for days 5 and 6 when there are no contacts).

In the following, we exploit the ability of the ANN to sample epidemic cascades from a posterior distribution to tackle three challenging epidemic inference problems: *i)* the *patient-zero* problem, in which the unique source of a partially observed epidemic outbreak has to be identified, *ii)* the *risk assessment* problem, in which the epidemic risk of each individual has to be estimated from partial information during the evolution of the epidemic process, and *iii)* the *inference of the epidemic parameters*. The results are described in the following sections, where the ANN performance in these tasks is compared with already existing methods in the field of epidemic inference. We also evaluate how the efficiency of the ANN algorithm depends on the size of the epidemic outbreak, measuring the number of generated epidemic samples necessary to obtain nearly optimal results.

Comparison with other techniques We compare our technique, Autoregressive Neural Network (ANN), against state of the art approaches for the problems that we consider, on both random graphs and real-world contact networks. The Soft Margin (SM) estimator previously described is applicable only to the *patient zero* problem. The Belief Propagation approach [37, 38], implemented in the SIB software [118, 151] provides exact inference on acyclic contact networks and performs very effectively on sparse network structures. For the *risk assessment* problem, we also compare ANN with the Simple Mean Field (SMF) and Contact Tracing (CT) algorithms, that have been described in [118]. For all the comparisons we run, we generate SIR epidemic cascades on the chosen contact graphs with a duration of $T = 15$ days, with Markovian parameters.

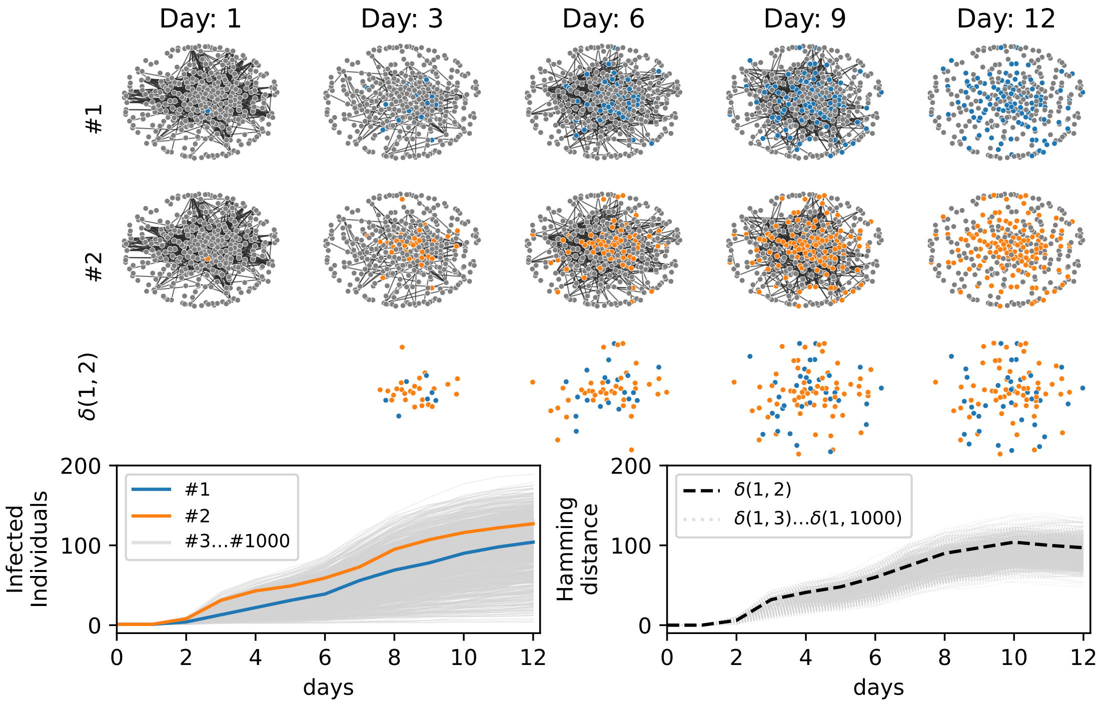

Figure 3.4: **Simulated epidemic cascades in the hospital contact network.** One thousand epidemic cascades simulated (with the same epidemic parameters) on a real contact graph measured in a hospital [45] (see section 3.6). The epidemics started from the same individual. Two samples (blue, and orange) of epidemic cascades are shown in the first and second rows of the figure. The third row represents the distance between them, where in this case the blue dots are the infected individuals present in the cascade 1 but not in cascade 2 and the orange ones are those present in cascade 2 but not in cascade 1. In the third row, the total number of blue and orange dots gives the Hamming distance between the two daily configurations. **Left-bottom plot.** Cumulative number of infected individuals for 1000 epidemic cascades started from the same individual. **Right-bottom plot.** Hamming distance ($\delta(1, i)$) between the cascade 1 and all the others $i \in [2,3 \dots 1000]$. Figure first shown in [158].

Because the final number of infected individuals in the chosen contact graphs fluctuates, we choose the parameters of the epidemic model for the results in such a way to have, on average, half of the individuals infected at the end of the epidemic propagation. This is done to reduce the cases where very few or a large fraction of individuals are infected, as, in these cases, the inference problems analysed become either too trivial or very hard to solve because of lack of information.

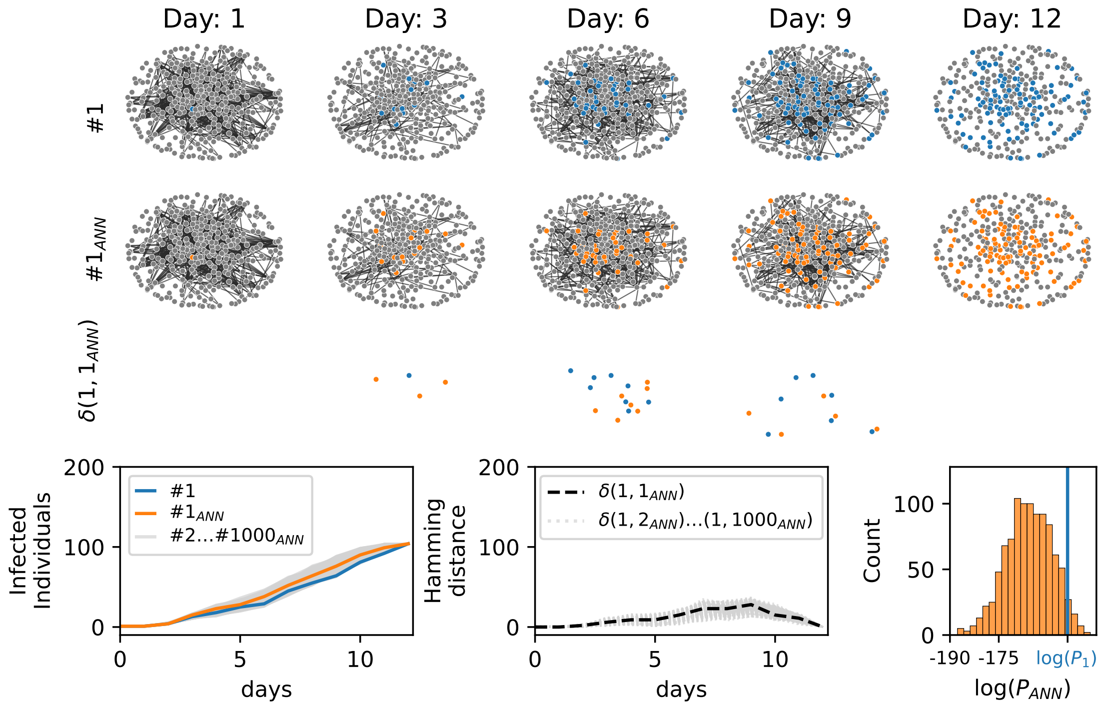

Figure 3.5: **Epidemics cascades generated by the ANN.** Epidemic cascades generated by the ANN trained on a posterior probability composed by a prior, the epidemic model that generate the blue cascade, and the evidence, its final configuration at day 12. The contact network is the **hospital** graph (see section 3.6). An example of the epidemic cascade generated is shown in second row (1_{ANN}, orange). The third row represents the Hamming distance between them, showing individuals infected in the cascade 1 but not in 1_{ANN} in blue, and the reverse (infected in cascade 1_{ANN} but not in 1) are shown in orange. **Left-bottom plot.** Cumulative number of infected individuals for epidemic cascade simulated (blue curve) and generated by the ANN ($i_{ANN} \in [1,2 \ldots 1000]$. **Central-bottom plot.** Hamming distance ($\delta(1, i_{ANN})$ between blue epidemic cascade and those generated by the ANN $i_{ANN} \in [1,2, \ldots 1000]$. **Right-bottom plot.** Distribution of the values of the prior probability of the generated epidemic cascades (P_{ANN}). The blue vertical line is the value of the prior probability of the blue cascade ($\log(P_i)$). Figure first shown in [158].

3.7.1 The patient zero problem

Given the exact knowledge of the final state of the epidemics at time T, the patient-zero problem consists in identifying the (possibly unique) source of the epidemics. In a Bayesian framework, this problem can be tackled by computing for each individual the marginal probability of being infected at time $t = 0$ given a set of observations $\mathcal{O}$. This quantity can be estimated from the posterior distribution (3.1) with all three algorithms (ANN, SIB, and SM) considered in this work. For each contact network (**RRG, proximity, work, hospital**), we considered 100 different realizations of the epidemic model with only one patient zero. The three algorithms were used to rank infected and recovered individuals in decreasing order according to the estimated probability of being infected at time zero for each epidemic realization. Fig. 3.6 displays, for each algorithm, the fraction of times, in 100 different realizations, the patient-zero is correctly identified. The left plots show the fraction of times it is correctly identified at the first position of the infected or recovered individuals ranked according to the algorithms. The right plots show the fraction of times the patient zero is found versus the fraction of infected or recovered individuals ranked by the algorithms considered. The ANN algorithm outperforms all the other methods as indicated by the larger area under the curve (AUC) obtained in all cases considered. The improvement is also evident when analyzing the fraction of patient zero correctly identified by each algorithm (left bar plots in Fig. 3.6). For example, in the hospital case, ANN correctly identifies the patient zero in the 74% of the instances, SIB in the 54% and SM in the 35% of them. In all cases, the ANN algorithm's performances are comparable to or better than those of the other approaches. The results on the patient zero problem reveal the ability of the ANN algorithm to efficiently generate epidemic cascades according to the posterior probability

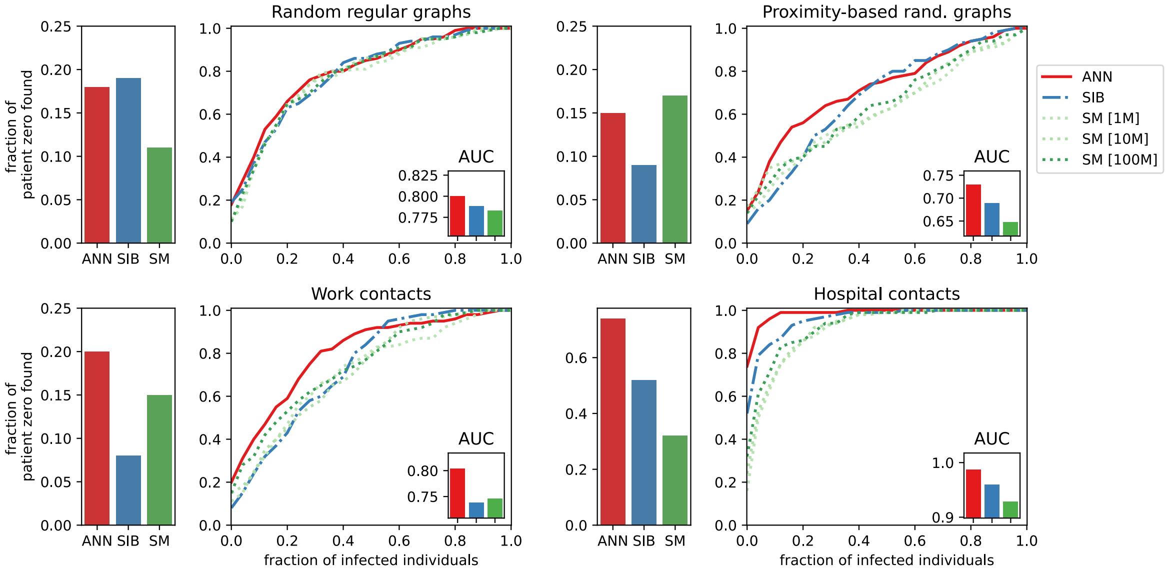

Figure 3.6: **Results of the patient zero problem.** The left bar plots, for each case, represent the fraction of times, in 100 different epidemic cascades, the patient zero is correctly identified at the first position of the ranking given by the algorithms. The right plots show the fraction of times the patient-zero is found (in 100 different epidemic cascades) in a fraction of infected or recovered individuals ranked according to the probability to be patient zero given by the three algorithms ANN, SIB, and SM (the values of the area under the curve [AUC] are shown in the insets). For the **RRG** we consider the following epidemic parameters $\lambda = 0.04$ and $\mu = 0.02$ and for **proximity** $\lambda = 0.03$, $\mu = 0.02$. In the case of the **work** graph, the parameters are $\gamma = 10^{-3}$, $\mu = 0.02$ and in the **hospital** case they are $\gamma = 2 \cdot 10^{-4}$, $\mu = 0.02$.

defined in Eq. 3.1.

3.7.2 Scaling properties with the size of the epidemic outbreak

From the results presented in the previous subsection, Autoregressive Neural Networks seems to be very effective in tackling classical epidemic inference problems, particularly on dense contact networks, where the performances of BP-based methods are expected to decrease. It is, however, critical to check how the convergence property of the learning processes scales with the size of the epidemic outbreak. For this analysis, we consider the patient zero problem on a tree contact network with a unique epidemic source and where the state of the system at the final time T is fully observed. With this choice, we ensure that the probability marginals computed by the SIB algorithm are exact; hence they can be taken as a reference to compare the performances of the other algorithms. The ANN algorithm with a second-order neighbours approximation is exact when the interaction graph is acyclic (see supplementary material for details), assuming that the architecture of the neural networks used is sufficiently expressive to capture the complexity of posterior probability. On the other hand, since the SM algorithm is based on a Monte Carlo technique, it can give estimates of marginal probabilities with arbitrary accuracy when a sufficiently large sample of epidemic cascades is generated.

In the case of complete observation of the final state, the larger the epidemic size (i.e., the total number of infected individuals at time T), the larger is the number of epidemic cascades that are compatible with the observation. For instance, in an epidemic realization of duration T time steps in which n_I individuals are tested infected and $N - n_I$ tested susceptible at time T, the number of epidemic configurations compatible with the observations scales as T^{n_I}. Both ANN and SM rely on sampling procedures, so their performances could suffer from convergence issues when the epidemic size (n_I) increases. We compute the total number of samples generated by the ANN during the learning process and the number of samples of epidemic cascades generated by SM in the Monte Carlo procedure. In both cases, we assume that convergence is reached when $\sum_i \left| P_{\text{algo}}(x_i^0 = I \mid \mathcal{O}) - P_{\text{sib}}(x_i^0 = I \mid \mathcal{O}) \right| < 0.1$ where *algo* is either ANN or SM, and $P_{\text{algo}}(x_i^0 = I \mid \mathcal{O})$ is the estimated marginal probability that individual i is infected at the initial time according to each method. The results on the scaling properties of the ANN and SM as a function of the epidemic size on a tree contact network with degree and depth both equal to 6 (**tree**) are shown in Figure 3.7. Here we set the duration of the epidemic cascades to $T = 15$ days. The ANN algorithm has a quasi-linear dependence with the epidemic size; conversely, the SM algorithm exhibits a very sharp increase in the number of simulations necessary for good estimates of the marginals, and already for epidemic sizes of order ten individuals, good estimates are difficult to obtain.

3.7.3 Epidemic risk assessment

The risk assessment problem consists in finding the individuals who have the highest probability of being infected at a specific time given a partial observation $\mathcal{O}$. For this particular task, we only consider epidemic realizations of the SIR model with $\mu = 0$ (i.e. only the states S and I are available) where half of the infected individuals are observed (at random) at the final time T.

The results of the risk-assessment analysis obtained by means of the ANN algorithm are

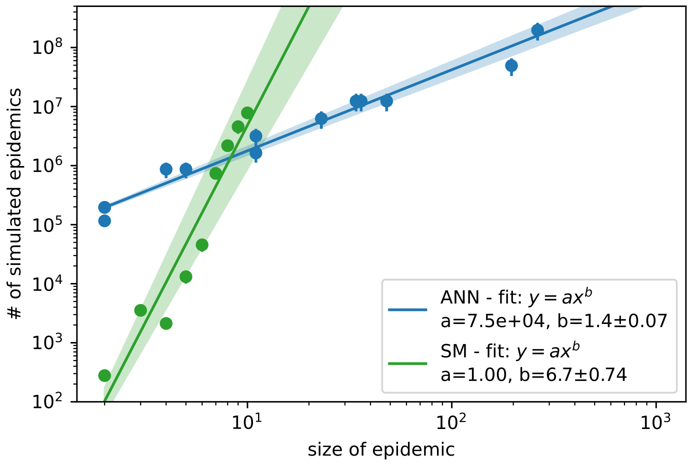

Figure 3.7: **Scaling properties with the size of epidemic cascades.** Number of samples generated by the ANN and SM algorithms to reach convergence. We consider the estimation of the marginal probabilities to be infected at initial time with interactions graphs given by a tree of degree and depth both equal to 6 and spanning 15 days of duration. The epidemic cascade are generated with $\mu = 0$ and different values of λ ($\lambda \in [0.1, 0.6]$). For the ANN algorithm, we consider the number of samples generated during the training, that is 10^3 samples for each annealing step. For each instance, we run the annealing process with 2^n number of steps with $n \in \{5,6,\ldots,18\}$. Each point is a single instance, and if the algorithm converges between 2^{n-1} and 2^n steps, we report the midpoint between the two values, multiplied by the number of samples extracted at each step number, as the number of samples for ANN. For the SM algorithm, each point in the plot is the average number of simulated epidemics necessary to reach convergence to a good estimate of the marginals (worst 10% results were discarded). No point is reported when more than ten infected individuals are observed, because more than 10% of the instances did not converge within $2 \cdot 10^8$ simulated epidemics. Figure first shown in [158].

compared with those provided by the BP-based method and two other risk evaluation algorithms that are proposed in [118]. The Simple Mean Field (SMF) algorithm is based on a mean-field description of the epidemic process in which information about the observed individuals is heuristically included. A more detailed explanation of the method is discussed in chapter 4. The Contact Tracing (CT) algorithm, instead, is a form of contact tracing that computes the individual risks according to the number of contacts each individual has had with observed infected individuals in the last $\tau = 5$ time steps.

A measure of the ability to correctly identify the unobserved infected individuals at the final time T is represented by the area under the Receiving Operating Characteristic (ROC) curve. This quantity, averaged over 100 instances of the epidemic realizations, is shown, for the methods considered above, in figure 3.8, for different contact networks (**RRG**, **proximity** and **work**). All algorithms perform similarly on random graphs, whereas ANN and SIB outperform the other two methods in the case of the **work** contact network.

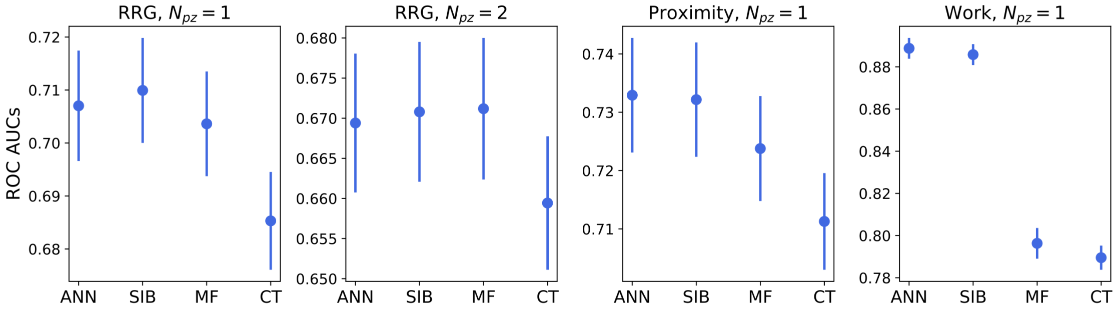

Figure 3.8: **Epidemic risk assessment results.** Area under the Receiving Operating Characteristic (ROC) curves for the risk assessment problem on random regular graphs (**RRG**, with both 1 and 2 patient zeros for the generation of test instances), on the **proximity** random graphs and **work** real-world contact network. The results are averaged over 100 different epidemic cascades generated with the same epidemic parameters. For each case, the ROC curve for the classification of the unobserved infected individuals at the final time is computed. In the **RRG** case, the epidemic parameters are $\lambda^{(1)}_{RRG} = 0.035$ for the single patient zero case and $\lambda^{(2)}_{RRG} = 0.03$ for the double patient zero case. For the **proximity** random graphs, $\lambda_{prox} = 0.03$. In the case of the **work** network, the model has rate of infection $\gamma_{\text{work}} = 10^{-3}$. In all instances considered we set $\mu = 0$.

3.7.4 Epidemic parameters inference

The parameters Λ governing the epidemic process can be simultaneously inferred during the learning process of the ANN algorithm using a heuristic method inspired by Expectation Maximization (see section 3.2). Other iterative algorithms, such as SIB, can incorporate such a parameter likelihood climbing step during their convergence [63]. A comparison between the performances of the two algorithms in learning the infectiousness parameter governing the spreading process on different contact networks (**tree**, **RRG**, **proximity** and **work**) is displayed in Fig. 3.9 (left plot), in which we adopt the same setting of the patient-zero problem where the states of all individuals are known at the final time T. The ANN algorithm largely outperforms SIB in **RRG** and **proximity** graphs, obtaining comparable results for the **tree** and **work** instances.

We also test the performance of parameters inference in a more challenging scenario where the population is split into two classes, with two different rates of infections γ_1, γ_2 (which could correspond, for instance, to a simplified scenario of vaccinated/not-vaccinated individuals). The states of all individuals at final time $T = 14$ are observed for ten epidemic cascades on the **hospital** contact network. Then we infer the parameters with two different epidemic models: in the first one, the population is correctly divided (we call this the *true* model); in the second, we split the population randomly (*null* model). The goal is to verify whether the *true* model has a larger likelihood than the *null* model, that is it can better explain the observations. In the central plot of Fig. 3.9, we observe how well the *true* model can infer the correct values of the infections rate of the two sub-populations. As expected, the two values of γ inferred with the *null* model are similar to each other but different from the correct ones. From the rightmost

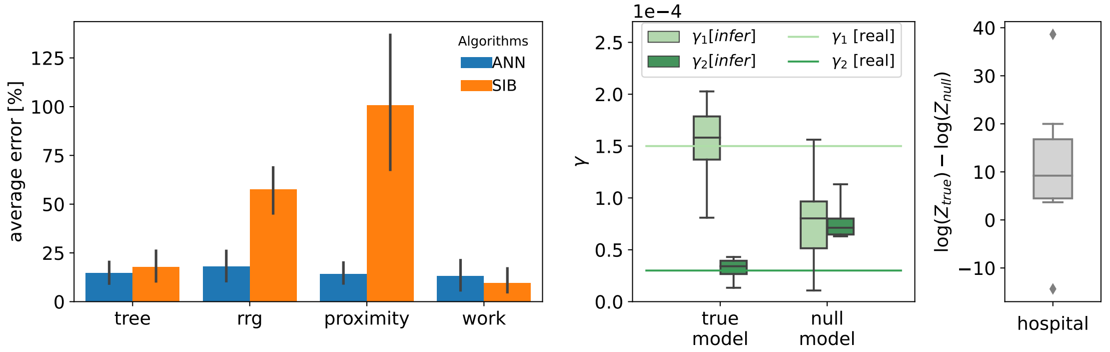

Figure 3.9: **Inference of epidemic infectiousness parameters. Left plot.** Average relative error in the inference of the infectiousness parameters over ten epidemic cascade per interaction graph. On **tree**, **RRG** and **proximity** networks, the discrete-time SIR model has infection probability respectively equal to $\lambda_{\text{tree}} = 0.35$, $\lambda_{\text{rrg}} = 0.04$, $\lambda_{\text{proximity}} = 0.03$. The **work** case has rate of infection $\gamma_{\text{work}} = 10^{-3}$. The initial conditions for the parameter learning process were set to $\lambda_{\text{init}} = 0.5$ for **tree**, $\lambda_{\text{init}} = 0.1$ for **RRG** and **proximity** networks and to $\gamma_{\text{init}} = 10^{-2}$ for the **work** network. **Central plot.** Box plot for the case of two classes of individuals with different rate of infection γ_1, γ_2 inferred by the ANN. We consider two inference model where the population is divided according the propagation model (*true* model) and randomly (*null* model), see the text for details. The *true* model is able to correctly infer the parameters with only ten different epidemic cascade. **Right plot.** Box blot of the log-likelihood difference between the *true* and *null* model. Figure first shown in [158].

plot of Fig. 3.9, we observe that the log-likelihood of the *true* model is much larger of the one of the *null* model, indicating the former better explains the observations. This example shows how the ANN approach can therefore be used to select the epidemic model that best explains the observations based on the estimate of their log-likelihood.

3.8 Robustness with respect to epidemic parameters

The scenarios in which we have run the ANN method depend on several parameters, which could change the results of our analysis. For instance, in the case of patient zero problems, we could vary the time T of the observed snapshot, the number of individuals, or the degree of the graphs in the case of synthetic interaction networks or the epidemic like the infectious or recovery parameters. All cases analysed show strong fluctuations, imposing to average over a large number of instances to have a proper statistical significance of the results and appreciate the difference in the performance of the methods compared. As specified above, we chose the infectious and recover parameters from having, on average, half of the infected individuals to decrease the instances where very few or almost all of them are infected (see Fig. 3.10, right plot). In these cases, the inference problems analysed become either trivial or impossible to solve. On the other hand, our method has a computational cost that spans several hours on modern GPU to reach convergence for each instance analysed, limiting the possibility of exploring the large dimensional space of the parameters of our inference problems. Nevertheless, we check how the performances of the method varies in the patient zero and the inference of parameters problems, with respect to the infectious parameters (λ or γ), in order to check the robustness of the results.

For the patient zero inference problem, we consider the **work** contact graph, and look at the performance of the different algorithms used, changing the infectivity γ. The figure 3.10 shows the area under the curves representing the fraction of times the patient zero is found, as in Figure 3.6 (right plots). The ANN method shows consistent performance with changing γ, giving the best score or on par with others algorithms.

Then we check the robustness of the inference of parameters in the case of of **RRG** interaction graph; the figure 3.11 shows that the inference of the infectiousness parameter λ remains good when it is changed from $\lambda = 0.02$ to $\lambda = 0.06$.

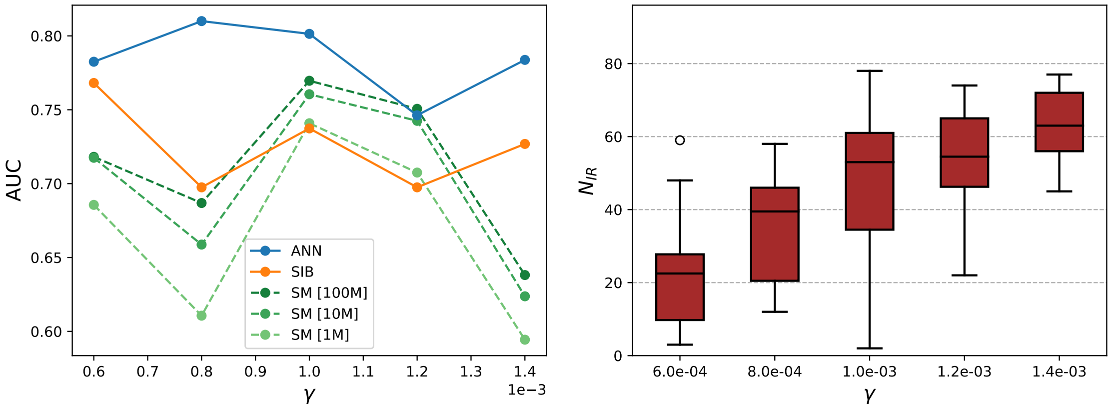

Figure 3.10: **Accuracy in finding the patient zero while varying infectivity on the work contact graph.** On the left pane, the performance of the algorithms is measured as the area under the curve (AUC) of the fraction of patient zeros found as a function of the fraction of considered individuals (the same kind of curves shown in Fig. 3.6). We consider 20 different epidemic realizations for the γ values $\gamma = 0.6 \cdot 10^{-3}$, $\gamma = 0.8 \cdot 10^{-3}$, $\gamma = 1.2 \cdot 10^{-3}$, $\gamma = 1.4 \cdot 10^{-3}$, and 100 realizations for $\gamma = 10^{-3}$, while the recovery rate is fixed $\mu = 0.02$. On the right pane, the boxplot shows the number of infected individuals with the different values of γ, in a total population of 95 individuals.

3.9 Approximations to the conditional dependencies

We now discuss the approximation to the dependency of the conditional probabilities for the individuals' trajectories. The probability of a whole epidemic cascade can be written in the following autoregressive expression:

$$q_\theta(\underline{\mathbf{x}}) = \prod_i q_i^{\theta_i}\left(\mathbf{x}_i|\underline{\mathbf{x}}_{<i}\right), \tag{3.23}$$

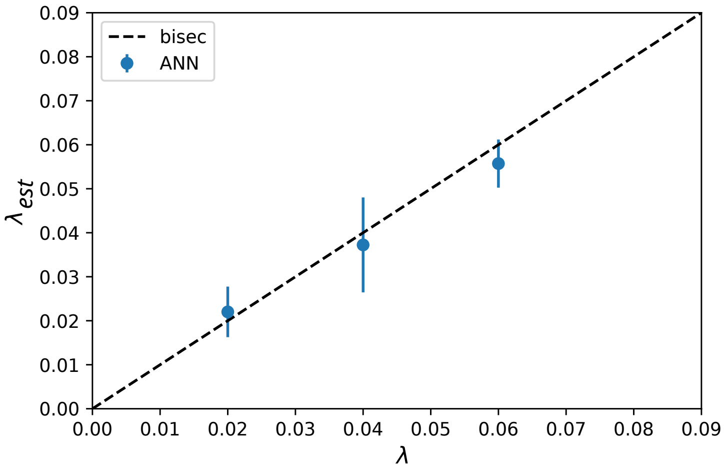

Figure 3.11: **Inference of parameters with different infectiousness**. Inference of infectiousness parameter λ in the **RRG** case (100 individuals interacting according a random regular graph with degree 10 for 15 days). The problem setting is equal to the left panel in Fig. 3.9, and is explained in subsection 3.2. The plot shows the infectious parameter λ_{est} inferred by ANN with the perfect inference line $\lambda_{est} = \lambda$ highlighted. ANN correctly estimates it within the error margin.

The dependencies in the factors $q_i^{\theta_i}(\mathbf{x}_i|\underline{\mathbf{x}}_{<i})$ can be reduced, in the case of acyclic contact networks, to only those corresponding to the next-nearest neighbours with lower index (we relegate the derivation to appendix A), but in the general case of contact networks with cycles it is just an approximation, which we employ for all contact networks analysed in the present work.

In order to check this approximation, we have run several tests on the patient zero problem, applied to the proximity random contact graphs. For this purpose, we train the ANN either (a) considering the full dependencies on all nodes with lower permutation index (we call this version *full graph*), (b) consider only the dependency on the first and second neighbours (called *next nearest neighbours* approximation) with lower permutation index, which is the approximation used in all the rest of this work, (c) considering only the first neighbours (*nearest neighbours*) with lower permutation index, or (d) ignoring the dependency on the rest of the graph (called *mean field* approximation). Figure 3.12 shows how the accuracy in finding the patient zero is influenced by the approximation chosen, and the respective reduction of the number of parameters in the network, with respect to the full graph case. We see that the *next nearest neighbours* approximation gives estimates which are on par with the full graph case, while employing less parameters, thus reducing the space and time needed for the training.

3.10 Discussion

The work presented in this chapter shows how significant individual-based epidemiological inference problems defined on contact networks can be successfully addressed using autoregressive neural networks. In problems such as patient zero detection and epidemic risk assessment,

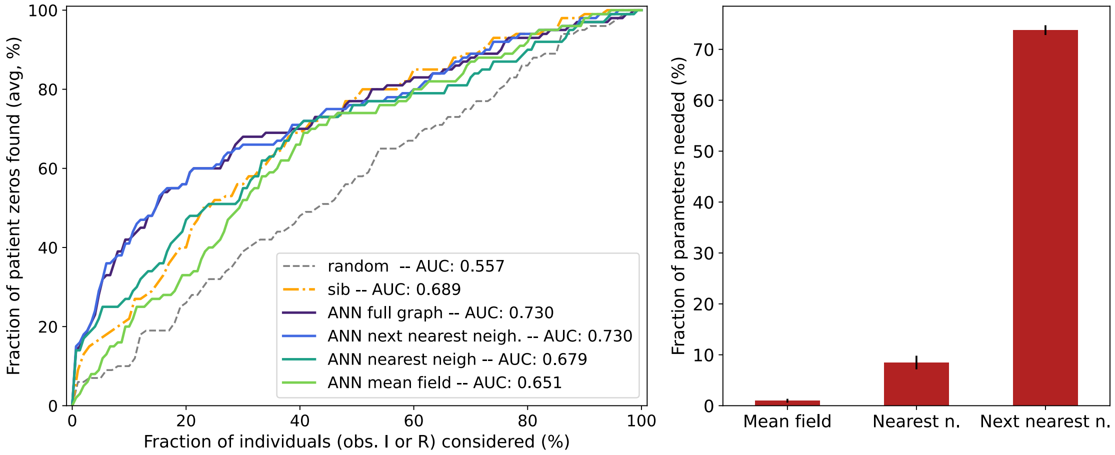

Figure 3.12: **Patient zero accuracy with different conditional dependence approximations**. **Left**: Accuracy in finding the patient zero with different approximations of the dependencies between individuals, compared with SIB (Belief Propagation) method and random guessing of the source among the infected. The "full graph" case is when no approximation on the dependencies is made. The average is made over 100 epidemic cascades on contacts network generated from a **proximity** random graph (see 3.6) with $N = 100$ individuals. The epidemic parameters used here are $\lambda = 0.03$ and $\mu = 0.02$, with $T = 15$ time instants. The legend also shows the Area under each curve (AUC). **Right:** Percentage of parameters needed, on average, when considering the approximation of the dependencies between individuals, with respect to the case of no approximation as described in section 3.9.

the method proposed in the chapter exploits the generative power of autoregressive neural networks to learn to generate epidemic realizations according to the epidemic model that are simultaneously compatible with the observations. When the model parameters are unknown, it can also infer them during the learning process. The approach is flexible enough to be easily applied to other epidemic inference problems and with different propagation models. Also, the proposed architectures for the autoregressive networks significantly reduce the number of necessary parameters with respect to vanilla implementation. Moreover, convergence properties are improved by means of a regularization method that exploits the introduction of a fictitious temperature and an associated annealing process.

According to the results obtained on three different problems (patient zero, risk assessment, and parameters inference) on both synthetic and real contact networks, the proposed method equals other state-of-the-art methods in the literature on epidemic inference, outperforming them in several cases. In particular, the ANN approach is computationally less demanding than standard Monte Carlo methods, as shown in Figure 3.7, where the number of samples generated to reach convergence scale almost linearly with the epidemic size. More efficient algorithms based on message-passing methods, like SIB, might experience convergence issues on dense contact networks like those measured in a hospital and a work office, and in these cases ANN provides significantly better results, as figure 3.6 shows. The framework presented in this chapter combines the high expressiveness of the neural networks to represent complex discrete variable probability distributions and the robustness of the gradient descent methods to train them. Moreover, the technique is a variational approach based on sampling of the distribution, which allows to compute an approximation of the log-likelihood, enabling model selection as shown in Fig. 3.9. On the other hand, like most neural networks approaches, the proposed framework suffers from some degree of arbitrariness in the choice of the neural network architecture and, consequently, the number of parameters. Also, an ordering of the variables has to be chosen, which could influence the quality of the approximations. While there is an optimal order for acyclic contact networks (see appendix A), it is unclear how to generalize this result on different systems. Although showing suitable scaling properties with system size, the method described so far reasonably needs improved architectural schemes and learning processes to be applied in epidemic inference problems regarding more than few thousand individuals.

For all these reasons, the method seems very promising for epidemic inference problems defined in small communities such as hospitals, workplaces, schools, and cruises, where contact data could be available. In such contexts, it could detect the source of an outbreak, measure the risk of individuals being infected to improve contact tracing, or estimate the channels of contagion and the infectivity of classes of people, thanks to the possibility of inferring the propagation parameters.

Chapter 4

Epidemic containment with statistical physics methods

In December 2019, an outbreak of a new kind of pneumonia was reported in Wuhan, China. This disease, which was later renamed COVID-19 (CoronaVirus Disease 19), caused great concerns because of its rapid spreading pace in China, reminding epidemiologists of the previous SARS (Severe Acute Respiratory Disease) outbreak of 2002-2003 [114]. The coronavirus responsible for the disease has been later named, because of the similarity to the original SARS virus, SARS-CoV-2, and it has spread around the world since. At the end of January the disease was declared a global health emergency [81], and on the 11th March 2020 it was declared a pandemic by the World Health Organization (WHO) [80].

The response of governments all around the world was to resort to nation-wide lock-downs [169]. While these were effective at containing the disease [145, 150], they also came at the huge economic and social cost of closing down schools and workplaces for non-essential businesses, banning social gatherings, and generally restricting travel, forcing people to stay at home [120]. In order to make the lifting of the lockdown possible, in order to restart the economy, while minimizing the number of infections, the spread of the virus had to be contained in another way. One such method is the so called *test-trace-isolate* measure, which relies on testing people with symptoms, then, if they turn out infected, find all the people that have been in contact with them and having them to be tested, too.

Before the new pandemic, the tracing of the contacts was done primarily through interviews and questionnaires, relying on each individual to remember the people they had met. This form of *manual contact tracing* works in the early stage of an epidemic outbreak and, when the number of relevant contacts is low (like with sexually-transmitted diseases), it has been found to be effective in controlling the spread of infections [14]. However, in the case of airborne diseases, such as COVID-19, where many people can be infected during a short time, lot of effort and resources are required to trace each infection [97]. Moreover, the relevant contacts for the spreading of the virus can become too many for a person to recall from memory.

To work around this problem and bring contact tracing to scale, several ways to record digitally the contacts between individuals have been developed, relying on the ever increasing ubiquitous nature of smartphones. To combat the spread of COVID-19, in 2020 numerous smartphone applications relying on Bluetooth Low Energy (BLE) to collect the contacts were released [92, 99, 108, 83, 69, 109] along with different protocols to perform *digital* tracing of the contacts of each individual that turned out to be infected [111, 73, 74, 77]. The fact that these applications could collect such detailed and personal information raised concerns for the users' privacy [103, 89], so some of the proposed algorithms rely on decentralized communication and storage protocols to mitigate this issue [104, 79, 73, 77, 74].

While it has been suggested that digital contact tracing can be effective in the containment of COVID-19, its efficiency when dealing with a large epidemic size is still disputed [85, 133, 123]. One issue is that when dealing with a large number of infections, the number of tests needed to check individuals that have been notified by digital contact tracing can be very high, surpassing test capacity. Also, testing can be expensive, especially when using Real Time Polymerase Chain Reaction (PCR), which, during the pandemic, was the preferred method for confirming a COVID-19 infection [110, 107]. When the cost of testing becomes unbearable, the reasonable choice is to put the newly-traced individuals into a very long quarantine, which, in the face of exponential growth of an outbreak, could very well result in large fraction of the population being placed in isolation [138].

In this chapter other methods are proposed to complement digital contact tracing, which use probabilistic inference techniques derived from statistical physics. This allows for the evaluation of the *risk* of each individual to contract the disease, which is a more useful measure when compared to the binary information about an individual who has or has not been in contact with an infected person (which was the only information used by previous contact tracing techniques). Throughout the chapter, it will be shown how these probabilistic methods perform in controlling outbreaks with respect to risk-aware contact tracing, in the context of limited testing capability. This will be shown first in simpler SIR epidemics on contact graphs, and then with the OpenABM model, which reproduces COVID-19-like epidemic cascades.

The work and results presented in the rest of the chapter have been first published as Baker et al. [118].

4.1 Risk inference with the SIR Model

As previously mentioned in the thesis, doing *risk inference* involves computing the risk of infection for all individuals at a particular time instant, usually the current one. In the scenario of controlling an ongoing epidemic outbreak, it's quite possible that at some day t, a certain number of tests have been performed in the previous days, and new tests have to be done on the population. Since it's highly improbable that there is enough testing capability to test a large part of the population for infection, it is then necessary to identify those individuals that have the highest probability of being infected. Moreover, early identification of infected individuals is much better than a late one, as it will prevent other individuals from getting infected. This is because if a person is found positive after the test, he/she is put into quarantine, limiting his/her contacts with other people, thus slowing down the propagation of the disease.

Let us suppose that the epidemic is spreading according to the SIR epidemic model. In this case, the posterior distribution at time t can be derived from the full SIR posterior distribution (1.24):

$$p\left(\mathbf{x}^t \middle| \mathcal{O}\right) \propto \sum_{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{t-1}} p\left(\mathbf{x}^0\right) \prod_{i=1}^{N} \prod_{\tau=1}^{t} p\left(x_i^\tau \middle| \mathbf{x}^{\tau-1}\right) \prod_{(i,\tau,\sigma)\in\mathcal{O}} p\left(z_i^\tau = \sigma \mid x_i^\tau\right) \tag{4.1}$$

where the observations collected so far are gathered in the set $\mathcal{O}$. This shows that it's necessary to marginalize the full epidemic cascade $\underline{\mathbf{x}}$ on all the previous time instants $\tau < t$, which involves, at a first glance, summing on an exponential number of terms (that are all the combinations of x_i^τ). A straightforward strategy to perform this evaluation would be simulating a large number of instances of the propagation, and then estimating $p\left(x_j^t = I\right)$ (for example) as the fraction of instances where individual j is in state I at time t. This idea, however, would require a centralized system, and large amounts of computing power, especially because the number of simulated cascades compatible with the observations is very low with moderate time windows and medium population sizes. This method is thus less appealing.

Here, instead, it will be shown how to use some approximate techniques from statistical physics that allow for a good estimate of $p\left(x_j^t\right)$, namely the Simple-Mean-Field (SMF) algorithm and the Belief Propagation (BP) algorithm. While the latter method has already been shown in the thesis, it will be reformulated to account for missing contacts and non-Markovian spreading parameters.

There are two main advantages of mean-field methods in this context. First, they are based on systems of equations for the marginals of interests defined in (4.1) (or similar related quantities), so they can directly provide the individual risks. Second, these systems can be (and typically are) efficiently solved iteratively, which fits well with a distributed approach in which computation is performed on individuals' cellphones, with a relatively small, though regular, exchange of information between individuals that have been in a contact.

As it will be shown, the BP method is much more accurate than SMF, but it is relatively more complex and, in a decentralized implementation, it would require the exchange of a larger amount of information between individuals.

4.1.1 Simple Mean Field approximation

The *Simple-Mean-Field* (SMF) method is a statistical physics inspired heuristic algorithm that computes the marginal probabilities through an iterative process. The starting point of the SMF heuristic is the Mean Field approximation of the SIR model, described in section 1.3.1. We recall here that this approximation gives that the probability of the state of individual i at time $t+1$ is given by:

$$p_{MF}\left(x_i^{t+1} = S\right) = p_{MF}\left(x_i^t = S\right) \prod_{k \neq i} \left(1 - p_{MF}\left(x_k^t = I\right) \lambda_{k\to i}(t)\right) \tag{4.2}$$

$$p_{MF}\left(x_i^{t+1} = R\right) = p_{MF}\left(x_i^t = R\right) + \mu_i p_{MF}\left(x_i^t = I\right) \tag{4.3}$$

for the states S and R, and that the probability of being infected is:

$$\begin{aligned} p_{MF}\left(x_i^{t+1} = I\right) &= 1 - p_{MF}\left(x_i^{t+1} = S\right) - p_{MF}\left(x_i^{t+1} = R\right) \\ &= p_{MF}\left(x_i^t = S\right)\left[1 - \prod_{k \neq i}\left(1 - p_{MF}\left(x_k^t = I\right)\lambda_{k\to i}(t)\right)\right] + \left(1 - \mu_i\right) p_{MF}\left(x_i^t = I\right) \end{aligned} \tag{4.4}$$

with the assumption that, in case k is not in contact with i at time t, $\lambda_{k\to i}(t) = 0$.

In practice, considering that the probability of transmission is small, a further approximation can be made to the MF algorithm, introducing the following linearised forms (which have been checked to hold in the regimes of epidemic propagation explored in the chapter):

$$\begin{aligned} p_{MF}\left(x_i^{t+1}=S\right) &= p_{MF}\left(x_i^t=S\right)\left(1-\sum_{k\neq i} p_{MF}\left(x_k^t=I\right)\lambda_{k\to i}(i)\right), \\ p_{MF}\left(x_i^{t+1}=R\right) &= p_{MF}\left(x_i^t=R\right)+\mu_i p_{MF}\left(x_i^t=I\right), \\ p_{MF}\left(x_i^{t+1}=I\right) &= p_{MF}\left(x_i^t=S\right)\sum_{k\neq i} p_{MF}\left(x_k^t=I\right)\lambda_{k\to i}(t)+ \\ &\quad +\left(1-\mu_i\right)p_{MF}\left(x_i^t=I\right). \end{aligned} \tag{4.5}$$

This approach offers several advantages: first, every individual j can estimate his/her probabilities $p_{MF}\left(x_i^t=S\right)$, $p_{MF}\left(x_i^t=I\right)$, $p_{MF}\left(x_i^t=R\right)$ every day, by updating the equations (4.5), and these probabilities can be stored in the user's phone. For the update, individual j needs to receive, during the contact with k, the information on $\lambda_{k\to j}(t)$ and the information from k about his estimates of $p_{MF}\left(x_k^t=S\right)$ and $p_{MF}\left(x_k^t=I\right)$. The value of $\lambda_{k\to j}(t)$ can be estimated with the standard contact-tracing information, which estimates the encounter duration within a certain distance, for instance based on Bluetooth signals between the phones of j and k. On top of this, the two individuals' phones should exchange their estimate of the probabilities. In this way, the information is fully distributed, and there is no need for a central system that stores the full information.

Suppose that, at time t_{obs} an individual i is tested or presents illness-associated syndromes. Then the state of i is known: $z_i^t\in\{S,I,R\}$ and

$$p_{MF}\left(x_i^{t_{obs}}=\sigma\right)=\delta_K\left(\sigma,z_i^t\right)$$

where δ_K is the Kronecker delta. In case of syndromes at time t_{obs} the probability $p_{MF}\left(x_i^{t_{obs}}\right)$ is updated on the basis of external medical data, namely the probability to be infected among all people presenting the same set of syndromes. As explained previously, the information about tests and syndromes must be propagated back in time and be used to update the risk levels of the contacts of person i in recent times. However, a simple inference heuristic that turns out to be quite efficient consists in adapting the mean-field equations (4.5) in order to take into account the results of tests and symptoms: assuming that we are estimating the probabilities for each individual i to be in each of the three states $\sigma\in\{S,I,R\}$ at a given time t, $p_{MF}\left(x_i^t=\sigma\right)$, it suffices to run the mean-field equations (4.5) starting at time $t-t_{MF}$ with the whole population in state S (setting to 0 the probability for the other states), and imposing the constraints due to the tests done in the interval $[t-t_{MF},t]$ as follows. If j is tested at a time t_{obs} in this interval, and the result is state z, then:

$$\text{if } z=S: \quad p_{MF}\left(x_i^{t'}=S\right)=1 \quad \text{for } t'\in[t-t_{MF},t_{\text{obs}}] \tag{4.6}$$

$$\text{if } z=I: \quad p_{MF}\left(x_i^{t'}=I\right)=1 \quad \text{for } t_{\text{obs}}-\tau_{MF}\leq t'\leq t_{\text{obs}} \tag{4.7}$$

$$\text{if } z=R: \quad p_{MF}\left(x_i^{t'}=R\right)=1 \quad \text{for } t'\geq t_{\text{obs}} \tag{4.8}$$

This inference procedure then depends on two parameters: τ_{MF} is the typical time between the infection and the testing consecutive to the apparition of syndromes, and t_{MF} is the integration time of the mean-field procedure.

4.1.2 Belief propagation inference of epidemic risk

The Belief Propagation (BP) method developed for epidemic inference (presented in Chapter 2) can also be applied for the computation of the risk of infection (4.1). However, it is now going to be derived again in a formulation in infection and recovery times $(\mathbf{t}, \mathbf{r})$ which is better suited to the problem, especially for a temporal network of contacts (see section 1.4) and for other cases which might occur while containing an outbreak (which will be explained later). This approach, although somewhat convoluted, is able to infer the propagation of the epidemic, treating the observations in a fully Bayesian way, thus achieving higher accuracy than SMF.

One major difference from the derivation in section 2.3 is that now it is assumed that the epidemic is spreading in continuous time, which means that contacts between every pair of individuals (i, j) can be collected into a finite set of time instants $\mathscr{T}_{ij} \subset \mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$. These contacts now define the time instants of interests, as it's easy to see that, since contagion is instantaneous, the time of infection of an individual will be $t_i \in \mathscr{T}_i = \bigcup_{j \in \partial i} \mathscr{T}_{ij}$. Then, the case $t_i = \infty$ means that i is never infected.

To add more flexibility to the method, it's also assumed that the probability of i infecting j, given that i is infected and j susceptible at time $s_{ij} \in \mathscr{T}_{ij}$, depends on the absolute time of transmission $s_{ij} \in \mathscr{T}_{ij}$, and the time of infection of i, t_i:

$$\lambda_{i \to j} \doteq \lambda_{ij}\left(s_{ij}, t_i\right),$$

keeping in mind that $\mathscr{T}_{ij} \equiv \mathscr{T}_{ji}$, so that also $s_{ji} \in \mathscr{T}_{ij}$. The difference with the previous approach of section 2.3 is that each s_{ij} now represents the absolute time of possible infection, instead of the delay from the time of infection. The recovery time $r_i \in \mathbb{R}_\infty$ is assumed to be drawn from a continuous distribution: since recovery only happens after infection, it's more useful to model the recovery delay $r_i - t_i$ and assume that its probability distribution is $R_i\left(r_i - t_i\right)$. Then, the transmission time s_{ij} is going to depend on both t_i and r_i, and its conditional probability distribution $S_{ij}\left(s_{ij} \middle| t_i, r_i\right)$ will be given by:

$$S_{ij}\left(s_{ij} \middle| t_i, r_i\right) = \mathbb{I}\left[t_i < s_{ij} < r_i\right] \lambda_{ij}\left(s_{ij}, t_i\right) \prod_{t_i < s < s_{ij}} \left[1 - \lambda_{ij}\left(s, t_i\right)\right] + \mathbb{I}\left[s_{ij} = \infty\right] \prod_{s \geq r_i} \left[1 - \lambda_{ij}\left(s, t_i\right)\right] \quad (4.9)$$

This truncated distribution is given by the fact that individual i will be able to transmit the disease in the open time interval (t_i, r_i) and, once the time r_i is reached, he/she won't be able to infect anyone, thus $s_{ij} = \infty$.

Within this formulation, the Markovian (i.e. memory-less) SIR model on a temporal network can be obtained by setting an infection probability independent from the infection time, $\lambda_{ij}\left(s_{ij}, t_i\right) \equiv \lambda_{ij}\left(s_{ij}\right)$, and an exponential recovery time distribution $R_i\left(r_i - t_i\right) = \mu_i e^{-\mu_i\left(r_i - t_i\right)}$.

Modelling auto-infections

The formulation presented so far represents a closed system in which infections can occur only through existing contacts between nodes. In order to account for the presence of an initial number of infected seeds, it is necessary to introduce the possibility that individuals can infect spontaneously, i.e. without having a contact with other infected people. Introducing such probability will be useful in the case in which full knowledge of the contact network is not possible (which will be discussed later in the chapter).

A simple solution is to add fictitious contacts to a node that is always infected. In particular, for each individual i and each time $t \in \mathcal{T}_i$, a contact with the node i_t^* is added, with an infection transmission variable $s_{i_t^* i}$ that has the probability $p_t\left(s_{i_t^* i} = t\right) = \gamma_i^t$ of infecting i if he/she is susceptible at that time (typically small), and $p_t\left(s_{i_t^* i} = \infty\right) = 1 - \gamma_i^t$ with the other case[1]. For ease of notation, all the auto-infection contacts probability will be grouped together in the factor

$$A_i\left(\mathbf{s}_{i*}\right) = A_i\left(\left\{s_{i_t^* i}\right\}_{t\in\mathcal{T}_i}\right) = \prod_{t\in\mathcal{T}_i\setminus\{\infty\}} p_t\left(s_{i_t^* i}\right).$$

For convenience, let us define $\partial^* i$ as the enlarged neighbourhood of i including all extra nodes $\{i_t^*\}_{t\in X_i}$. Given $\{s_{ki}\}$ for $k \in \partial^* i$, the infection time t_i satisfies, in a deterministic way:

$$t_i = \min_{k\in\partial^* i} s_{ki} \tag{4.10}$$

Then, we can write the joint probability distribution of $\mathbf{t}$, $\mathbf{r}$ and $\mathbf{s}$:

$$p\left(\mathbf{t}, \mathbf{r}, \mathbf{s}\right) \propto \prod_i \delta\left(t_i, \min_{k\in\partial^* i} s_{ki}\right) A_i\left(\mathbf{s}_{i^*}\right) R_i\left(r_i - t_i\right) \prod_{(ij)} S_{ij}\left(s_{ij} \middle| t_i, r_i\right) \tag{4.11}$$

where the delta functions enforce the dynamical constraint (4.10), and the product over (ij) runs over all pairs of individuals, including the fictitious auto-infections nodes.

Factor graph representation

Combining all the previous terms with the observations $\mathcal{O}$, the application of Bayes' rule gives the following posterior distribution:

$$p\left(\mathbf{t}, \mathbf{r}, \mathbf{s} \,|\, \mathcal{O}\right) \propto \prod_i \delta(t_i, \min_{k\in\partial^* i} s_{ki}) A_i\left(\mathbf{s}_{i^*}\right) R_i\left(r_i - t_i\right) p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right) \prod_{(ij)} S_{ij}\left(s_{ij} \middle| t_i, r_i\right) \tag{4.12}$$

in which $p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right)$ is the likelihood of all the observations $\mathcal{O}_i$ that have been made about the state of individual i.

The naive interpretation of this probability distribution as a graphical model would introduce many short cycles which were not present in the original contact graph. For example, pairs $\left(t_i, s_{ji}\right)$, $\left(t_i, s_{ij}\right)$, $\left(t_j, s_{ij}\right)$, $\left(t_j, s_{ji}\right)$ share respectively factors with indices $i, (ij), j, (ji)$, effectively forming a small cycle. This is then solved (differently than in section 2.3) by regrouping several factors into a single Ψ_i, and considering $\left(s_{ij}, s_{ji}\right)$ as a single variable:

$$\Psi_i\left(t_i, r_i, \{s_{ki}, s_{ik}\}_{k\in\partial^* i}, \mathcal{O}_i\right) = \delta(t_i, \min_{k\in\partial^* i} s_{ki}) A_i\left(\mathbf{s}_{i^*}\right) R_i\left(r_i - t_i\right) p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right) \prod_{j\in\partial i} S_{ij}\left(s_{ij} \middle| t_i, r_i\right) \tag{4.13}$$

which results in a factor graph for (4.12) in which variables $\left(s_{ij}, s_{ji}\right)$ have degree two and live in the middle of the factors Ψ_i and Ψ_j, i.e. a topology that closely follows the one of the original contact network:

$$p\left(\mathbf{t}, \mathbf{r}, \mathbf{s} \,|\, \mathcal{O}\right) = \frac{1}{Z(\mathcal{O})} \prod_i \Psi_i\left(t_i, r_i, \{s_{ki}, s_{ik}\}_{k\in\partial^* i}, \mathcal{O}_i\right) \tag{4.14}$$

---

[1] A particularly interesting case is with $\gamma_i^0 = \gamma \to 0$ and $\gamma_i^t = 0$ for $t > 0$: in this case individuals can be self-infected only at time 0, representing a closed system with a single unknown seed at time $t = 0$.

The derivation of the message passing equations follows the approach described in section 2.2, and gives rise to messages $m_{ij}\left(s_{ij}, s_{ji}\right)$ that goes from ψ_i to ψ_j, and $m_{ji}\left(s_{ij}, s_{ji}\right)$ which will go in the opposite direction. The full derivation of the messages and the beliefs for the infection time, $b\left(t_i\right)$, which represent effectively the risk of infection for each individual, is discussed in Appendix B. The code for this method has been published on GitHub [151].

Finite-window approximation

In this BP-based epidemic tracing scheme, the number of messages exchanged between two individuals grows quadratically with the number of temporal contacts occurred between them. However, only recent contacts are important to determine marginal probabilities at present time, therefore keeping only a short time window (about two or three weeks) is sufficient to obtain quasi-optimal results. For better accuracy, information about contacts and observations at the dropped times is included approximately as simple factorized priors applied at the start of the window. This prior contains the posterior probability at the first non-dropped time computed only using contacts and observations at the dropped time (and the prior computed in the previous step). All simulations have been performed using a 21 days time window.

4.2 Mitigation of proximity-based epidemics

In this section, the theoretical capabilities of the SMF and BP techniques are going to be tested first, in the context of SIR epidemic cascades on consecutive time instants with homogeneous parameters (i.e. spreading according to equations (1.20) where $\lambda_{ij}^t = \lambda$ and $\mu_i = \mu$). These epidemics spread in a population of N individuals, with a contact graph that is dynamical, and generated as follows: the individuals are distributed uniformly in a square of side $\sqrt{N}$, and at each time step a contact can be established between two individuals i and j with a probability $e^{-d_{ij}/\ell}$, where d_{ij} is the Euclidean distance between the points and ℓ is a parameter that controls the density of the contact graph. These graphs are called *proximity*-based in the following because of this generation procedure.

Using this spreading model, it's possible to test the accuracy of the statistical inference algorithms in the most ideal situation, as the prior for the inference corresponds to the propagation model. In order to see the efficacy in the containment of epidemics, an intervention framework is developed following this procedure: at each time instant, spreading occurs according to the above model, then each algorithm is run to determine the risk of individuals at that time. Afterwards, the n_r individuals with the highest estimate of the risk are tested, and, if found infected, are confined (which effectively means that they cannot have any future contacts), to stop transmission of the disease. Results of the test are assumed to be available on the same time step and are included in the observations used to adjust the probabilities of risk on the next time step. In order to save tests, they are applied only to individuals who have not previously been found infected. Also, in order to model a reaction time to the introduction of a new disease in the population, a delay in the start of testing is considered (t_{start} time instants have to pass). At the beginning of each simulation, every individual is susceptible, except for a number n_{pz} which are infected (the ”patients-zero”). In the following, each time step will correspond to a day.

For estimating the risk, for the Simple Mean Field method, the probability of an individual being infected, $p_{MF}\left(x_i^t = I\right)$, is used for the ranking, as computed by the algorithm. For the BP method, instead, the probability of infection in the last δ_{rank} days is used (see Appendix B), as

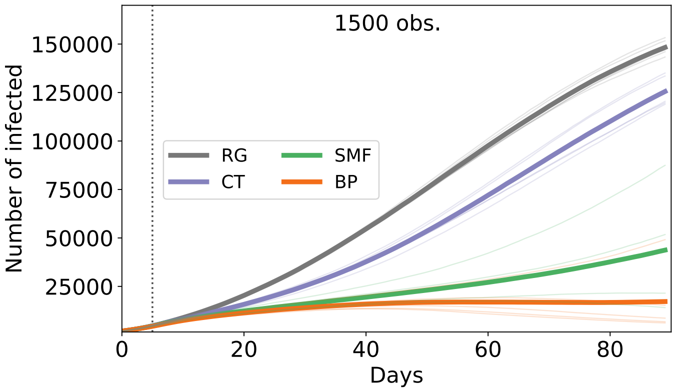

Figure 4.1: SIR model on proximity-based random network with 6 contacts on average per day and 500 000 individuals. In the plot we show the average numbers (bold lines) of infected individuals versus time among three different realizations (thin lines) of the epidemics with 200 patients zero. The epidemic parameters are the same as used by the inference algorithms: $\lambda = 0.05$, $\mu = 0.02$. For the first 10 days the evolution is free, afterwards 1500 tests are applied each day according to each algorithm. Also, 50% of the infected individuals are considered severely symptomatic on each day, and are thus observed as infected 5 days after their infection. The parameters used for these simulations are $\tau = 5$ for both SMF and CT, and $t_{\mathrm{MF}} = 15$ for MF. Image first shown in [118].

prioritizing recent infections can be more effective as it helps containing the "boundary" of an ongoing outbreak. In the following, $\delta_{\text{rank}} = 10$ days.

Other than BP and SMF, an implementation of Contact Tracing (CT) is included for comparison. This method calculates the risk of infection of each individual at time t by counting the number of contacts with confirmed positive people that happened during the time interval $[t - \tau, t)$. In this way, it represents a slight improvement on the "infectious contact notification" method which was implemented as digital contact tracing after the first wave of COVID-19, which does not account for the number of infectious contacts, but only for their presence. Moreover, the last ranking strategy considered for comparison is Random Guessing (RG), in which the risk of each individual is drawn at random.

This simulation framework, applied for three months to a population of 500 000 individuals yields large fluctuations across runs (Figure 4.1). Nevertheless, one sees a very clear signal indicating that the statistical physics-based inference methods, SMF and BP, largely improve upon the usual CT, itself better than RG. The best inference method is clearly BP, but the simpler SMF is also quite successful. Even in this pessimistic regime, i.e. where a large fraction of the population gets infected, both risk inference methods significantly slow down the epidemic spread, when compared to classic contact tracing.

4.2.1 Robustness of the mitigation

Let us check now how robust the risk evaluation methods are when the spreading parameters are different than the ones given to the algorithms. For this objective, three experiments have been devised. For the following cases, we'll consider twenty realizations of the SIR epidemics on the proximity model of 10^5 individuals (same as above), and apply our intervention protocol from time 10 to time 50.

In the first experiment, we run the epidemic dynamics using a probability of infection $\lambda = 0.05$ and a probability of recovery $\mu = 0.02$ and perform the inference procedure using several values of mismatched probability of infection, in the range $[0.001, 0.1]$. To compare the performances in the epidemic mitigation we measure the number of infected and recovered individuals at time 50, shown in Fig. 4.2(a). The random ranker (RG) and contact tracing (CT) do not use the probability of infections λ, so they are insensible to its variation. The SMF algorithm performs equally well in the underestimation regime of λ, and the performances slowly decay as long as the probability of infection is overestimated. Remarkably, the BP algorithm performs equally well in the range of λs considered.

In a second experiment, we perform the inference step using the SIR model with $(\lambda, \mu) = (0.05,0.02)$, while the λ_{ij}^{t}'s, for each time t, associated with the true contacts are now random variables distributed according to a normal distribution with mean 0.05 and standard deviation $\sigma \in [0.005, 0.03]$. The latter simulation may mimic a realistic situation where the Bluetooth signal used by the app is affected by noise. The performances of all methods are shown in Figure 4.2(b). CT and RG performances are independently of the value of σ, by construction, and their results only vary due to the increasing heterogeneity of the true probability of infection. For the other cases, we notice that even when the standard deviation of the probability of infection reaches the value 0.03, that is when σ is almost equal to the mean value, the number of individuals that got infected when using SMF or BP slightly increases if compared to the small noise regime (i.e. for $\sigma = 0.005$). However, the linear regression curves, fitting the mean values over the twenty realizations of the dynamics, appear almost flat in the studied regime, suggesting that the inference-based methods are notably robust to a noisy detection of the contacts.

Finally, Let us treat the case where only a certain fraction of contacts is available to the risk assessment methods. The set-up of this experiment is the one used for Figure 4.1. Specifically, a fraction of the links of the network is randomly extracted (81%, 49%, and 36%) for each time step of the dynamics, and given to the ranking algorithms.

The results are shown in Figure 4.3: for all the ranking procedures, the number of infected individuals changes only slightly in all cases but 36%, suggesting that all methods are remarkably robust against partial information of the network contacts. Surprisingly, only for a very small fraction of the known contacts (i.e. 0.36), the spreading is notably affected. This behavior can be explained by the features of the proximity models: close neighborhoods are repeatedly connected in short time windows, hence, even when a strong pruning is performed, the information stored in the detected links suffices to perform an effective inference.

4.3 The OpenABM model

While the performance is good when the outbreaks follow the SIR model, real world epidemics have several features that are not captured by this model. In fact, since the aim of this chapter is to show the potential application of probabilistic methods like SMF and BP to COVID-19-like epidemics, it is important to test those on a model that is able to show features of COVID-19 spreading. For this task, an agent based model is required, which is able to track the spread of the disease at the individual level. In this sense, agent-based models provide stylized but sufficiently reliable representations of the actual contact networks on which contagion between

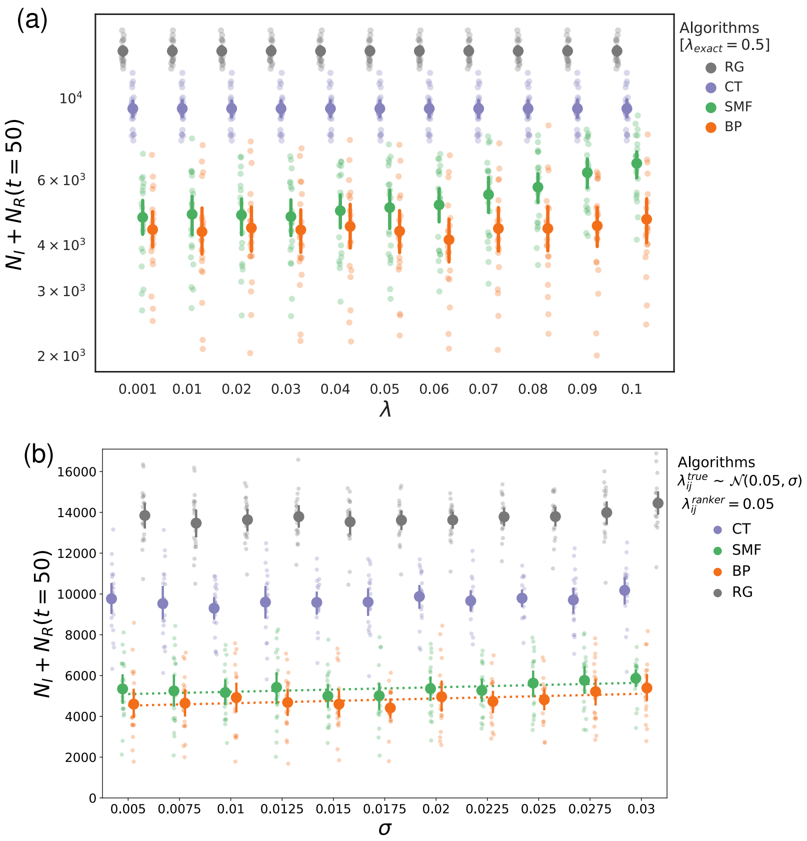

Figure 4.2: Average number of infected and recovered individuals among a population of 10^5 at time 50, while interventions start at time 10. In (a) the epidemic dynamics use a probability of infection $\lambda = 0.05$ and a probability of recovery $\mu = 0.02$, instead the inference algorithms use different values for λ shown in the x-axis. In (b) the ranking strategies use $\lambda = 0.05$ while the true probability of infection is a random variable distributed according to a Gaussian distribution of mean 0.05 and a standard deviation σ in the range $[0.005, 0.03]$ (x-axis). The number of total infected (I+R) individuals slightly increases when the distribution of the λs becomes broader as suggested by the linear regression curves in figure (b). Figure first shown in [118].

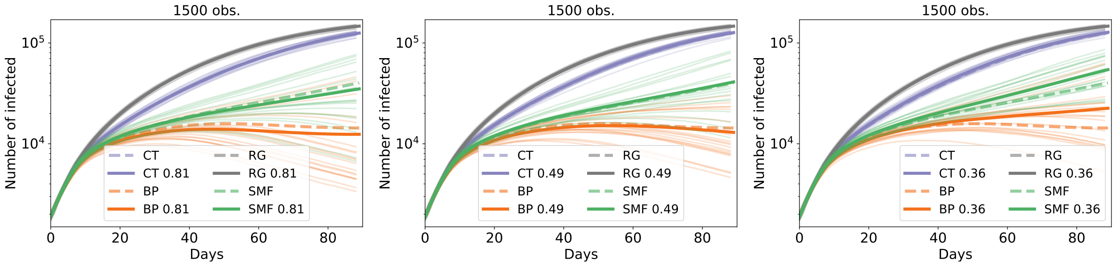

Figure 4.3: **Containing with pruning of contacts**. Number of infected individuals are shown w.r.t. time for proximity model (same as Figure 4.1). Thick lines are the averages, while thin lines are single realizations. From left to right, the percentage of links used within the inference procedure is changed: 81%, 49% and 36%. All inference-based methods are remarkably robust against a partial knowledge of the contact network: indeed, the number of infected individuals notably increases with respect to the noiseless case (dashed lines) only when 36% of the contacts is kept. Figure first shown in [118].

individuals could take place, thus becoming a natural and necessary tool for analyzing the effects of interventions strategies. The OpenABM-Covid19 model by R. Hinch et al. [131] (called simply OpenABM for the rest of the thesis) is the one used in this chapter, and, apart from capturing features of real contact networks, its aim is to model real epidemiological aspects of COVID-19 spreading. This section describes some characteristics of the model that contribute to that objective.

For starters, the N individuals in the OpenABM model are divided into 9 age classes, ranging from (0-9) to (80+) years. For each of these groups, both the contacts and the transmission have different parameters. The distribution of the population in each age group is constructed in such a way to reflect UK demographics data [131]. Each individual is represented by a node of a multi-layered network and takes part in three different subnets describing different social contexts:

- The **household network** includes the interactions with family members at home and is a fully connected graph. Household sizes range from 1 (living alone) to 6, and their distribution on the population is matched to UK demographics.

- The **workplace network** models the interactions inside schools, offices, similar workplaces for adults and recurrent social activities for elderly people. Each network is constructed as a Watts-Strogatz small-world model [11]. Daily interactions in this network vary according to the age class.

- The **random network** is included to consider interactions that occur once in a while, independently, for each day. The number of random contacts for each individual is the same each day and it is drawn from a negative-binomial distribution. The mean number of connections within the population, however, depends on the age class. The negative-binomial distribution is used to take into account the presence of super-spreaders in the network.

For the infection dynamics, the OpenABM model follows a state evolution derived from the

SEIR model (section 1.3.2) instead of the SIR, as the former is best suited for COVID-19 because of the incubation period of the disease [96], with each time instant corresponding to a single day. Instead of 4 states, however, there are 11, accounting for symptomatic status, hospitalization and intensive care. The states and the possible transitions are shown in Figure 4.4, where also the probabilities of moving from one state to the next are indicated. A susceptible individual thus may become asymptomatic (A), mildly symptomatic (SM) or severely symptomatic (SS). In the last case, he/she may need to be hospitalized, and then, be taken to intensive care units (ICUs). Only in these cases he/she may die, but it is always possible to recover from the disease (although, more slowly from the ICU). The recovered (R) and dead (D) states are the only absorbing ones. The time each individual takes for each transition is drawn from a gamma distribution (except for t_{hosp}, which follows a shifted Bernoulli distribution). The parameters for all of these transitions have been fitted for COVID-19 progression, and can be found in tables S7 and S8 of [131].

It is important to note that there is no exposed (E) state once an individual is infected. This absence is justified by the fact that the transmission of the disease is modelled in such a way that prevents an individual passing the disease the immediately after infection. This is achieved through the dependence of the daily infection probability $\lambda_{ij}(t)$ on a (discretized) gamma distribution that explicitly depends on the time passed since the infection. Also, the infection probability depends on the symptomatic state of the potential infector, the age of the potential infected and the intrinsic infectiousness level of the virus. The interaction network is also important: the infection strength of the contacts inside households is double that of other environments, to account for the typically longer duration of domestic interactions.

The model provides the possibility of intervention in order to slow down and, if possible, contain the epidemic outbreak. In particular, it is possible to introduce interventions of increasing severity, from case-based measures (e.g. quarantine for individuals which are positive to swab tests and their housemates) to mobility restrictions for some categories of individuals and lockdown scenarios. Moreover, the OpenABM model is very appropriate for the implementation of contact tracing strategies. This is due to the fact that, assuming that the contacts are collected through a smartphone application, the model also provides for the possibility of varying the adoption fraction of the app within the population, possibly introducing different percentages of adoption in different age groups of individuals.

4.4 Epidemic containment of OpenABM outbreaks

In this section, the previously-described statistical physics based methods are applied for the containment of COVID-19-like epidemics generated with the OpenABM model. The framework used for this task is the one previously described: starting from a certain day, on each day risk evaluation is performed using the information from tests performed until the previous day, and the people with the highest risk of infection (the top n_r ones) are tested and, if found positive for COVID infection, are quarantined. The new tests results (which, again, are assumed to be given without delay) are used in the risk evaluations of the following days. Again, tests are not performed on previously-discovered infected individual to save capacity, and at the beginning every individual is susceptible while a small number (n_{pz}) are infected.

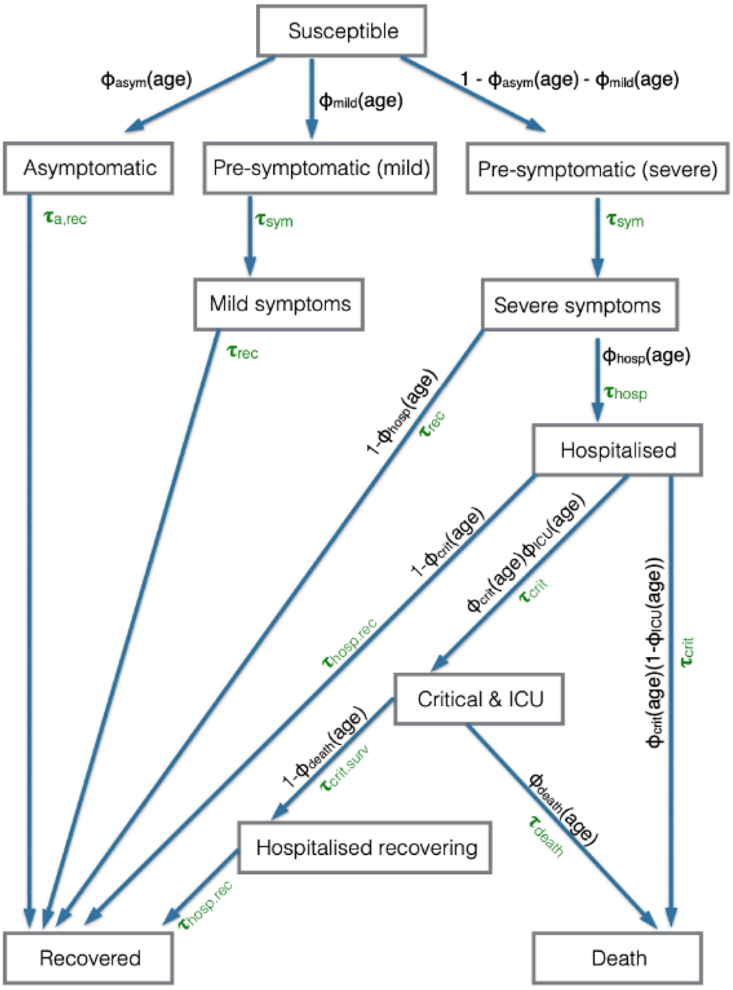

Figure 4.4: **Disease progression in OpenABM**. Image taken from [131] under the Creative Commons Attribution License, with no modifications applied. The ϕ are the probabilities of traversing each transition, from the previous state, the τ are instead the time taken for the transition, each following (different) gamma distribution. These parameters can be found in tables S7 and S8 of [131].

4.4.1 Reduction to the SIR model

In order to use the aforementioned algorithms (SMF, BP), an estimate of the infectiousness and the recovery probability distribution has to be found. These parameters are obtained from a fit of the experimental data produced by the OpenABM model. Since the SMF algorithm employs the Markov SIR model, disregarding the dependence on the infection time of the infectiousness, only two parameters (λ, μ are required). For the BP method instead it is possible to use non-Markovian SIR parameters, and both the infectiousness and the recovery distribution are fitted with a Gamma distribution.

For estimating the infectiousness, of all the parameters of the OpenABM model that give a contribution, only the base information of the contact network can be derived from contact tracing (household contact are twice more infectious), while the rest can be estimated from averages on the whole population. By including this, the probability distribution used for BP was Gamma(k_I=5.76, $\mu_I = 0.96$), rescaled by 0.25. For MF, instead, the fit value was $\lambda = 0.02$. The network information of contact weight g ($g = 2$ for the household, $g = 1$ else) was given to all of the risk evaluation strategies, including BP and MF.

The estimation of the recovery distribution requires identifying all the states from which an individual stops being infectious, which in the OpenABM case includes all the hospitalized states (which include all those reachable from the hospitalized state in Figure 4.4). For the equivalent SIR model, all these are mapped into the R state (along with Recovered and Dead states). Then, the distribution of the recovery time is obtained by averaging on all paths that lead to the equivalent R states, using the distribution of the ages in the population. A fit of the obtained curve gives a Gamma distribution with parameters $k_R = 10$ and $\mu_R = 0.57$ for the BP method, and a recovery probability for the Markov SIR model of $\mu = 1/12$, which is used in SMF. Also, the values used for other parameters of the SMF method are $\tau_{\text{MF}} = 5$ and $t_{\text{MF}} = 10$.

4.4.2 Testing the algorithms

We will now discuss the performance of the each method of risk evaluation. In the following, unless stated otherwise, the OpenABM model is run on 500,000 individuals with for 100 days. It will be assumed that all individuals that develop severe symptoms (SS) are immediately tested, and that $\rho_{SM} = 50\%$ (on average) of the new mildly symptomatic ones self-report for testing each day. Each day, each risk evaluation strategy (RG, CT, SMF, BP) is going to run according to the intervention framework described in section 4.2, with a fixed number n_{obs} of tests available, which is equal among different strategies. It is important to note that the number of symptomatic individuals that can be observed is not fixed, and it is not going to impact the number of tests available to the ranking method.

From these settings, two scenarios are derived, differing only in the number of patient zeros n_{pz} and the starting day of intervention t_{start}: scenario A has $n_{pz} = 50$ and $t_{\text{start}} = 10$, while for scenario B $n_{pz} = 20$ and $t_{\text{start}} = 7$. For the time being, Let us assume that the results of the test are noise-less and the full knowledge of the contact network is available, which corresponds to an ideal scenario in which all individuals in the population have application enabling the collection of contacts (contact tracing app) installed on the smartphone, and the results of the test are 100% reliable.

Figure 4.5 shows the results of the containment for three independent realizations of the epidemics (thick lines indicate the mean number of infected individuals). The number of tests performed per day n_{obs} is varied from 625 to 5000, with four different values shown. It is possible to see that the size of the epidemics, represented by the number of infected individuals, is significantly reduced with the statistical-physics derived algorithms if compared to random testing and also to the classic contact tracing, even with few tests. While the classic contact tracing is able to slow down the epidemic, BP and SMF algorithms are able to reduce much more the spread of the virus and even stop the propagation, while needing lower number of observations per day. The BP-based method is the most efficient, as it is able to stop the epidemic spreading even with just 1250 observations per day. With higher testing, SMF is able to do the same, although it achieves full suppression of the outbreak at later times. Only in the case with $n_{obs} = 5000$ the Contact Tracing method is able to slowly reduce the number of infected people, instead of just shifting its peak to later times.

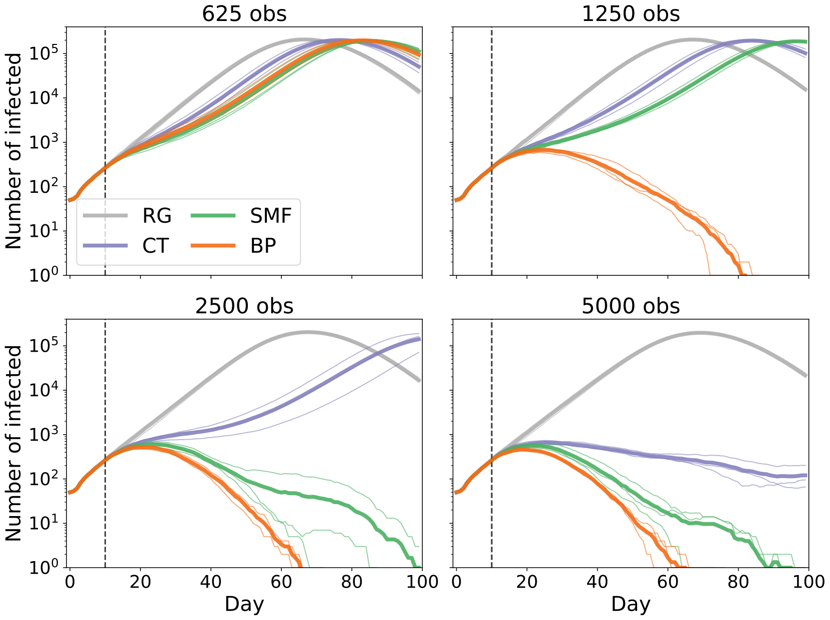

Figure 4.5: **Epidemic containment in the OpenABM model in scenario A** (5×10^5 individuals, $n_{pz} = 50$ and $t_{\text{start}} = 10$), with different values of the number of daily observations n_{obs} Thin lines represent the results for single instances of the epidemics, while the thick lines represent their averages. These four cases are selected within the daily observation range [625,5000] to stress the qualitative differences among the methods. Here only tested positive individuals, and not their cohabitants, are confined. Figure first shown in [118].

It is also possible to imagine an intervention strategy in which, in order to contain more severely the spreading of the disease, when an individual tests as positive for infection, his/her whole household (HH) is confined, too. The results for this more aggressive strategy in scenario A are shown in Figure 4.6. Confining the households enables the SMF-based algorithm to reach the same performance as BP. Also, the Contact Tracing method is able to contain the epidemic with a lower number of daily tests when compared to the previous case (no HH confinement).

The same confinement settings have also been run in scenario B, with household confinement included. This scenario, in which there is a lower number of starting patient zeros, and intervention starts 3 days earlier, might represent cases when an outbreak is discovered in the earlier stages. Results are shown in Figure 4.7, for both household and non-household confinement, and indicate that the epidemic outbreak is easier to contain. In fact, it can be observed that with just 500 daily tests, the BP-based method is able to fully contain the outbreak. The SMF-based algorithm also manages to completely stop the epidemic spreading, but needs double the number of tests. It also manages to do so with significant delay (≈ 20 days) when compared with BP.

Confinement of the households (HH) also has a much higher impact in this scenario. For example, the SMF method is able to contain the epidemic now in the $n_{obs} = 500$ case, while in non-household confinement it is only able to delay the peak of infections. The same happens with CT, but in the higher testing regime with $n_{obs} = 1000$.

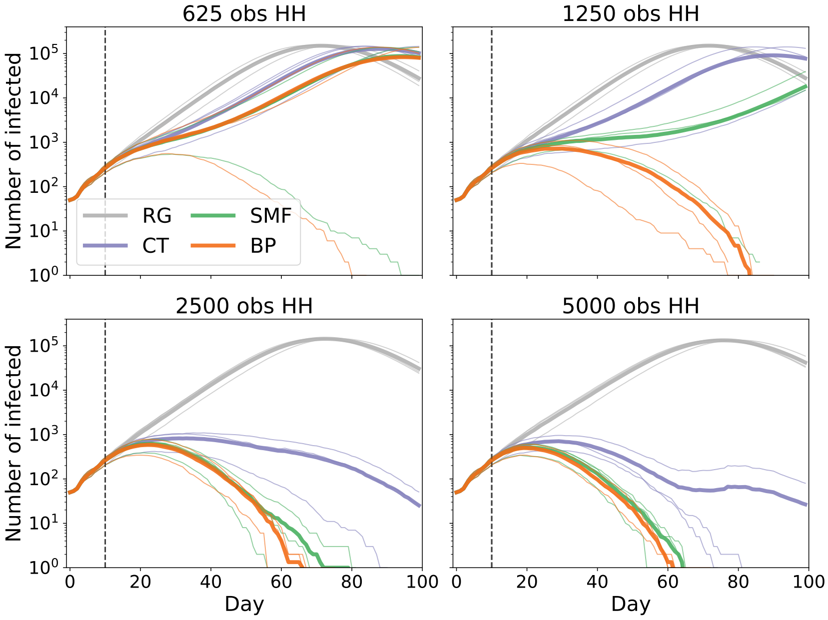

Figure 4.6: **Epidemic containment in OpenABM with household confinement, scenario A**. This figure shows the same experiments as in Figure 4.5, with the only difference being that once an infected individual is found, his/her whole household is confined. Thin lines represent the results for single instances, while the thick lines are the averages. It can be seen here that the confinement of the households helps in slowing down the spread of the virus, enabling (for example) CT to contain the outbreak with a lower number of tests. Figure first shown in [118].

Overall, the results obtained in scenario B confirm the patterns seen in scenario A, with the differences being the number of tests required (1000 daily tests instead of 5000) for MF and BP to both achieve full containment. However, while in scenario A with $n_{obs} = 5000$ the classic contact tracing is able to slow down the spreading, in scenario B with $n_{obs} = 1000$ this is not the case (confinement of the households is needed for CT to reduce the number of infected individuals). Overall, both statistical physics-based methods are able to make better use of the limited number of observations available.

4.4.3 Stability to false negative testing

The results obtained so far reflect the ideal case in which the tests are 100% accurate, however this is often not the case in real scenarios. Let us now consider when the tests are subject to some noise. The testing scenario considered here is the following: while susceptible individuals always test negative for infection, when an individual is infected, it is possible that his/her test turns out negative (susceptible), with a probability (or false negative rate, FNR). These settings correspond to a test specificity of 1 (no false positives) and a sensitivity of 1 − FNR. The reason for this choice is that while a non-negligible FNR means that less infectious individuals are found, and thus there is a higher chance for the outbreak to go out of control, a non-zero false positive rate (FPR) could give the opposite result, as more individuals get tested as infected and quarantined. This makes investigating the case FNR > 0 more interesting than the dual case

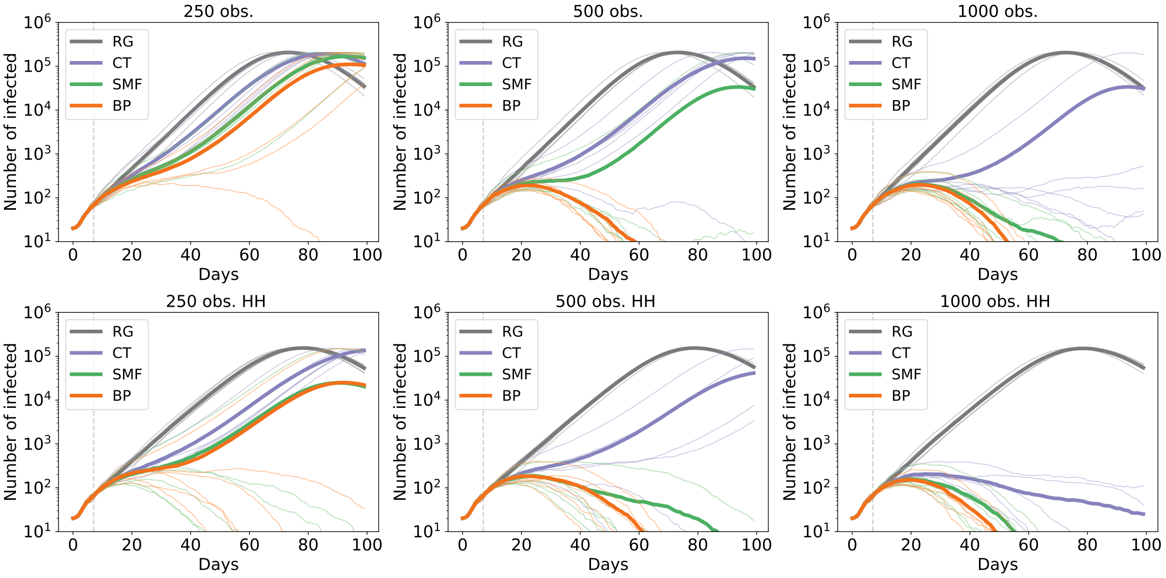

Figure 4.7: **Epidemic containment in the OpenABM model in scenario B** (5×10^5 individuals, $n_{pz} = 20$ and $t_{\text{start}} = 7$), with different values of the number of daily observations n_{obs}. Here both the results with and without confinement of the households are included, the former on the bottom row, the latter on top. Thin lines represent the results for 6 single instances derived by changing the seed of the OpenABM simulation, while the thick lines represent their averages. Figure first shown in [118].

$\text{FPR} > 0$.

While observations might give more information, this setting is the simplest case to test inaccuracies. The BP-based method already includes the possibility of noisy observations in the formulation, but the SMF method does not, in order to keep it as simple as possible. Figure 4.8 shows the results for containment in scenario A ($n_{pz} = 50$, $t_{\text{start}} = 10$), with $n_{obs} = 2500$ observations per day, with different values of the false negative rate, from very low ($\text{FNR} = 0.09$) to significant error rate ($\text{FNR} = 0.4$). It can be seen how the increasing noise makes controlling the epidemic significantly more difficult, as all algorithms (except RG) start by containing the epidemic with lower values of the FNR, and eventually lose the ability of keeping the outbreak in check. The SMF and CT-based methods are able to avoid exponential growth of the infections until $\text{FNR} = 0.19$, while BP manages to do so until $\text{FNR} = 0.31$, losing control at higher values. However, these results indicate how BP-based inference can be very resistant to noise, managing to significantly reduce the number of infected and even completely stop to epidemics even with higher FNR. The same kind of simulations is run with scenario B, obtaining the results in Figure 4.9. The picture is qualitatively the same as scenario A, with SMF and BP performing better in the containment of the epidemic, when compared with classic contact tracing, and BP being the most resistant method to the false negative rates of tests.

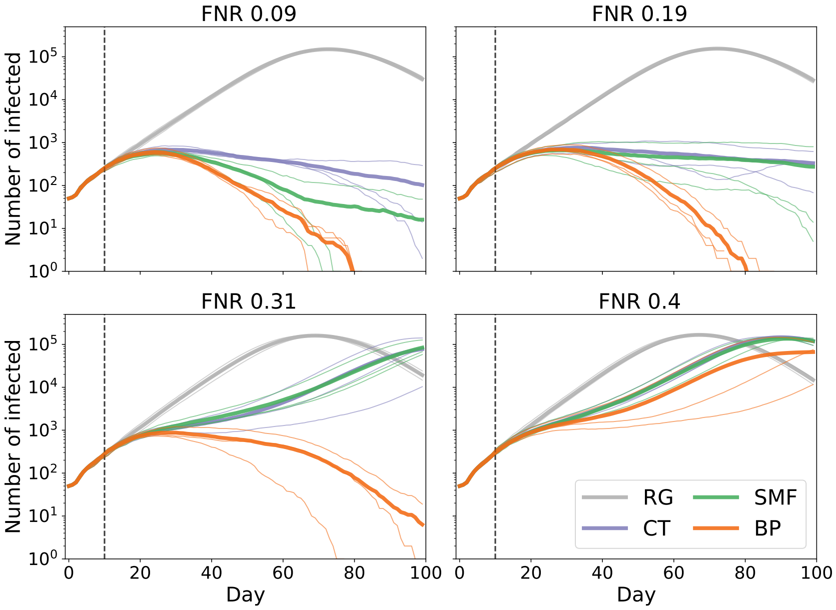

Figure 4.8: **Effect of inaccurate tests in scenario A**. Here, the intervention protocol and the initial conditions are the same as in Figure 4.5 ($n_{pz} = 50$, $t_{\text{start}} = 10$), with 2500 daily tests and household confinement. Containment experiments are shown for 4 different values of the false negative rate (FNR). Thin lines represent single instances, thick lines are the averages. Figure first shown in [118].

4.4.4 Effect of app adoption

Finally, Let us discuss the case when it is not possible to know all of the contacts between the individuals, as a result of the fact that part of the population does not have the possibility to track their contacts trough the smartphone. This might be either because they don't have the smartphone application installed, or they don't own a smartphone. Partial knowledge of the contact network might then hinder the containment of the disease: in fact, when an individual is not taking part in the digital collection of contacts, all of his/her links with others are invisible to a ranking algorithm, like he/she was effectively removed from the network.

This setting is going to be investigated now, by varying the adoption fraction (AF) of the contact collection app in the OpenABM model. Random individuals are then hidden in the contact network, so that only a fraction ρ_{AF} of them, on average, is visible. The adoption fraction is gradually reduced, from 90% to just 60%, resulting in a contact reduction that varies significantly, ranging from 19% (with AF = 0.9) to 64% (with AF = 0.6).

The results for the experiments done in scenario A ($n_{pz} = 50$, $t_{\text{start}} = 10$) are shown in Figure 4.10, and those for scenario B ($n_{pz} = 20$, $t_{\text{start}} = 7$) are in Figure 4.11. Overall, they show that the performance in containment in significantly affected by the adoption fraction. However, while the statistical physics-based method are unable to contain the disease, they manage to delay and lower the peak of the number of infected, way better than the classical contact tracing. This "flattening of the curve" can give several benefits, like having more time to prepare the healthcare system. In scenario B, however, the difference is not as large in scenario A: this could be motivated by the lower number of tests (5 times more in scenario A), indicating that a high

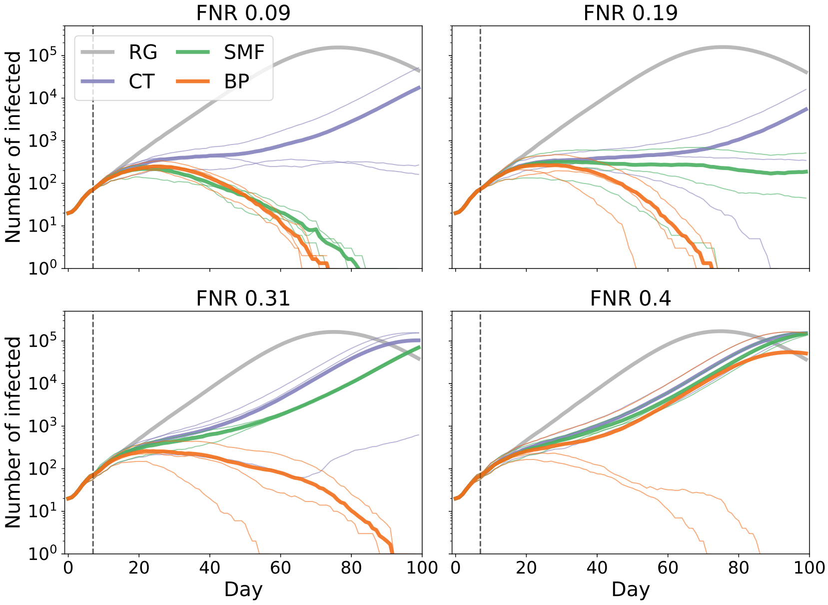

Figure 4.9: **Effect of inaccurate tests in scenario B** Here, the intervention protocol and the initial conditions are the same as in Figure 4.7 ($n_{pz} = 20$, $t_{\text{start}} = 7$), with 1000 daily tests and household confinement. Containment experiments are shown for the same values of the false negative rate (FNR) as in Figure 4.8. Thin lines represent single instances, thick lines are the averages.

testing rate is important when knowledge of the contact network is partial.

4.5 Discussion

In this chapter, it has been shown how the mean field approximations of SIR epidemic spreading on contact networks can be used for the containment of the outbreaks of different diseases. Two methods, with different level of approximation, have been developed: the Simple Mean Field algorithm, and the Belief Propagation-based approximation. While the former is simpler to derive, and faster in practical application, the latter is more accurate, owing both to the approximation of the epidemic process and the Bayesian treatment of the observations.

These probabilistic risk inference methods have been first compared with the classical contact tracing when the spreading follows the same SIR model with homogeneous parameters. In this case, it has been shown that, when it is not possible to conduct exhaustive testing in the population, they are able to slow down the epidemic propagation in a much more effective way than the simple contact tracing. The risk inference methods have also been tested in presence of a mismatch between the spreading and the inference parameters: this analysis showed that both the SMF and BP method are resistant to changes in parameters, with BP being the most resistant of the two.

Then, these methods have been applied to a different spreading model, OpenABM, which has been developed with the aim of describing COVID-19-like epidemic outbreaks using agent-based simulation. It has been shown in the chapter that the disease progression of this model

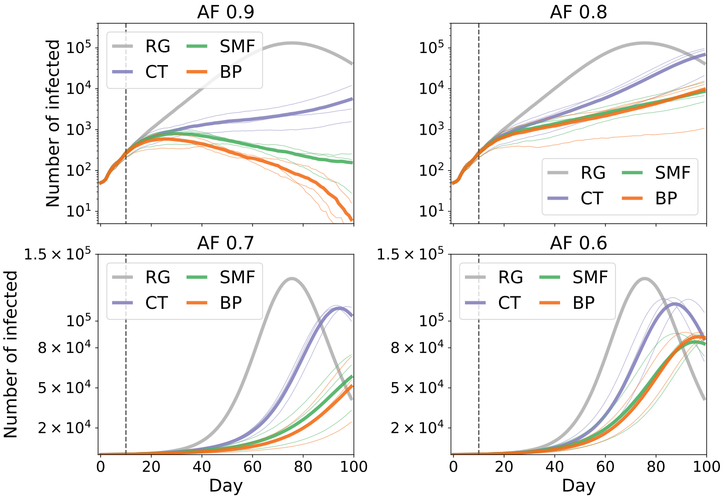

Figure 4.10: **Effect of app adoption in scenario A** The intervention scenario is the same as in figure 4.6 ($n_{pz} = 50$, $t_{\text{start}} = 10$), with the highest number of test (5000) and the confinement of the households. In these experiments the adoption fraction of the contact tracing app is changed, starting from 0.9 until AF = 0.6. The bottom plots are in linear instead of log-linear scale in order to show the difference of BP and SMF with respect to CT. Thin lines represent the 3 single instances, thick lines are the averages.

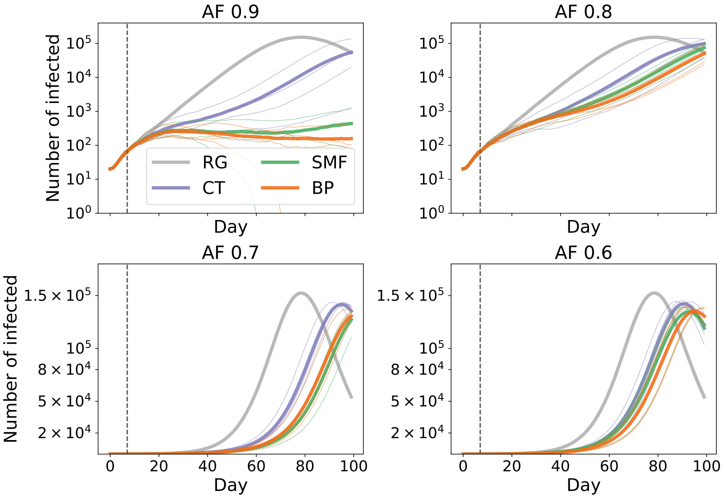

Figure 4.11: **Effect of app adoption in scenario B** The intervention scenario is the same as in figure 4.7 ($n_{pz} = 20$, $t_{\text{start}} = 7$), with the highest number of test (1000) and the confinement of the households. In these experiments the adoption fraction of the contact tracing app is changed, starting from 0.9 until AF = 0.6. The bottom plots are in linear instead of log-linear scale in order to better show the difference between the methods. Thin lines represent the 4 single instances, thick lines are the averages.

has many characteristics that make it very different from an SIR model on a contact network. However, in spite of this mismatch between the spreading model and the one used for inference (which is much simpler), the results obtained show how employing these algorithms provides large improvements over classical contact tracing techniques, when the epidemic is growing and the number of observations is relatively small compared to the population size (which is compatible with real testing capacity of many countries).

Of the two algorithms proposed in the chapter, the most performing is the Belief Propagation-based one. This is mainly because a) the approximation used is more precise than the SMF in inferring the risk of infection (although more suited to sparse contact networks) and b) the treatment of the observations is much more accurate, because of the use of the Bayesian posterior. This last factor is particularly important when the results of the tests is affected by noise, as shown above. The BP method is able to take into account this inaccuracy, while the SMF algorithm is not. Another advantage of BP is that it can deal with non-markovian SIR spreading dynamics; this feature is essential to capture some specific properties observed in the SARS-CoV-2 transmission. It could also be possible to learn the spreading parameters with BP, as done in [38], however this is left for future work.

On the other hand, the SMF algorithm is a simpler approximation of the SIR process, trading accuracy for simplicity and ease of computation. When the number of daily observations is high enough, it can achieve the same containment efficacy of BP (while still performing significantly better than classical contact tracing). It is, however, quite affected by testing noise.

Another inherent advantage of obtaining a good infection risk estimate, as provided by these mean-field algorithms, is the possibility of implementing a diversified intervention strategy, for example giving different suggestions to people (reduction of contacts, self-isolation, self-testing) according to some thresholds of the individual's risk value.

On a more computational standpoint, the deployment of these methods in a smartphone application would require each individual sending and receiving messages with other individuals. We remark that the volume of daily exchanged messages per pair of individuals in the two proposed methods is constant with respect to both the population size and time: in particular, for the BP approach this is achieved by defining a fixed time-window over which the inference step is carried out, as discussed in Section 4.1.2. A rough estimation of this volume gives about 1kB for MF and 1MB for BP per individual on each day (assuming ~ 10 daily contacts): this volume is very small when compared with normal data usage, and the computational load on an a user's phone would be minimal. With regard to privacy, it is worth emphasizing that the proposed inference methods are in principle more protective than the manual tracing. In fact, by identifying individuals who have the largest probability of being infected through a cumulative process by which information is integrated, the direct attribution of potential infection events to a given individual is made much harder. Moreover, the messages can be distributed and encrypted, eliminating the need to upload the contacts on a public server. Details of such fully privacy preserving implementation, along the lines of [111], are left for future work. Implementation of the SMF and BP risk estimation algorithms and all the tests that follow can be found at [88].

Both risk inference methods proposed in the chapter are severely affected by the lack of knowledge of the contact network: when individuals do not install the contact collection app, their contacts cannot be identified. This makes the risk inference less effective, leading in turn to a worse mitigation of the epidemic. Despite the fact that a delay of the epidemic peak can be observed even at smaller values of the adoption fraction, previously deployed applications for COVID-19 tracing still had adoption levels insufficient for these methods to be effective in containing the epidemic spreading [113, 154].

Finally, it is important to consider that while the agent-based model considered in this chapter is very detailed from the point of view of the disease progression, it exploits synthetic contacts networks for the disease transmission from person to person. The resulting interaction dynamics within the population are however quite realistic, due to the many networks of contacts (household, workplace, and random) and to the calibration of the model. In order to show the performance of the risk inference methods and interventions on real world contact networks, it could be possible to perform experiments with virtual epidemics, running on volunteers' smartphones, virtually confining individuals as the (virtual) disease spreads. Using the collected contacts a posteriori, instead, would make the obtained epidemic outbreaks less realistic, as applying interventions would change the interaction dynamics. This kind of virtual epidemic experiments has been already performed: some notable examples include the BBC Four Pandemic [93] and Operation Outbreak [170].

Chapter 5

Capturing epidemic features of COVID-19 spreading

In the previous chapter it has been shown that risk inference methods derived from statistical physics can be used to mitigate epidemics, and can also be applied to models for the spreading of COVID-19. Then, it is important to understand how these methods can be helpful in the reconstruction of the features of the spreading of the disease. Moreover, it is interesting to compare them with more evolved forms of contact tracing that employ the informations of the tests in a smarter way.

Also, one limitation of the work of previous chapter is that, with respect to COVID-19 epidemic spreading, only one model has been considered, which has its own set of assumptions in the contact model and transmission of the disease. In this chapter, two additional models will be used to test the inference methods in these diverse settings. As will be explained in the following, the models taken under consideration use very different assumptions for the spreading of the disease and the infection dynamics.

5.1 Motivation

Much work has been done recently on investigating the effectiveness of digital contact tracing (DCT), which is supplementing and replacing ”classical” manual contact tracing (MCT) using smartphone applications. In fact, early computational models and theoretical results on DCT have shown that it alone can ensure epidemic containment in large populations only in the presence of high adoption [85, 94, 123, 167, 141]. Despite some controversial results [168], the analysis of data obtained from early implementations of DCT applications has revealed that they really contribute to epidemic containment, providing an additional quantitative and qualitative gain over MCT [105, 155, 148, 140, 139]. However, massive contact tracing comes with the risk of the proliferation of exposure notifications and quarantines, which may result in high cost-to-benefit ratio [75, 117, 119, 123, 125]. This is certainly the case, as the size of the outbreak grows, when notification of exposure is provided as a consequence of every single contact with tested positive individuals, regardless of the proximity or the duration of the contact, or the viral load of the infected individual. In fact, current DCT implementations, such as Corona-Warn-App [82] based on the Google/Apple API protocol [73], can use such additional information to compute risk levels and provide more individually-tailored tiered notifications. This is a first step towards improving the efficacy of DCT and reducing notification redundancy. A major improvement in this direction would be represented by inference-based contact tracing methods, which could naturally account for multiple exposures [71, 128, 118, 142]. Using a Bayesian framework to include all the available observations of positive (and negative) tested individuals, an efficient and distributed method based on Belief Propagation has been put forward, as shown in Chapter 4, to compute the individual probabilities of being infected that contribute to the individual risk levels provided on the app.

Due to the peculiarity of COVID-19 of displaying overdispersion in secondary infections [29, 116, 84, 95, 91, 72, 152, 136], it was claimed that backward tracing can significantly increase the number of traceable individuals compared to forward tracing [112, 122], because positive tested individuals are more likely to come from contagion clusters than to generate them [135]. Countries like Japan [100], South Korea [97] and Uruguay [107] are credited with successfully implementing backward tracing in their contact tracing campaigns. This reveals a second limitation of current app-based digital contact tracing implementations, as they primarily perform forward tracing [127], meaning that the individuals being tracked are almost only those who could have been infected by a tested positive individual. Innovative digital contact tracing methods based on statistical inference, such as the one shown in the previous chapter, which ground their predictive power on reconstructing causal relationships in transmission paths [37], are instead expected to more efficiently discover backward traces and capture super-spreading events. This is particularly effective when a possibly large number of cheap, low-sensitivity rapid tests is available [132], as the prior information about the sensitivity of the tests can be included in the Bayesian probabilistic approach [118].

5.2 Agent-based models for COVID-19

In the realm of mathematical models for epidemic spreading, agent-based models are able to provide an individual-level description of the spreading of diseases, with sufficiently reliable representations of the actual contact networks on which contagion between individuals could take place. Among the agent-based models proposed since the beginning of the COVID-19 pandemic [70, 90, 131, 134, 162, 161], some of them can be considered exemplary for formulating a critical analysis of the containment capabilities of the different contact tracing methods and evaluate their cost-to-benefit ratio. In this chapter, apart from the already presented OpenABM model (see section 4.3), we will employ the Covasim model by Kerr et al. [134], and the Spatiotemporal Epidemic Model (StEM) by Lorch et al. [162].

All the models considered are derived from the Susceptible-Exposed-Infected-Recovered (SEIR) model, in which several states are included to account for various stages of infectiousness and disease severity. All the models considered are capable to reproduce the empirically observed non-Poissonian statistics and overdispersion in contacts patterns and individual viral loads. Since the OpenABM model has already been described in detail, in the following the other two models will be described.

5.2.1 Spatiotemporal Epidemic Model

The epidemic simulator put forward in Lorch et al. [162] consists of two interconnected processes: a mobility simulation, in which the individuals can visit a number of sites and come in contact with each other by being in the same place at the same time, and a proper epidemic simulation, in which the virus spreads through the population due to some initially infected individuals and the aforementioned contacts.

In this model, particular attention is devoted to producing mobility data that are as realistic as possible. Once a city has been chosen, the first step of the mobility generator consists in retrieving the population data, i.e. the population density (when publicly available, otherwise individuals are located uniformly on the map), the age groups, and household composition distributions of the overall population. These quantities are used to constitute and locate the households on a map. The next step is to model the sites that the individuals can visit during the mobility simulation. These venues belong to one of the following categories: education (e.g. schools, universities), social activities (e.g. bars, restaurants, cafés), business (e.g. offices, shops), supermarkets, and bus stops. Similarly to the households assignment, the number of sites of each type and the location of each site are decided using the real distribution of these venues in the city: this is accomplished using OpenStreetMap [165], which is able to provide geolocation data for each of these place categories. The visits to specific places are then modelled assuming that people visit only a modest subset of the possible venues, with a probability that decreases with the site-household distances.

The epidemic model used is a generalization of the SEIR model in continuous time, with multiple infected states to account for the possibility of distinguishing among pre-symptomatic, symptomatic, and asymptomatic individuals. The state evolution is modelled as a set of counting processes and simulated using stochastic differential equations (SDE) with jumps (as this dynamic requires discrete state transitions in continuous time). Only exposure events exploit the mobility data by taking into account the concurrent presence of an infector in a pre-symptomatic, symptomatic or asymptomatic state, and a susceptible individual in a certain venue. The exposure counting process considers, together with a transmission rate that depends on the state of the infector, a venue-dependent exposure rate and a kernel term that quantifies possible environmental transmissions (due to the presence of the virus in the air or on the surfaces). It is worth noting that the presence of super-spreaders naturally occurs in the dynamic, indeed, the number of transmissions is characterized by an over-dispersed distribution (see Figure 5.5).

The other events are individual-dependent with log-normally distributed typical transition times, whose parameters reflect the data extracted from clinical COVID-19 literature.

For sake of simplicity, our simulations exploit daily mobility and epidemiological dynamics data to build a test-and-isolate intervention measurement similar to what has been presented in the previous chapter. The code used for these simulations is reported in the GitHub fork [159].

5.2.2 The Covasim model

The Covasim model [134] is a special variant of the SEIR model in discrete time, with several states dividing infectious individuals in asymptomatic (AS) and symptomatic categories (the latter is subdivided into pre-symptomatic (PS), mildly (MS), severely symptomatic (SS) and critical (C), see Figure 5.1). The death of individuals is also accounted as a state which is reachable when the individual is in the critical stage of the disease. The transition times between the aforementioned states are drawn from log-normal distributions with different parameters (described in detail in [134]).

The model can be used with several types of contact networks, and the contacts can be divided into different layers. As default, the simulations run on simple Erdös-Rényi (ER) random graph [62] with fixed average degree. However, as done in [133], in all the simulations conducted in this work the Synthpops package (described in [134]) is used in conjunction with King County epidemic data to obtain realistic contact networks. With this procedure, several contact networks are randomly generated as ER graphs, each representing different environments (houses, workplaces, schools, and community spaces) and then modified to accounting for statistics such as employment age in schools and workplaces. Also, a network representing long-term care facilities (LTCF) is created by randomly sampling elderly people, whose interactions are removed from the houses and workplaces networks, and younger workers. All of the contact networks generated in this way are static, but during the simulations, the community contacts are updated each day. Also, we limit the size of the generated population to either 50,000 or 70,000 people for practical reasons.

In the Covasim model, it's possible to simulate epidemics for a very large population by employing its ”dynamic rescaling” feature, as done in [133]. With this technique, a single infected agent represents more than one individual, this number varying during the simulation, depending on the number of infected agents [134]. This features is turned off in our simulations, as it is not relevant for the investigation of individual contact tracing techniques. Also, epidemic parameters for the model are taken from [133], where they have been derived using King County epidemic data.

Pre-emptive quarantine

It is important to note that in the Covasim model individuals who are notified of a known infected contact enter into a pre-emptive quarantine state that is accompanied by a reduction in the infectivity (both for the source and target individuals of a possible infection) of $1 - f_Q$ (so that with $f_Q = 0.2$ we have a reduction of 80%).

This intervention thus does not take into account whether an individual is infected or not, and it ends when the individual is tested for infection: if the result is positive, he/she is put into full isolation, otherwise he/she is released from quarantine. However, if an individual is not tested in the following 14 days after notification, he/she is released from quarantine, too.

In practice, while f_Q could be different for different contact layers (e.g. higher for school and workplace networks, and lower for the house), in this chapter it will be fixed and equal among all contacts, for the sake of simplicity.

Infection modelling

The infectivity of single contacts in the Covasim model depends on a series of factors given by global parameters, the status of the individual (including quarantine/isolation), individual parameters (relative transmissibility T_{rel} and susceptibility S_{rel}, for example) and the contact layer. The exact formula giving the probability of infection from a source infectious individual i to target susceptible individual j is:

$$\lambda_{i\to j} = \beta_g \beta_c \beta_{lay} T^{(i)}_{\mathrm{rel}} V_{\mathrm{load}} \left(t - t^{(i)}_{\mathrm{inf}}\right) M^{(i)}_Q M^{(i)}_{\mathrm{Asy}} M^{(i)}_{\mathrm{Iso}} S^{(j)}_{\mathrm{rel}} M^{(j)}_Q \left(1 - I^{(j)}\right) \tag{5.1}$$

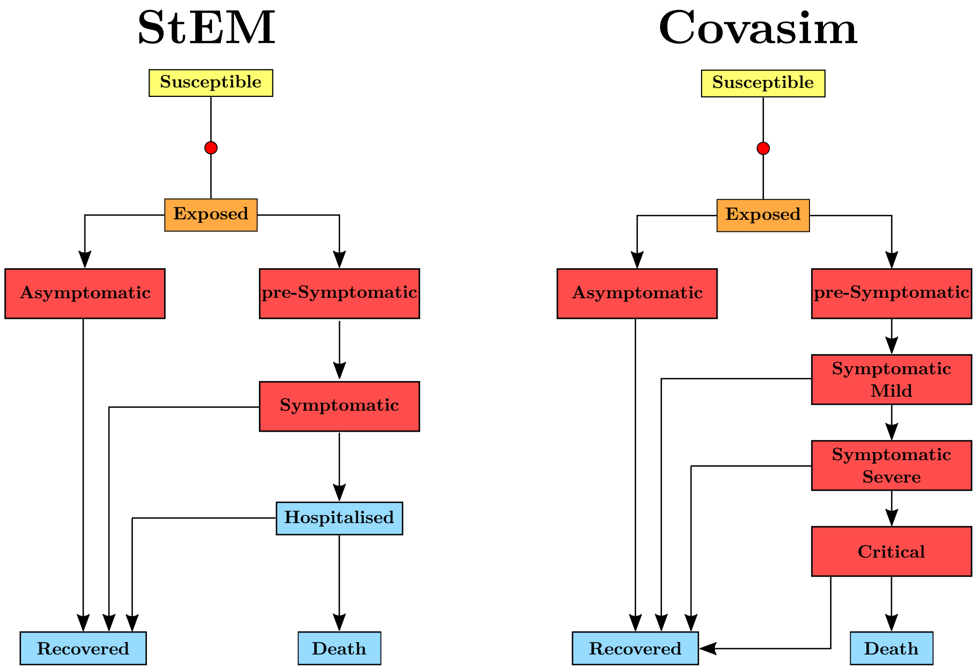

Figure 5.1: **Schematic representation of the epidemic dynamics** in the new models considered in this chapter. Colour codes identify how these states are considered with respect to the SEIR model: red represents Infectious and light blue stands for Recovered. Infection of the individual is graphically represented by a red dot, and the probability to undergo one of the different infection pathways is always age-dependent on the three models.

Where β_g is the global β given to the simulation, β_{lay} varies with the contact layer (household, long-term care facilities, work, and schools contacts being respectively 10 times, 5 times, 2 times more infectious than community contacts, as done in [133]), β_c is the strength of each contact, I is the immunity level of the target individual. The M factors model the effect of the state of an individual on the infectivity: if individual i is in quarantine, then $M_Q^{(i)} = f_Q$, the quarantine factor, otherwise $M_Q^{(i)} = 1$. The same logic is applied to $M_{\text{Iso}}^{(i)}$ and $M_{\text{Asy}}^{(i)}$, which are different than 1 when individual i is in isolation (following diagnosis) or is asymptomatic, with factor parameters f_{Iso} and f_{Asym}, respectively. The T_{rel} and S_{rel} parameters are fixed during the simulations, and the former is used as the *default viral load* as explained in [134](see panel (b) of figure 5.5 for the distribution of T_{rel}). In all the simulations $f_{\text{Asym}} = 1$, while both f_{Iso}, which models the reduction of infectiousness after isolation of an individual, and f_Q are fixed to $f_{Iso} = 0$ and $f_Q = 0.6$ in the containment experiments. During all the simulations, also, the immunity levels stay fixed once an individual recovers from the disease so that it is impossible to have reinfections.

Viral load depends on the time elapsed since infection: $V_{\text{load}}(t - t_{\text{inf}})$, taking value only in two different levels: in fact, it falls to half its initial value after a number of days corresponding to 30% of the individual duration time of the disease or 4 days, whichever is shorter[134].

5.3 Mapping of the models into SIR parameters

In order to apply the Simple-Mean-Field (SMF) and Belief Propagation (BP) methods for the inference of the risk of infection, equivalent SIR model parameters have to be derived for each agent-based model. In the following, it will be shown how this procedure was carried out in the two models.

5.3.1 Recovery probability

Let us begin from the Covasim model. Since in this model there are Exposed and Infectious states, they will both be mapped to the Infected state of the equivalent SIR model. Therefore, the recovery time of the SIR, that is the time it takes to pass from the Infected to the Recovered state, corresponds in the Covasim model to the total delay from becoming Exposed (E) to entering either the Recovered (R) or the Dead (D) state. This means that all the possible paths for the disease progression need to be considered for the calculation. Each path will reach the R or D state with a different delay, and that can be computed by summing the transition times according to the disease progression (shown in Figure 5.1):

$$\tau_r^A = \tau_{E\to AS} + \tau_{AS\to R} \tag{5.2a}$$

$$\tau_r^M = \tau_{E\to PS} + \tau_{PS\to MS} + \tau_{MS\to R} \tag{5.2b}$$

$$\tau_r^S = \tau_{E\to PS} + \tau_{PS\to MS} + \tau_{MS\to SS} + \tau_{SS\to R} \tag{5.2c}$$

$$\tau_r^C = \tau_{E\to PS} + \tau_{PS\to MS} + \tau_{MS\to SS} + \tau_{SS\to C} + \tau_{C\to R} \tag{5.2d}$$

$$\tau_d^C = \tau_{E\to PS} + \tau_{PS\to MS} + \tau_{MS\to SS} + \tau_{SS\to C} + \tau_{C\to D} \tag{5.2e}$$

where the + here indicates the sum of random variables, with each variable representing the delay of the corresponding state transition (*AS* for the asymptomatic state, *PS* for pre-symptomatic, *MS* for mild symptomatic, *SS* for severe symptomatic and *C* for critical). Then, in order to obtain an equivalent distribution for the SIR time of recovery, each path needs to be weighted by the probability of an individual having that disease progression, and since the path probabilities depend on the age of the individual, the results needs to be averaged on the age distribution $p(a)$. The final distribution is then computed as:

$$\begin{aligned} p(\tau_R) = \Big\langle & \phi_a p\left(\tau_r^A\right) + (1-\phi_a)\,\phi_R^{SM} p\left(\tau_r^M\right) + \\ & + (1-\phi_a)\Big[\left(1-\phi_R^{SM}\right)\phi_R^{SS} p\left(\tau_r^S\right) + \left(1-\phi_R^{SM}\right)\left(1-\phi_R^{SS}\right)\phi_R^C p\left(\tau_r^C\right) + \\ & + \left(1-\phi_R^{SM}\right)\left(1-\phi_R^{SS}\right)\left(1-\phi_R^C\right) p\left(\tau_d^C\right)\Big]\Big\rangle_{p(a)} \end{aligned} \tag{5.3}$$

The distribution obtained in this way can be applied to a non-Markovian SIR model. This is supported by the BP method, and for simplicity reasons the the $p(\tau_R)$ is passed to the method by fitting it to a gamma distribution. Overall, the distribution of τ_R has mean $\langle\tau_R\rangle = 14$ and standard deviation $\sigma_R = 2.37$. The SMF method instead relies on the Markovian SIR model, thus needing a (time-independent) recovery probability μ_R. The value of this probability is obtained by a geometric distribution fit of $p(\tau_R)$ (resulting in the estimate $\mu_R = 0.071$).

The same procedure is applied to the StEM model to obtain the distribution of the recovery delay $p(\tau_R)$. In this case, however, there is also the Hospitalised state, in which an individual is already unable to transmit the virus to others, thus making it equivalent to the Recovered state in the SIR model (see left pane of Figure 5.1). The recovery delay distribution obtained has $\langle \tau_R \rangle = 18.7$ and $\sigma_R = 2.37$, while the equivalent recovery probability for the SMF method is $\mu_R = 0.083$. The recovery delay distributions obtained in Covasim and StEM are shown in the right pane of Figure 5.2, along with the one from the OpenABM model.

5.3.2 Infection probability

In order to give an estimate to the algorithms of the infection probability in Covasim, the parameters shown in Eq. 5.1 have to be given to the ranking algorithms. The factors β_c, β_{lay} and the M_Q, M_{Iso} and M_{Asy} modifiers will be given as weight $w^t_{i\to j}$ of each contact. This leaves β_g, that is a global simulation parameter. The $S^{(j)}_{rel}$ and $T^{(i)}_{rel}$ depend on the individual, so they cannot be known in a contact tracing scenario. Therefore, it is necessary to give an estimate based on their distribution, obtained from a SynthPops population of $N = 225 \times 10^3$ individuals generated as described in section 5.2.2. For the $S^{(j)}_{rel}$, since the distribution obtained is fairly uniform, the average value is given to the contact tracing methods (which is ≈ 1). For the $T^{(i)}_{rel}$, however, in order to take into account the skewness of the distribution (see panel b of figure 5.5), it's necessary to give the median $T^M_{rel} = 0.41$, instead of the average $\langle T_{\text{rel}} \rangle \approx 1$, to the algorithms.

It's possible to account for the time dependence of the V_{load} factor only in the BP algorithm, that is able to work with a non-Markovian SIR Model. This information is also combined with the waiting time for the transition $E \to I$, as in the E state in individual is already infected. These two temporal effects are modeled with two sigmoid functions: the first one is fitted on the distribution of the $E \to I$ transition time, and the second one on the distribution of the time of the switch from the high value V_H to the low value V_L of the V_{load}.

Therefore, the probability of infection for each contact for the contact tracing methods can be written as:

$$\lambda^{\text{BP}}_{i\to j}(t) = \lambda^{\text{BP}}_0 \Lambda\left(t - t^{(i)}_{\text{inf}}\right) w^t_{i\to j} = \beta_g T^M_{rel} \Lambda\left(t - t^{(i)}_{\text{inf}}\right) w^t_{i\to j} \tag{5.4}$$

with $w^t_{i\to j}$ being the weight of each contact that is given by the model, and

$$\Lambda(t) = V_H S(g_H t - t_H) + (V_L - V_H) S(g_L t - t_L) \tag{5.5}$$

with S the sigmoid function, $g_H = 1.21$, $t_H = 4.71$, $g_L = 1.05$ and $t_L = 7.28$. For the SMF algorithm, instead,

$$\lambda^{\text{MF}}_{i\to j}(t) = \lambda^{\text{MF}}_0 w^t_{i\to j}$$

where λ^{MF}_0 is the average of all the factors in Eq. (5.4) excluding the weight $w_{i\to j}$.

In the StEM model, the probability of an infectious individual transmitting the disease to a susceptible one depends on the time of residence in one of the mobility sites [162]. In this case, no evident weight of each contact can be passed to the ranker, so the expression of the infectiousness reduces to the prefactor (λ^{BP}_0 or λ^{MF}_0). The distribution of the time of the $E \to I$ state transition can be given to the BP method as

$$\Lambda(t) = S(t - t_0),$$

with $t_0 = 3.22$, and the values of the prefactors are found through averages obtained on a typical realization of the epidemic dynamics with a large population size.

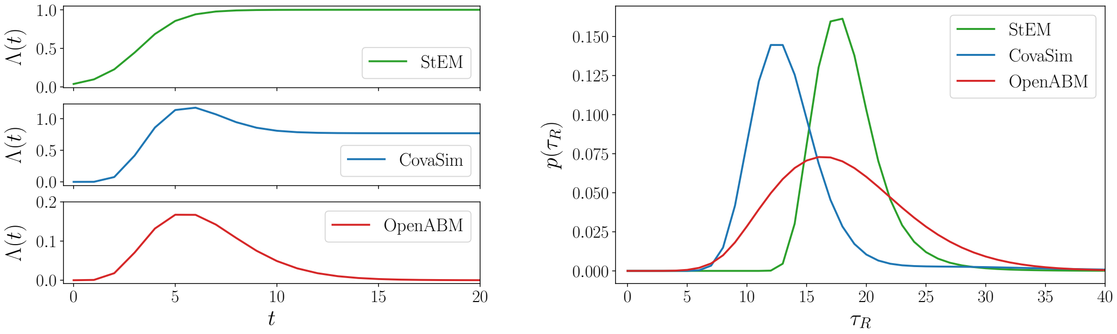

Figure 5.2: Left: Time-dependent contribution to the infection probability $\Lambda(t)$ for Covasim and StEM, where t denotes the time elapsed since infection. Right: probability distribution of the recovery time in the effective SIR reduction for the agent-based models. Both time-scales are in days. The same curves are also shown for the OpenABM model for reference.

5.4 Epidemic containment in Covasim and StEM

We now discuss how to perform epidemic containment with the Covasim and StEM models, as has been shown in the previous chapter with OpenABM. As previously explained, the aim is to show how the methods derived from statistical physics, Simple-Mean-Field (SMF), and Belief Propagation (BP), perform in the containment of epidemic cascades generated with the two agent-based models. These methods have been derived in Sections 4.1.1 and 4.1.2 of the previous chapter, respectively, and will be compared with two different implementations of digital contact tracing: the first is a naive form, called simply DCT, in which all people that have been in contact with a confirmed infectious individual are given the same rank; the second one, called informed contact tracing (ICT), takes into account all the test results, both positive and negative.

Then, during a simulation run of each model, each algorithm is run every day (after an initial waiting period of t_{st} days) in order to produce a ranking of all individuals on the risk of being infected. As previously done with the OpenABM model, on each run of the method the results of the tests done in the previous days are available to the ranking method as observations, along with the contacts up to the current day. These contacts are supposed to be collected with a contact detection app, running on each individual smartphone. Only when both individuals have this app, the contact is detected and given to the algorithms. After the ranking has been obtained, a number of tests N_{test} is performed on the individuals who have the highest risk of being infected according to the algorithm, and who haven't previously found infected. Also, each day some of the symptomatic individuals are tested, with probability p_{sympt}. Positively tested individuals enter isolation, thus avoiding infecting others.

In this containment protocol, the tests results are affected by an error due to a non-zero false-negative rate v_n. A relatively large value of false-negative rate can be associated with the outcome of rapid COVID-19 tests which provides an inaccurate, but fast and low-cost, way to ascertain contagion. The value, which is set to of $v_n = 0.285$ in the present simulations, is a rough estimate computed from data published in [126, 129].

A standard testing strategy, applicable to all contact-tracing methods, requires that the number of tests suggested by each ranker, and performed on a daily basis, is fixed at $N_{\text{test}} = 400$. An alternative testing strategy can be employed with probabilistic contact-tracing methods, such as BP and SMF, and consists in observing the state of individuals whose estimated probability of being infected is larger than a threshold value p_{th} ($p_{th} = 5 \times 10^{-4}$ for StEM and 1.5×10^{-2} or 2.5×10^{-2} for Covasim). As a consequence, in this latter case, the number of ranker-based tests changes adaptively in time.

For the Covasim model, a more effective version of manual contact tracing is available, called Test-Trace-Quarantine (TTQ), which has been employed in [133]. This method exploits the pre-emptive quarantine (PQ) available in Covasim, so that when manual contact tracing is performed, the individuals reached by the MCT enter PQ, awaiting being tested. In the TTQ strategy, then, individuals are tested with a probability that depends on their quarantine status, and also by their symptoms: symptomatic individuals have higher probability of being tested with respect to non-symptomatic ones, and those who have just entered PQ also have higher probability than those who are not quarantined or are already in quarantine. Then, the number of tests performed each day with the TTQ strategy is not fixed: in order to compare it with the rankers-based testing, a version with fixed number of tests, called TTQ-N, has been created where the individuals to be tested are sampled according to their probabilities. A first set of preliminary results is shown in Figure 5.3, for the case when all the population participates in the collection of the contacts (i.e. when the adoption fraction (AF) of the contact collection app is 100%). While these results are not definitive, they indicate that probabilistic inference algorithms are able to contain the epidemic cascades better then the two version of the contact tracing used (DCT and ICT). In the StEM model, the BP method is the only one able to achieve an effective reproduction number R_t below 1, and its threshold testing version (indicated as BP(p_{th})) is able to stop the epidemic outbreaks earlier and with a lower number of total tests. In the Covasim model, both SMF and BP are able to stop the epidemic spreading before the final time. Threshold testing with these two algorithms then leads to a degradation of the containment performance, but with a much lower number of tests. While the SMF(p_{th}) method consumes less tests than BP(p_{th}), it is able to contain much less, too, while the performance of BP(p_{th}) is close to its fixed daily test version.

The same containment framework is also tested when the adoption fraction (AF) of the contact collection app is 80%. In this case, the contact tracing methods have partial knowledge of the contact network, and it becomes harder to contain the epidemic outbreaks, as shown in Figure 5.4. As a consequence, it can be seen that now the threshold testing version of BP requires many more tests to keep the epidemic outbreak in check in the StEM model. Instead, on the Covasim model both BP and SMF show a qualitatively better containment than contact tracing methods or TTQ, with their threshold testing version requiring again less tests but with a smaller performance gap than the AF = 100% case.

Overall, the probabilistic risk evaluation methods seem more effective on the Covasim model than on the StEM: this could be due, among other factors, on the structure of the contact networks created by the models. In fact, while the Covasim model generates simple (albeit slightly modified) random graphs, the contact graphs generated by the mobility simulations in StEM are much more complex, making the reconstruction of the epidemic spreading more difficult. Also, it's clear that both Digital Contact Tracing and Informed Contact Tracing do not perform well in all the simulations shown, and this is also due to the fact that they are easily fooled by the false negatives in testing.

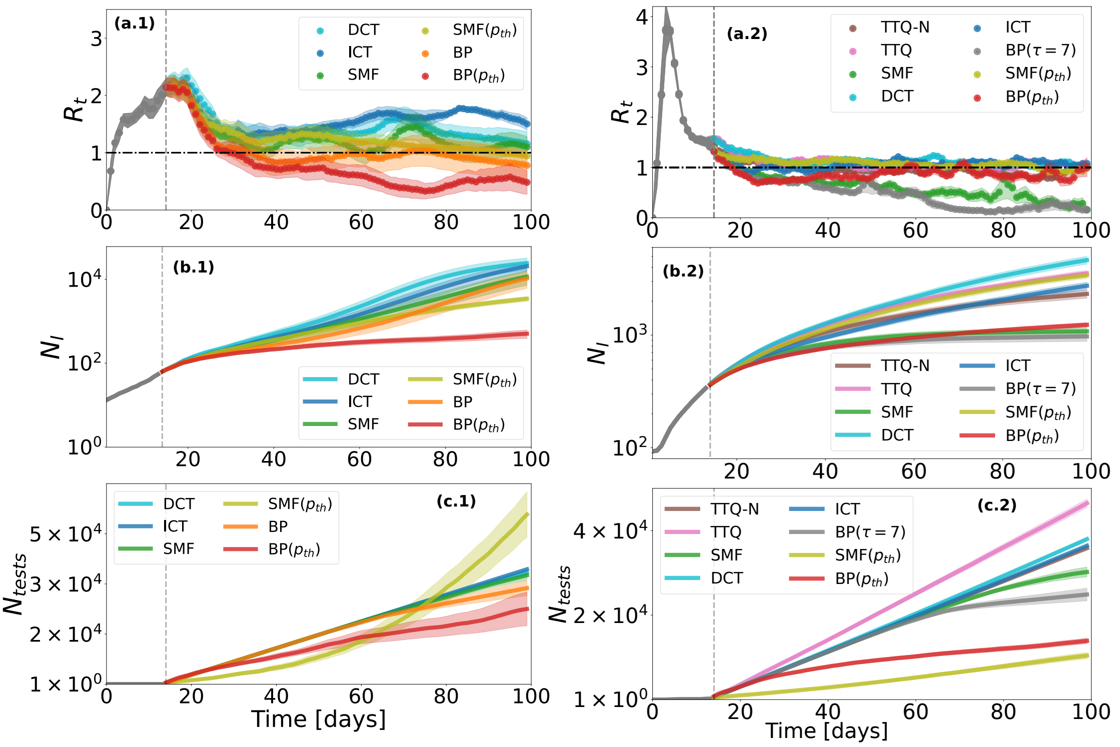

Figure 5.3: **Epidemic mitigation results for StEM and Covasim for 100% app adoption**: for each model (left for StEM, right for Covasim) the effective reproduction number R_t, the cumulative number of infected individuals N_I and the cumulative number of tests done N_{tests} are reported. For the StEM model, the simulations are done on the city of Tübingen, with a population of $N \approx 90\,000$ individuals. The number of patient zeros N_{pz} is 12 (3 in the exposed state, 4 in the asymptomatic state, and 5 are pre-symptomatic individuals). A large fraction of the symptomatic individuals is observed on a daily basis ($p_{sym} = 80\%$). In StEM, households are confined whenever a member is tested positive. For the Covasim model, the population size is $N = 70\,000$, and there are $N_{pz} = 66$ patient zeros, all starting in the exposed state. Also, a fraction $p_{sym} = 50\%$ of the symptomatic individuals is observed each day. Lines reflect the average behaviour of the metrics computed from a number of realizations (10 for StEM, 20 for Covasim) of the simulations while error bars report the associated standard error. The p_{th} indicates that threshold testing is employed ($p_{th} = 5{\times}10^{-4}$ for StEM and $2.5{\times}10^{-2}$ for Covasim). The BP results for Covasim have been obtained while computing the risk of infection as the sum of the infection time beliefs in the last $\tau = 7$ days (see appendix B).

5.5 Overdispersion and super-spreaders

This section is devoted to showing how probabilistic-based tracing methods encompass the ability to effectively detect super-spreaders. For sake of clarity, and because of the diversity of the agent-based models used as proving ground, it is first necessary to precisely define what is meant by super-spreading transmission in this work.

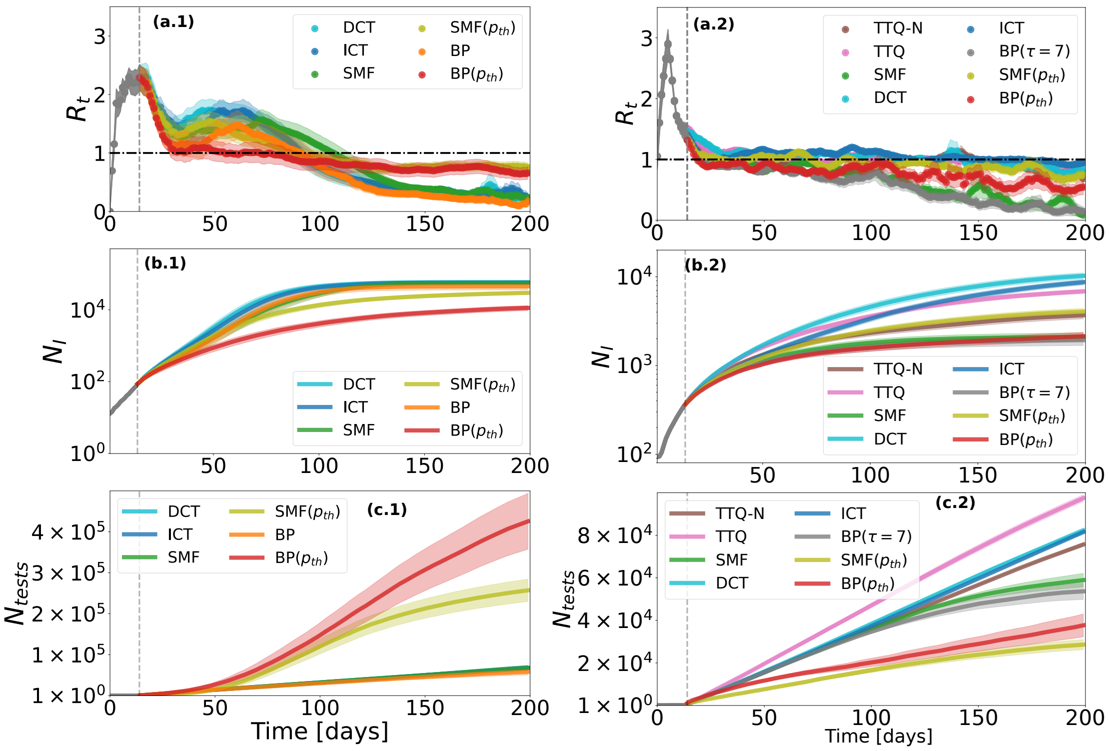

Figure 5.4: **Epidemic mitigation results for StEM and Covasim for 80% app adoption**. For each model (left for StEM, right for Covasim) the effective reproduction number R_t, the cumulative number of infected individuals N_I and the cumulative number of tests done N_{tests} are reported. The parameters of the simulations are the same as in Figure 5.3, with the difference that now the fraction of people using the contact collection app is just 80%. Also, in this case, threshold probabilities for testing are adjusted for the Covasim model ($p_{th} = 1.5 \times 10^{-2}$).

In OpenABM [131], super-spreading events arise as a result of over-dispersed contacts in one of the three artificial contact networks used to model interactions within a population. Similarly, in StEM [162], over-dispersion arises naturally from the contact graph, as a result of the realistic mobility simulation that exploits geolocalized data within a city. In both cases, the empirical distribution of the number of infections shows a significant non-Poissonian behavior, i.e. with a variance to mean ratio (VMR) larger than one (see Figure 5.5). For these two models, individuals infecting at least 7 contacts within their infectious time window are identified as super-spreaders, in accordance with the definition provided in Wong and Collins [115]. Conversely, in Covasim [134], the over-dispersion of infections arises directly by virtue of the individual's viral load, which is drawn from a fat-tailed distribution (see Figure 5.5b): super-spreaders can therefore be identified by looking at the individual relative transmission intensity T_{rel} (see equation 5.1), a quenched parameter not accessible to the tracing methods. In particular, in each simulation, individuals displaying $T_{\text{rel}} \geq 5$ were classified as super-spreaders.

The ability of the different contact tracing methods to detect super-spreaders among the infected individuals was measured according to the following numerical experiment. For each epidemic realization, the propagation was left free to evolve without intervention up to a time T, whereupon the contact tracing methods are applied only once and the corresponding ranking of potentially infected individuals is collected. The value of T is here chosen of the order of few weeks, i.e. the typical time-window for which contact information can be retained in the digital contact tracing applications [118]. To gain some information about the infectious transmissions, a fixed fraction of the symptomatic individuals was observed daily (see caption of Figure 5.6 for additional details). Individuals identified by means of the different contact tracing methods, and ranked according to their epidemic risk, are then classified according to their real state of infection, obtaining corresponding ROC curves. To fairly study the detection of super-spreaders (and not any other infected individual), only the subset composed by (*a posteriori* determined and non-observed) super-spreaders and susceptible individuals at time T was considered (see Figure 5.6a for a schematic representation of the setting). Super-spreaders recovering before time T are not taken into account (although their number is very small after $T \sim 20$ days).

Figures 5.6b-5.6d display the distributions of the AUC obtained from the same contact tracing methods used above, over many epidemic realizations, for StEM, Covasim, and OpenABM. In all the three models, probabilistic methods (SMF and BP) turn out to better differentiate between non-infected and super-spreaders individuals as suggested by both the empirical distribution of the AUC (the mass of the distribution is significantly shifted to large values for SMF and BP) and the average value of the AUC as shown in the legend. The distributions associated with ICT, DCT (and TTQ for Covasim) predictions are conversely concentrated at low values, confirming that super-spreaders exposures cannot be effectively traced by non-probabilistic algorithms.

5.6 Backward, one-step and multi-step propagations

One of the intrinsic difficulties in performing contact tracing lies in identifying the direction of the infection among confirmed cases. In particular, identifying the so-called backward propagations, e.g. the source of infections, may help in identifying the super-spreaders and, generally, in mitigating the epidemic spreading [112, 122] at difference with the detection of forward propagations, e.g. the new infections, that can be considered an easier task. To further highlight the benefits carried by probabilistic contact-tracing methods, such as SMF and BP, it is interesting to estimate the ability of the different contact-tracing methods to identify secondary and tertiary infections, i.e. new (multi-step) infections at distances two and three from the observed individuals in the propagation history.

The same experimental setting adopted in Figure 5.6 is used: for each epidemic realization, the propagation is left free to evolve without intervention up to a time T and a small fraction of the symptomatic individuals is observed on a daily basis. The backward propagators are defined to be the sources of infections of the observed symptomatic individuals (blue dots in Fig. 5.7a.1), whereas forward propagators are defined to be their secondary infections (green nodes in Fig. 5.7b.1). The new infections at two and three steps from the observed nodes are displayed as orange nodes in the example in Fig. 5.7c.1.

To quantify the performances of the contact tracing methods, a comparison is performed between the areas under the ROC curves associated with the classification of the infected individuals in a restricted array, where the false positive set is composed of all the non-infected

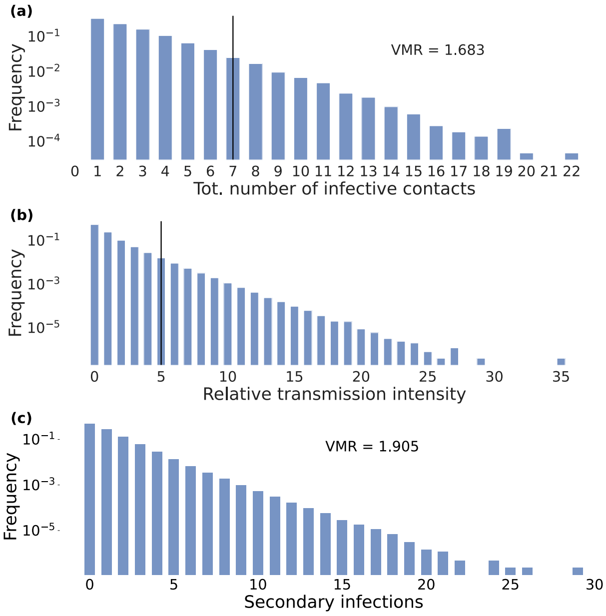

Figure 5.5: **Secondary infection capability distributions**: Histograms of the secondary infection distribution for (a) the Spatio-temporal model, and (c) OpenABM. The threshold used to identify the super-spreaders is 7 in both models. The variance-to-mean (VMR) ratio is larger than one in the two cases (the precise values are reported in panels (a) and (c)). In panel (b) the empirical distribution of the relative transmission intensity T_{rel} is shown for the Covasim model. For this model, a super-spreaders in defined as an infector that has at least $T_{\text{rel}} = 5$.

individuals (light grey nodes in Fig. 5.7a.1–5.7c.1) while the true positive set contains the backward infectors, the forward infections or infections at steps two and three further along the contagion chain. Other infected individuals that do not belong to one of these three categories are not considered (e.g. tested-positive individuals, i.e. red nodes in Fig. 5.7a.1–5.7c.1). Although with different performances on the three epidemic models, results in Figure 5.7 show that probabilistic models such as BP and SMF are very effective in identifying forward and backward propagations. For StEM (panels a.2–c.2) and OpenABM (panels a.4–c.4), the probabilistic contact-tracing methods outperform the other ones, especially when seeking for backward and multi-step propagations. In Covasim (panels a.3–c.3), probabilistic methods seem to play a crucial role only in detecting multi-step propagations, because doing that using simpler contact-tracing methods can be computationally excessively demanding.

5.7 Discussion

While social distancing, widespread use of masks and other hygiene measures are of the utmost importance to contain the spreading of the SARS-CoV2 virus, contact tracing represents the only non-pharmaceutical mitigation policy that can prevent the deployment of socio-economically

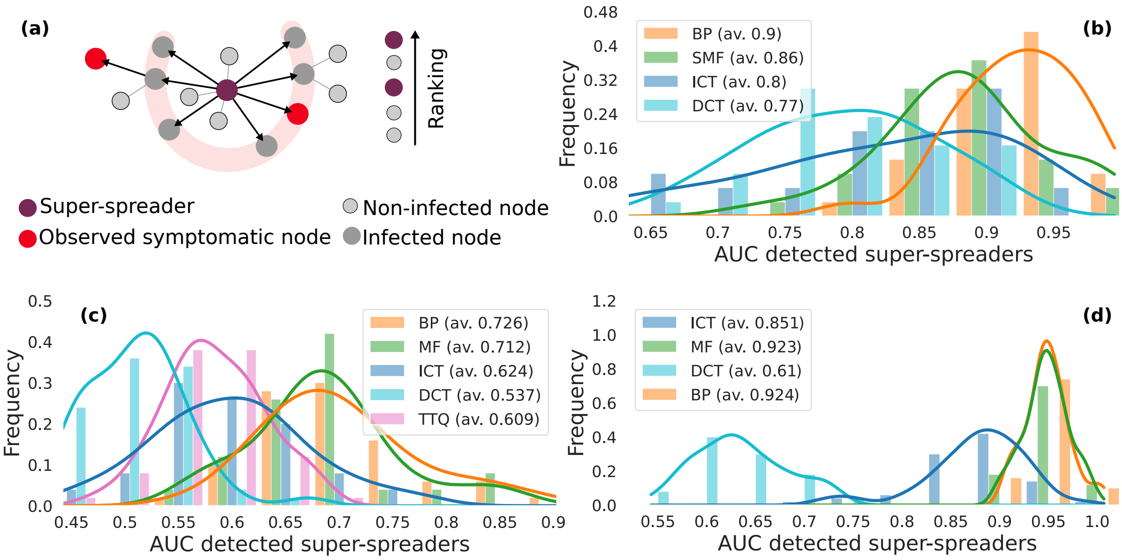

Figure 5.6: **Detection of the super-spreaders.** (a) Schematic representation of the ranking setup. A posteriori, the super-spreaders individuals (the purple nodes) are identified as those responsible for over-dispersed transmissions (see the main text for a proper definition for the three models), here marked as dark grey nodes within the pink shadow. ROC curves are built only for a subset of the individuals composed of the true super-spreaders and susceptible individuals (light grey nodes). Information about the epidemic dynamics entirely comes from the contact network and the daily observation of a fixed fraction of symptomatic (red nodes). The statistics of the AUC associated with the ROC curves obtained by different methods are shown for (b) StEM, (c) Covasim and (d) OpenABM. Lines are kernel density estimation plots used as guides for the eyes, while mean AUC values are reported in the legend. The duration of the free epidemic propagation before estimation is set to $T = 10$ for StEM, $T = 20$ for Covasim and OpenABM. The number of initially infected individuals is set to $N_{\text{pz}} = 200$ for StEM, $N_{\text{pz}} = 66$ for Covasim and $N_{\text{pz}} = 100$ for OpenABM. The fraction of observed symptomatic individuals is set to $p_{\text{sym}} = 0.1$ for StEM and for Covasim, while for OpenABM all severe symptomatic individuals are observed ($p_{\text{ssym}} = 1.0$) together with a fraction $p_{\text{msym}} = 0.3$ of mild ones.

impactful measures such as lockdowns. In particular, digital contact tracing overcomes the limitation of manual contact tracing by encompassing the ability to detect pre-symptomatic and asymptomatic individuals outside of close and known relationships with tested individuals, a key aspect in the prevention of COVID-19. The major drawback of current implementations of digital contact tracing is that, for large outbreaks, the number of exposure notifications delivered drastically grows, thus the number of candidate individuals to be tested and possibly confined, rendering the overall procedure impractical. A possible solution to this problem lies in improving the risk assessment of the population, through the reconstruction of the contagion channels and therefore through a more accurate estimate of the individual probability of infection. Test procedures guided by probabilistic contact tracing have recently proven to be more effective than the current standard implementations of digital contact tracing.

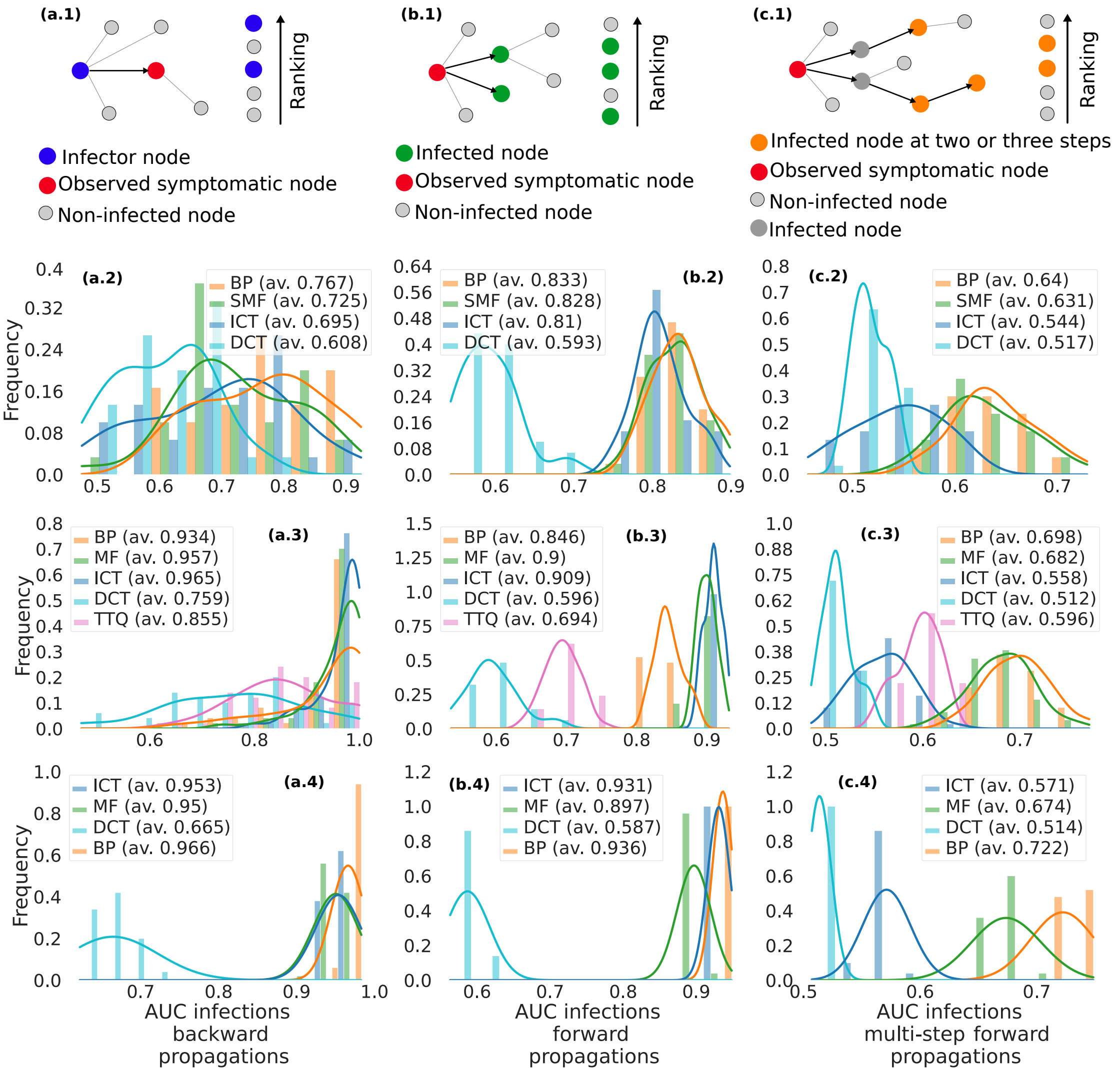

Figure 5.7: **Detection of the backward, one-step, and multi-step forward propagators.** Panels (a.1), (b.1), (c.1), and (d.1) show a schematic representation of the backward, one-step, and multi-step forward propagations respectively. In each case, only symptomatic individuals are observed (red dots), and the coloured nodes (backward, and one-step and multi-step forward) are put as positive for detection of infection, and the non-infected ones are included as negatives. The rest of the setup is the same as Figure 5.6, regarding the symptomatic observation probabilities, the number of initially infected individuals N_{pz} and the duration of free epidemic propagation T. The statistics of the AUC associated with the ROC curves obtained by different methods are shown for StEM in the second row, for Covasim in the third row and for OpenABM in the bottom row. Lines are kernel density estimation plots used as guides for the eyes, while mean AUC values are reported in the legend.

In this chapter, a quantitative analysis of the extent to which such procedures outperform other methods and the reasons behind this has been conducted, using as testing ground three different epidemic models developed for COVID-19. While the results obtained are not final, they strongly indicate that probabilistic contact-tracing methods can successfully mitigate ongoing outbreaks. This is evidenced by both the rapid decrease of the effective reproduction number below the critical value of one and a considerably smaller cumulative number of infected compared to other methods. This higher containing efficacy also comes with a similar or significantly lower deployment of testing resources. In particular, there is evidence that employing probabilistic contact tracing with threshold testing can reduce significantly the level of testing required for the containment of an ongoing epidemic outbreak.

The numerical experiments also revealed that probabilistic methods are more suitable than others for the detection of super-spreading events, which are pointed to as one of the important mechanisms for the spreading of SARS-CoV-2 virus. Epidemics characterized by overdispersion, both by innate variety of transmissibility and by heterogeneity of the contact network, are expected to become more common in the future, necessitating the development of tracing methods that can exploit such peculiarities. Because of super-spreaders, one would also work backwards to identify who first infected the observed case, as it is likely that most people have been infected by someone who also infected other individuals. In this respect, probabilistic contact tracing methods were found to outperform other methods in correctly reconstructing infection channels by backward tracing. Although their performances has been evaluated here in regimes where the trace-test and isolate technique can successfully contain most outbreaks, these methods are sufficiently flexible to be applied in more general situations and cope with other measures for COVID-19 prevention and mitigation, especially with the presence of vaccinated individuals.

The probabilistic contact tracing methods used in the chapter rely on the collection of contacts via a smartphone application or other devices: while results show that epidemic containment could be possible with just 80% of the population installing the app, the level of adoption during the COVID-19 pandemic has been much lower in western countries [163]. However, this is a subject of active research, with some work pointing to ways to increase the app adoption, like better public campaigns for promotion and the inclusion of other useful features that could transform the app into a gateway for public health informations and services [121, 163, 149].

Chapter 6

Dynamic cavity methods in epidemics

The Belief Propagation approach to individual-based epidemic processes can be classified as a generalized mean-field method, because the epidemic trajectories of the neighbours of a given individual are assumed to be conditionally independent. This hypothesis is correct only when the dynamical process takes place on a contact network with a tree-like topology, but the method has proved to be very effective also in epidemic inference problems defined on loopy contact networks. This is the same assumption of the dynamic cavity approach [24, 27], which has been applied to both recurrent and non-recurrent epidemic processes [46, 44, 166]. When dealing with pure time-forward epidemic dynamics without observations, in the case of non recurrent individual states, the BP/dynamic cavity approach simplifies into a dynamic message passing technique that has been extensively used to study spreading processes on networks [44, 35]. Even the simpler individual-based mean-field method, also known as the N-intertwined model, has been shown to provide moderately good approximations to time-forward epidemic dynamics [20]. Recently, the individual-based mean-field method has been employed to propose a very simple inference method where the observations of individual states are heuristically taken into account [118]. Despite the absence of a correct Bayesian framework and the simplicity of the approach, this method provides moderately good results for epidemic risk assessment.

In this chapter, a novel generalized mean field approximation for Bayesian epidemic inference and risk assessment is proposed. For simplicity, the method is developed for the Susceptible-Infected (SI) model but its generalization to epidemic models including additional states, such as Recovered and Exposed states, is possible.

6.1 Methodology

In this section, the individual-based SI model on a contact network is introduced, and then rewritten in a way compatible for the dynamic cavity approach. Then, the dynamic cavity equations are derived, and the small coupling expansion is obtained. From this expansion, an approximation leads to the mean field formulation of the dynamic cavity method.

6.1.1 Definition of the stochastic epidemic model and observations

The simplest non-trivial model employed in epidemic inference is the discrete-time stochastic SI model. Let us consider then the dynamics of the SI model on a population of N individuals over a temporal window of T time steps (e.g. days). The daily contacts are supposed to be directly encoded in the set of parameters specifying the infection transmission, with λ_{ij}^t being the infection probability along the directed edge from individual i to individual j at time t. The epidemic state of the population at time t is represented by a binary array $\mathbf{x}^t = (x_1^t, \dots, x_N^t)$, with $x_i^t = 0$ (resp. $x_i^t = 1$) meaning that i is a Susceptible (resp. Infected) individual at time t. The epidemic model is assumed to be Markovian, although this hypothesis can be relaxed. In the Markovian setup, the time evolution of the probability $p_t\left[\mathbf{x}^t\right]$ that the population is in state $\mathbf{x}^t$ at time t is given in terms of the master equation

$$p_{t+1}\left[\mathbf{x}^{t+1}\right] = \sum_{\mathbf{x}^t} W\left[\mathbf{x}^{t+1}|\mathbf{x}^t\right] p_t\left[\mathbf{x}^t\right] \tag{6.1}$$

with transition rates $W\left[\mathbf{x}^{t+1}|\mathbf{x}^t\right] = \prod_i W_i\left(x_i^{t+1}|\mathbf{x}^t\right)$ where

$$W_i\left(x_i^{t+1} = 1|\mathbf{x}^t\right) = x_i^t + \left(1 - x_i^t\right)\left[1 - \prod_j \left(1 - \lambda_{ji}^t x_j^t\right)\right] \tag{6.2a}$$

$$W_i\left(x_i^{t+1} = 0|\mathbf{x}^t\right) = \left(1 - x_i^t\right)\prod_j \left(1 - \lambda_{ji}^t x_j^t\right). \tag{6.2b}$$

It is convenient to introduce local fields $h_i^t = \sum_j \nu_{ji}^t x_j^t$, with $\nu_{ji}^t = \log\left(1 - \lambda_{ji}^t\right)$, such that

$$\prod_j \left(1 - \lambda_{ji}^t x_j^t\right) = \prod_j \left(1 - \lambda_{ji}^t\right)^{x_j^t} = e^{h_i^t}, \tag{6.3}$$

and use them to provide an equivalent description to the master equation, based on a system of discrete-time stochastic maps

$$x_i^{t+1} = x_i^t + \left(1 - x_i^t\right) r_i^t \tag{6.4}$$

in which r_i^t is a Bernoulli random variable with parameter $1 - e^{h_i^t}$, i.e.

$$P\left[r_i^t|h_i^t\right] = \text{Bernoulli}\left(1 - e^{h_i^t}\right). \tag{6.5}$$

The likelihood of the model can be defined by a set $\mathcal{O}$ of statistically independent observations, each of them providing information about the state of a certain node i at the corresponding observation time. The most general scenario admits multiple observations on the same node i (at different times), encoded in the vectors O_i, and uncertainty on the outcome of the tests, the latter being eventually quantified by false positive rate ν_p and/or false negative rates ν_n. If node i is observed at time τ_{μ_i}, the corresponding likelihood over its epidemic trajectory $x_i = (x_i^0, \dots, x_i^T)$ reads

$$p\left(O_i^{\tau_{\mu_i}} \mid x_i\right) = \begin{cases} \left(1 - \nu_p\right)\delta_{\mathrm{K}}\left(x_i^{\tau_{\mu_i}} - 0\right) + \nu_n \delta_{\mathrm{K}}\left(x_i^{\tau_{\mu_i}} - 1\right) & \text{if } O_i^{\tau_{\mu_i}} = 0 \\ \\ \nu_p\, \delta_{\mathrm{K}}\left(x_i^{\tau_{\mu_i}} - 0\right) + \left(1 - \nu_n\right)\delta_{\mathrm{K}}\left(x_i^{\tau_{\mu_i}} - 1\right) & \text{if } O_i^{\tau_{\mu_i}} = 1 \end{cases} \tag{6.6}$$

where $\delta_{\mathrm{K}}(\cdot)$ is the Kronecker delta function. The total likelihood over the full set of observations $\mathcal{O} = \{O_i\}_{i=1}^{N}$ can be rewritten as

$$p(\mathcal{O} \mid X) = \prod_i p(O_i \mid x_i) = \prod_i \prod_{\mu_i} p\left(O_i^{\tau_{\mu_i}} \mid x_i\right) \tag{6.7}$$

where each term in the last equation has the form (6.6). At the rightmost hand side, the second product runs over all the observations on node i. In the above equation the quantity X is a short-hand notation to indicate the trajectories of all nodes, namely $X = \{x_1, x_2, \dots, x_N\} = \{\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^T\}$. In the case of perfectly accurate tests, in which $\nu_p = \nu_n = 0$, the effect of the observations is to enforce the dynamical trajectories to be compatible with the observed states. The posterior probability of the trajectory $\mathbf{X}$ can be expressed using Bayes' theorem as follows

$$p(\mathbf{X}|\mathcal{O}) = \frac{1}{p(\mathcal{O})} p(\mathbf{X})\, p(\mathcal{O} \mid X) \tag{6.8a}$$

$$= \frac{1}{p(\mathcal{O})} \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t, h_i^t} P\left[r_i^t | h_i^t\right] \delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t - \left(1 - x_i^t\right) r_i^t\right) \delta\left(h_i^t - \sum_j v_{ji}^t x_j^t \right) \right] \right\} p(O_i \mid x_i) \tag{6.8b}$$

$$\propto \prod_i \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t, h_i^t} P\left[r_i^t | h_i^t\right] \delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t - \left(1 - x_i^t\right) r_i^t\right) \delta\left(h_i^t - \sum_j v_{ji}^t x_j^t \right) p\left(O_i^t \mid x_i^t\right) \right] \right\} \tag{6.8c}$$

where in the last expression it is assumed the simplifying notation that the conditional probability $p\left(O_i^t \mid x_i^t\right) = 1$ also in the case in which there is no observation of the state of individual i at time t, i.e. $\nexists\, \mu_i$ such that $t = \tau_{\mu_i}$. The same will be assumed in the rest of the paper.

The Bayesian inference problem consists in evaluating marginals of the posterior distribution $p(\mathbf{X}|\mathcal{O})$, such as the quantity $p\left(x_i^t = x|\mathcal{O}\right)$ representing the posterior probability that individual i is in state $x \in \{0,1\}$ at time t given the set of the available observations $\mathcal{O}$. The posterior distribution is, in general, intractable but it is the starting point for the derivation of approximate inference methods.

6.1.2 The Dynamic Cavity Equations for the SI model with observations

Starting from the posterior probability weight in Eq. (6.8c) and employing a cavity argument by removing the node j (and the corresponding trajectory $x_j = (x_j^0, \dots, x_j^T)$), the marginal probability $c_{ij}\left[x_i, s_i|\mathcal{O}\right]$, representing the probability of the pair of variable-field trajectories (x_i, s_i) on node i in the cavity graph, can be computed by means of the following *dynamic cavity equations*:

$$\begin{aligned} c_{ij}\left[x_i, s_i|\mathcal{O}\right] = \frac{1}{Z_{ij}\left[\mathcal{O}\right]} p(x_i^0) \sum_{x_{\partial i \setminus j}} \Bigg\{ &\left[\prod_{k \in \partial i \setminus j} c_{ki}\left[x_k, v_{ik} x_i|\mathcal{O}\right] \right] \\ &\times \prod_{t=0}^{T-1} \left[\delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t\right) e^{s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t} + \delta_{\mathrm{K}}\left(x_i^{t+1} - 1\right) \left(1 - e^{s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t}\right) \right] p\left(O_i^t | x_i^t\right) \Bigg\}, \end{aligned} \tag{6.9}$$

where $x_{\partial i \setminus j} = \{x_k\}_{k \in \partial i \setminus j}$ is the set of trajectories on neighbors of i except for j, the notation $v_{ik} x_i$ stands for the array $(v_{ik}^1 x_i^1, \dots, v_{ik}^T x_i^T)$, and $Z_{ij}\left[\mathcal{O}\right]$ is a normalization term for the cavity marginals. The equations (6.9), for which a derivation from Eq. (6.8c) is provided in Appendix C, are exact when the underlying interaction graph (the time-independent projection of the contact network) is a tree and provide a rather good approximation of the posterior distribution on sparse loopy graphs. In agreement with the formulation of Belief Propagation equations on epidemic trajectories, it is convenient to interpret s_i as a proxy for the trajectory of the missing neighbouring node j in the cavity graph (more precisely $s_i \propto v_{ji} x_j$), and then normalize the marginal over both arguments, by defining

$$\begin{aligned} Z_{ij}\left[\mathcal{O}\right] = \sum_{x_i, s_i} p(x_i^0) \sum_{x_{\partial i \setminus j}} \Bigg\{ \Bigg[\prod_{k \in \partial i \setminus j} c_{ki}\left[x_k, v_{ik} x_i | \mathcal{O}\right] \Bigg] \\ \times \prod_{t=0}^{T-1} \left[\delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t\right) e^{s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t} + \delta_{\mathrm{K}}\left(x_i^{t+1} - 1\right) \left(1 - e^{s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t}\right) \right] p\left(O_i^t | x_i^t\right) \Bigg\}. \end{aligned} \tag{6.10}$$

Completing the cavity and computing the total marginal over i gives the posterior marginal probability of one-site trajectories

$$\begin{aligned} p_i\left(x_i | \mathcal{O}\right) \propto p(x_i^0) \sum_{x_{\partial i}} \Bigg(\prod_{k \in \partial i} c_{ki}\left[x_k, x_i v_{ik} | \mathcal{O}\right] \Bigg) \\ \times \prod_{t=0}^{T-1} \left\{ \left[\delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t\right) e^{\sum_{k \in \partial i} x_k^t v_{ki}^t} + \delta_{\mathrm{K}}\left(x_i^{t+1} - 1\right) \left(1 - e^{\sum_{k \in \partial i} x_k^t v_{ki}^t}\right) \right] p\left(O_i^t | x_i^t\right) \right\} \end{aligned} \tag{6.11}$$

6.1.3 Small-coupling expansion

It is convenient to express the cavity marginal in Eq. (6.9) in terms of the conjugate field trajectory $\hat{h}_i = (\hat{h}_i^1, \dots, \hat{h}_i^T)$, by introducing its Fourier transform (see Appendix C), which leads to the following expression for the dynamic cavity equations

$$\begin{aligned} c_{ij}\left[x_i, \hat{h}_i | \mathcal{O}\right] = \frac{p\left(x_i^0\right)}{Z_{ij}[\mathcal{O}]} \prod_{t=0}^{T-1} \left\{ \left[\delta\left(\hat{h}_i^t - \mathrm{i}\right) \left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right) \delta\left(x_i^{t+1} - 1\right) \right] p\left(O_i^t | x_i^t\right) \right\} \\ \times \prod_{k \in \partial i \setminus j} \left[\sum_{x_k} \int D\hat{h}_k c_{ki}\left[x_k, \hat{h}_k | \mathcal{O}\right] e^{-\mathrm{i} \sum_t \left(x_k^t v_{ki}^t \hat{h}_i^t + x_i^t v_{ik}^t \hat{h}_k^t\right)} \right] \end{aligned} \tag{6.12}$$

In the spirit of Plefka's approach [6, 48], one can add a perturbative parameter α in the argument of the exponential regarding the interaction term, and perform their Taylor expansion up to a certain order in α. To the second order in α, we get

$$\begin{aligned} c_{ij}\left[x_i,\hat{h}_i|\mathcal{O}\right] &= \frac{p\left(x_i^0\right)}{Z_{ij}[\mathcal{O}]}\prod_{t=0}^{T-1}\left\{\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta\left(x_i^{t+1}-x_i^t\right)-\delta\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1}-1\right)\right]p\left(O_i^t|x_i^t\right)\right\} \\ &\quad\times\prod_{k\in\partial i\setminus j}\left[\sum_{x_k}\int D\hat{h}_k c_{ki}\left[x_k,\hat{h}_k|\mathcal{O}\right]e^{-\mathrm{i}\alpha\sum_t\left(x_k^t v_{ki}^t\hat{h}_i^t+x_i^t v_{ik}^t\hat{h}_k^t\right)}\right] && (6.13\text{a}) \\ &\approx \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}[\mathcal{O}]}\prod_{t=0}^{T-1}\left\{\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta\left(x_i^{t+1}-x_i^t\right)-\delta\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1}-1\right)\right]p\left(O_i^t|x_i^t\right)\right\} \\ &\quad\times\prod_{k\in\partial i\setminus j}\sum_{x_k}\int D\hat{h}_k c_{ki}\left[x_k,\hat{h}_k|\mathcal{O}\right]\left\{1-\mathrm{i}\alpha\sum_t\left(x_k^t v_{ki}^t\hat{h}_i^t+x_i^t v_{ik}^t\hat{h}_k^t\right)\right. \\ &\qquad\left.+\frac{\alpha^2}{2}\sum_{t,t'}\left(x_k^t v_{ki}^t\left(-\mathrm{i}\hat{h}_i^t\right)+x_i^t v_{ik}^t\left(-\mathrm{i}\hat{h}_k^t\right)\right)\left(x_k^{t'}v_{ki}^{t'}\left(-\mathrm{i}\hat{h}_i^{t'}\right)+x_i^{t'}v_{ik}^{t'}\left(-\mathrm{i}\hat{h}_k^{t'}\right)\right)\right\}, && (6.13\text{b}) \end{aligned}$$

where $Z_{ij}[\mathcal{O}]$ is also consistently approximated at the second order by $\tilde{Z}_{ij}[\mathcal{O}]$. By performing the averages over the dynamic cavity marginals c_{ki}, one obtains

$$\begin{aligned} c_{ij}\left[x_i,\hat{h}_i|\mathcal{O}\right] &= \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}[\mathcal{O}]}\prod_{t=0}^{T-1}\left\{\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta\left(x_i^{t+1}-x_i^t\right)-\delta\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1}-1\right)\right]p\left(O_i^t|x_i^t\right)\right\} \\ &\quad\times\prod_{k\in\partial i\setminus j}\left\{1+\alpha\sum_t\left(m_{k\setminus i}^t v_{ki}^t\left(-\mathrm{i}\hat{h}_i^t\right)+x_i^t v_{ik}^t\mu_{k\setminus i}^t\right)+\frac{\alpha^2}{2}\sum_{t,t'}V_{k\setminus i}^{(2)}\left(t,t'\right)\right\} && (6.14\text{a}) \\ &\approx \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}[\mathcal{O}]}\prod_{t=0}^{T-1}\left\{\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta\left(x_i^{t+1}-x_i^t\right)-\delta\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1}-1\right)\right]p\left(O_i^t|x_i^t\right)\right\} \\ &\quad\times\prod_t\exp\left\{\alpha\sum_{k\in\partial i\setminus j}\left(m_{k\setminus i}^t v_{ki}^t\left(-\mathrm{i}\hat{h}_i^t\right)+x_i^t v_{ik}^t\mu_{k\setminus i}^t\right)+\frac{\alpha^2}{2}\sum_{k\in\partial i\setminus j}\sum_{t'}V_{k\setminus i}^{(2)}\left(t,t'\right)\right\}, && (6.14\text{b}) \end{aligned}$$

with the second order term $V_{k\setminus i}^{(2)}$ being

$$V_{k\setminus i}^{(2)}\left(t,t'\right)=-v_{ki}^t\hat{h}_i^t v_{ki}^{t'}\hat{h}_i^{t'}C_{k\setminus i}^{tt'}+v_{ki}^t\left(-\mathrm{i}\hat{h}_i^t\right)R_{k\setminus i}^{tt'}x_i^{t'}v_{ik}^{t'}+v_{ki}^{t'}\left(-\mathrm{i}\hat{h}_i^{t'}\right)R_{k\setminus i}^{t't}x_i^t v_{ik}^t+B_{k\setminus i}^{tt'}x_i^t v_{ik}^t x_i^{t'}v_{ik}^{t'} \quad (6.15)$$

In the above equations, a set of one-time and two-time cavity quantities were defined by the relations

$$\begin{aligned} m_{k\setminus i}^t = m_{k\setminus i}^t[\mathcal{O}] &= \sum_{x_k}\int D\hat{h}_k c_{ki}\left[x_k,\hat{h}_k|\mathcal{O}\right]x_k^t \\ &= \sum_{x_k}c_{ki}\left[x_k,s_k=0|\mathcal{O}\right]x_k^t && (6.16\text{a}) \\ \mu_{k\setminus i}^t = \mu_{k\setminus i}^t[\mathcal{O}] &= \sum_{x_k}\int D\hat{h}_k c_{ki}\left[x_k,\hat{h}_k|\mathcal{O}\right]\left(-\mathrm{i}\hat{h}_k^t\right) \\ &= \sum_{x_k}\frac{\delta}{\delta s_k^t}c_{ki}\left[x_k,s_k|\mathcal{O}\right]\Bigg|_{s_k=0}. && (6.16\text{b}) \end{aligned}$$

and

$$\begin{aligned}
C_{k\setminus i}^{tt'} = C_{k\setminus i}^{t,t'}[\mathcal{O}] &= \sum_{x_k} \int D\hat{h}_k c_{ki} \left[x_k, \hat{h}_k | \mathcal{O}\right] x_k^t x_k^{t'} \\
&= \sum_{x_k} c_{ki} \left[x_k, s_k = 0 | \mathcal{O}\right] x_k^t x_k^{t'} \qquad (6.17a) \\
R_{k\setminus i}^{tt'} = R_{k\setminus i}^{tt'}[\mathcal{O}] &= \sum_{x_k} \int D\hat{h}_k c_{ki} \left[x_k, \hat{h}_k | \mathcal{O}\right] x_k^t \left(-\mathrm{i}\hat{h}_k^{t'}\right) \\
&= \frac{\delta}{\delta s_k^{t'}} \sum_{x_k} x_k^t c_{ki} \left[x_k, s_k | \mathcal{O}\right] \Big|_{s_k=0} \qquad (6.17b) \\
B_{k\setminus i}^{tt'} = B_{k\setminus i}^{tt'}[\mathcal{O}] &= \sum_{x_k} \int D\hat{h}_k c_{ki} \left[x_k, \hat{h}_k | \mathcal{O}\right] \left(-\mathrm{i}\hat{h}_k^t\right) \left(-\mathrm{i}\hat{h}_k^{t'}\right) \\
&= \frac{\delta^2}{\delta s_k^t \partial s_k^{t'}} \sum_{x_k} c_{ki} \left[x_k, s_k | \mathcal{O}\right] \Big|_{s_k=0} . \qquad (6.17c)
\end{aligned}$$

The quantity $m_{k\setminus i}^t$ measures the average probability that node k is infected at time t in the absence of the interaction with i. Similarly, $C_{k\setminus i}^{tt'}$ represents the two-time autocorrelation function of node k in the cavity graph, the quantity $R_{k\setminus i}^{tt'}$ is the response function on node k at time t' to a perturbation due to an infinitesimal external field acting on node k at time t'. The remaining two quantities $\mu_{k\setminus i}^t$ and $B_{k\setminus i}^{tt'}$ are of less intuitive interpretation, as they measure the mean and temporal correlations of fluctuations around the unperturbed single-site statistics in the cavity graph. A direct calculation of the quantity $\sum_{x_k} \int D\hat{h}_k(-\mathrm{i}\hat{h}_k^t) c_{ki}[x_k, \hat{h}_k | \mathcal{O} = \emptyset]$ shows that, in the absence of observations, $\mu_{k\setminus i}^t = 0$. Similarly, $B_{k\setminus i}^{tt'} = 0$ in the absence of observations. This result, due to the causality of the dynamical process, does not hold any more when some observations are included. For notational convenience, the implicit dependence of all marginals and normalization constants on the set of observations $\mathcal{O}$ will not be further reported in the following.

6.1.4 Mean-field approximation

A mean-field approximation can be obtained neglecting the second-order terms in Eq. (6.14b) and focusing on the effects of the first-order ones. The expression of the dynamic cavity equations simplifies as follows (setting $\alpha = 1$):

$$\begin{aligned}
c_{ij}\left[x_i, s_i\right] = \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}} \prod_{t=0}^{T-1} \Bigg[\int d\hat{h}_i^t e^{-\mathrm{i}\hat{h}_i^t \left(s_i^t + \sum_{k\in\partial i\setminus j} m_{k\setminus i}^t v_{ki}^t\right)} e^{\sum_{k\in\partial i\setminus j} x_i^t v_{ik}^t \mu_{k\setminus i}^t} \\
\times \Big\{ \delta\left(\hat{h}_i^t - \mathrm{i}\right) \left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right) \delta\left(x_i^{t+1} - 1\right) \Big\} p\left(O_i^t | x_i^t\right) \Bigg] \qquad (6.18a) \\
\propto \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}} \prod_{t=0}^{T-1} \Bigg[\Big\{ \delta\left(x_i^{t+1} - x_i^t\right) e^{s_i^t + \sum_{k\in\partial i\setminus j} m_{k\setminus i}^t v_{ki}^t} + \delta\left(x_i^{t+1} - 1\right) \Big[1 - e^{s_i^t + \sum_{k\in\partial i\setminus j} m_{k\setminus i}^t v_{ki}^t}\Big] \Big] \Big\} \\
\times e^{\sum_{k\in\partial i\setminus j} x_i^t v_{ik}^t \mu_{k\setminus i}^t} p\left(O_i^t | x_i^t\right) \Bigg] \qquad (6.18b)
\end{aligned}$$

where, using the definitions in Eqs. (6.16a) and (6.16b), the two quantities $m_{i\setminus j}^t$ and $\mu_{i\setminus j}^t$ turn out to satisfy the self-consistent equations

$$
\begin{aligned}
m_{i\setminus j}^t = \sum_{x_i} x_i^t \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}} \prod_{t'=0}^{T-1} \Bigg[\Bigg\{ \delta\left(x_i^{t'+1} - x_i^{t'}\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t'} \nu_{ki}^{t'}} + \delta\left(x_i^{t'+1} - 1\right) \left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t'} \nu_{ki}^{t'}}\right] \Bigg\} \\
\times e^{\sum_{k\in\partial i\setminus j} x_i^{t'} \nu_{ik}^{t'} \mu_{k\setminus i}^{t'}} p\left(O_i^t \mid x_i^t\right) \Bigg]
\end{aligned}
\tag{6.19}
$$

and

$$
\begin{aligned}
\mu_{i\setminus j}^t = \sum_{x_i} \frac{p\left(x_i^0\right)}{\tilde{Z}_{ij}} \left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) e^{\sum_{k\in\partial i\setminus j}\left(m_{k\setminus i}^t \nu_{ki}^t + x_i^t \nu_{ik}^t \mu_{k\setminus i}^t\right)} \\
\times \prod_{t'\neq t} \Bigg\{ \delta\left(x_i^{t'+1} - x_i^{t'}\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t'} \nu_{ki}^{t'}} + \delta\left(x_i^{t'+1} - 1\right) \left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t'} \nu_{ki}^{t'}}\right] \Bigg\} \\
\times e^{\sum_{k\in\partial i\setminus j} x_i^{t'} \nu_{ik}^{t'} \mu_{k\setminus i}^{t'}} p\left(O_i^t \mid x_i^t\right) \Bigg] .
\end{aligned}
\tag{6.20}
$$

The normalization constant is chosen to ensure that the time-dependent cavity mean $m_{i\setminus j}^t$ is normalized in the cavity graph, that is

$$
\begin{aligned}
\tilde{Z}_{ij} = \sum_{x_i} p\left(x_i^0\right) \prod_{t=0}^{T-1} \Bigg[\Bigg\{ \delta\left(x_i^{t+1} - x_i^t\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^t \nu_{ki}^t} + \delta\left(x_i^{t+1} - 1\right) \left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^t \nu_{ki}^t}\right] \Bigg\} \\
\times e^{\sum_{k\in\partial i\setminus j} x_i^t \nu_{ik}^t \mu_{k\setminus i}^t} p\left(O_i^t \mid x_i^t\right) \Bigg] .
\end{aligned}
\tag{6.21}
$$

In addition, the total time-dependent marginal m_i^t of the posterior distribution on the full graph is given by

$$
m_i^t = \frac{1}{\tilde{Z}_i} \sum_{x_i} p\left(x_i^0\right) x_i^t \prod_{t=0}^{T-1} \left[\left\{ \delta\left(x_i^{t+1} - x_i^t\right) e^{\sum_{k\in\partial i} m_{k\setminus i}^t \nu_{ki}^t} + \delta\left(x_i^{t+1} - 1\right) \left[1 - e^{\sum_{k\in\partial i} m_{k\setminus i}^t \nu_{ki}^t}\right] \right\} e^{\sum_{k\in\partial i} x_i^t \nu_{ik}^t \mu_{k\setminus i}^t} p\left(O_i^t \mid x_i^t\right) \right]
\tag{6.22}
$$

with

$$
\tilde{Z}_i = \sum_{x_i} p(x_i^0) \prod_{t=0}^{T-1} \left[\left\{ \delta\left(x_i^{t+1} - x_i^t\right) e^{\sum_{k\in\partial i} m_{k\setminus i}^t \nu_{ki}^t} + \delta\left(x_i^{t+1} - 1\right) \left[1 - e^{\sum_{k\in\partial i} m_{k\setminus i}^t \nu_{ki}^t}\right] \right\} e^{\sum_{k\in\partial i} x_i^t \nu_{ik}^t \mu_{k\setminus i}^t} p\left(O_i^t \mid x_i^t\right) \right] .
\tag{6.23}
$$

Equations (6.19) and (6.20) represent the set of self-consistent equations defining a non-causal (dynamic) MF approximation, that is valid for small values of the couplings. The dynamical equations are indeed of mean-field type, since correlations are neglected, but in the presence of observations they describe a non-causal dynamical process. Because of the cavity construction, the fundamental unknown of the equations, the one-time cavity marginals $m_{i\setminus j}^t$ and the one-time cavity fields $\mu_{i\setminus j}^t$, are defined by means of local self-consistent conditions, which can be implemented using a message-passing update scheme. A computational bottleneck of Eqs. (6.19)-(6.20) is represented by the partial trace over single-site trajectories x_i, that requires $O(2^T)$ operations, meaning that a complete update of all cavity quantities requires $O\left(|E| T 2^T\right)$, where $|E|$ is the total number of non-zero weighted directed edges on the interaction graph. This computational cost can be reduced by employing different representations of the epidemic state, as is done in the following section.

6.2 Infection times representation

In order to reduce the amount of computations required for the message-passing update, it's useful to pass to the infection time representation, as done in previous chapters. States x_i^t are defined for $t \in \{0,1,\dots,T\}$, so it has $T+1$ values, and together form the time trajectory of the individual. This can be described by the time of infection $t_i \in \{0,1,\dots,T\} \cup \{\infty\}$ Then, the correspondence is simply

$$x_i^t = \mathbb{I}\left[t \geq t_i\right] \tag{6.24}$$

where $\mathbb{I}$ is the indicator function. The following cases are shown as example of trajectories with the corresponding value of t_i:

$$\begin{aligned}
(0,0,\dots,0,0) &\to t_i = \infty \\
(0,0,\dots,0,1) &\to t_i = T \\
(0,0,\dots,1,1) &\to t_i = T-1 \\
(0,1,\dots,1,1) &\to t_i = 1 \\
(1,1,\dots,1,1) &\to t_i = 0
\end{aligned}$$

With this change of representation, it's possible to rewrite the equations for the cavity messages. Equation (6.19) becomes:

$$m_{i\setminus j}^t = \frac{1}{\tilde{Z}_{ij}} \sum_{t_i=0}^{t} p\left(O_i \,|\, t_i\right) \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r \nu_{ki}^r}\right] \left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^r \mu_{k\setminus i}^r} \tag{6.25}$$

with the normalization $\tilde{Z}_{ij}$ (6.21) now reading:

$$\tilde{Z}_{ij} = \sum_{t_i=0}^{T+1} p\left(O_i \,|\, t_i\right) \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r \nu_{ki}^r}\right] \left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^r \mu_{k\setminus i}^r} \tag{6.26}$$

and with $p\left(O_i \,|\, t_i\right)$ the probability of getting the set of observations O_i when the individual has been infected at time t_i. The case of no infection $t_i = \infty$ can be modelled in practice with a finite infection time $T_\infty > T$ (and is considered actually $T_\infty = T+1$). The function ϕ is defined as the following:

$$\phi\left(t_i\right) = \begin{cases} 1 & 1 \leq t_i \leq T \\ 0 & \text{else} \end{cases}$$

For the cavity fields μ, the result is:

$$\begin{aligned}
\mu_{i\setminus j}^t = {} & \frac{1}{\tilde{Z}_{ij}} \sum_{t_i=t}^{T-1} p\left(O_i \,|\, t_i\right) \pi_i^0\left(t_i+2\right) \left[\prod_{r=0}^{t_i} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r \nu_{ki}^r}\right] \left[1 - \phi\left(t_i+2\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i+1} \nu_{ki}^{t_i+1}}\right] \prod_{r=t_i+2}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^r \mu_{k\setminus i}^r} + \\
& - \frac{\mathbb{I}\left[0 \leq t \leq T\right]}{\tilde{Z}_{ij}} p\left(O_i \,|\, t+1\right) \pi_i^0\left(t+1\right) \left[\prod_{r=0}^{t} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r \nu_{ki}^r}\right] \prod_{r=t+1}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^r \mu_{k\setminus i}^r}
\end{aligned} \tag{6.27}$$

These equations show the separation of the flow of information in epidemic time: for the computation of the $m_{i\setminus j}$, the messages of the neighbours $m_{k\setminus i}$ are used up until time $t_i - 2$, while for the times $r \geq t_i$, the information of the cavity fields $\mu_{k\setminus i}$ is used. The central factor in equations (6.25) and (6.26) represents the correction that needs to be included if the individual has been infected at $t_i > 0$ (i.e., when it's not a patient zero).

The marginals can also be written in the infection times representation, applying the same treatment as to the cavity messages:

$$m_i^t = \frac{1}{Z_i} \sum_{t_i=0}^{t} p\left(O_i \,|\, t_i\right) \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i} m_{k\setminus i}^r \nu_{ki}^r}\right] \left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i} \nu_{ik}^r \mu_{k\setminus i}^r} \tag{6.28}$$

with Z_i as

$$Z_i = \sum_{t_i=0}^{T+1} p\left(O_i \,|\, t_i\right) \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i} m_{k\setminus i}^r \nu_{ki}^r}\right] \left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i} \nu_{ik}^r \mu_{k\setminus i}^r} \tag{6.29}$$

The derivation of these equations is described in Appendix D, along with the details for the implementation of the method. Using the above equations for the message passing makes it possible to reduce the computational cost of one update to $O(|E|T)$. However, this kind of representation is only suited to irreversible epidemic processes, like the SI and SIR models.

6.3 Recovering time-forward dynamics

Causality-breaking is a consequence of the existence of observations at later times, that have to be taken into account in the mathematical model by a flux of information flowing backward in time and conditioning the whole history of the process. This property reflects in the existence of non-trivial values for the one-time cavity fields $\mu_{i\setminus j}^t$. One can recover usual causal time-forward mean-field dynamics when no observation is present and, consequently, all cavity fields $\mu_{i\setminus j}^t$ vanish. To show this, it is convenient to start from a particular form of the update equations for the cavity marginals $m_{i\setminus j}^t$ (see App.C.1 for a derivation),

$$m_{i\setminus j}^t = m_{i\setminus j}^{t-1} + \left(1 - m_{i\setminus j}^{t-1}\right) \frac{\left\{1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right\} \rho_{t\leftarrow}^{i\setminus j}(1)}{e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}} \rho_{t\leftarrow}^{i\setminus j}(0) + \left\{1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right\} \rho_{t\leftarrow}^{i\setminus j}(1)}, \tag{6.30}$$

where the messages $\rho_{t\leftarrow}^{i\setminus j}(x)$ represent the backward probability that state $x_i^t = x \in \{0,1\}$ given the dynamic constraints and the observations in the future. In the absence of observations (on all nodes at all times $t' \geq t$) the backward probability is balanced, i.e. $\rho_{t\leftarrow}^{i\setminus j}(0) = \rho_{t\leftarrow}^{i\setminus j}(1)$, and (6.30) reduces to the usual time-forward mean-field equations,

$$m_{i\setminus j}^t = m_{i\setminus j}^{t-1} + \left(1 - m_{i\setminus j}^{t-1}\right)\left\{1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right\} \tag{6.31a}$$

$$= m_{i\setminus j}^{t-1} + \left(1 - m_{i\setminus j}^{t-1}\right)\left\{1 - \prod_{k\in\partial i\setminus j} \left(1 - \lambda_{ki}^{t-1}\right)^{m_{k\setminus i}^{t-1}}\right\}. \tag{6.31b}$$

For the total marginals m_i^t, we obtain

$$\begin{aligned} m_i^t &= m_i^{t-1} + \left(1 - m_i^{t-1}\right)\left\{1 - e^{\sum_{j\in\partial i} m_{j\setminus i}^{t-1} \nu_{ji}^{t-1}}\right\} && (6.32a) \\ &= m_i^{t-1} + \left(1 - m_i^{t-1}\right)\left\{1 - \prod_{j\in\partial i}\left(1 - \lambda_{ji}^{t-1}\right)^{m_{j\setminus i}^{t-1}}\right\} && (6.32b) \\ &\approx m_i^{t-1} + \left(1 - m_i^{t-1}\right)\sum_{j\in\partial i} \lambda_{ji}^{t-1} m_{j\setminus i}^{t-1}. && (6.32c) \end{aligned}$$

We checked that applying the message passing with the infection time representation (Eqs. (6.25) - (6.29)), converges to the same result obtained by running the above equation (6.31a): in particular, since the latter does not require an iterative scheme, it is sufficient to run it starting from $t = 0$ in a single swipe.

A comparison between the Small Coupling Dynamic Cavity (SCDC) method proposed in the chapter, the Individual-based Mean Field approximation (IMF) described in section 1.3.1 and the Belief Propagation (BP) method from chapter 4, together with the average of Monte Carlo (MC) simulations (10^5 at most), is shown in Figure 6.1 for the forward dynamics. For all the methods, the same prior is used, with each individual being infected at the initial time $t = 0$ with probability $\gamma_0 = {}^1\!/_N$, independent on others. As an overall remark, it is important to notice how all MF-like methods tend to overestimate the number of infected individuals: in particular, SCDC seems to have better performance than Individual-based Mean Field but worse than Belief Propagation, whose result is exact in the case of acyclic graphs (like the Bethe lattice shown in the top-left panel). This effect is much more evident at later times, and shows that mean field methods tend to represent epidemics with at least one infected individual in the initial condition: in fact, a significant fraction of Monte Carlo simulations have no patient zeros, and removing those yields a result that is much closer to the mean field methods. The Belief Propagation method is able to reproduce exactly the MC result in the case of the Bethe lattice, due to the absence of loops in the graph. Its performance degrades in the other cases, such as the regular random (RRG) and proximity random graphs, when the contact graph contains more and more loops, and becomes closer to the MF method.

6.4 Risk inference performance

In this section, the performance of the new method in the setting of risk inference will be analysed. The risk inference problem has already been discussed in section 4.1, where two methods for the approximate solution have been derived, the Simple Mean Field (SMF) and Belief Propagation (BP) algorithms. In this section, the method proposed in this chapter, SCDC, will be compared with these two methods.

In order to perform this task, in both the BP and SCDC method the marginal probabilities produced by the algorithms will be used, after running the message passing until convergence. For the SCDC method, however, if after several iterations (more then 1600) it has not converged, the marginal probability of infection is computed as an average over many iterations of the message passing.

In order to thoroughly test the SCDC method, several epidemic cascades have been generated with the SI model, on both synthetic and real-world contact networks. Then, observations

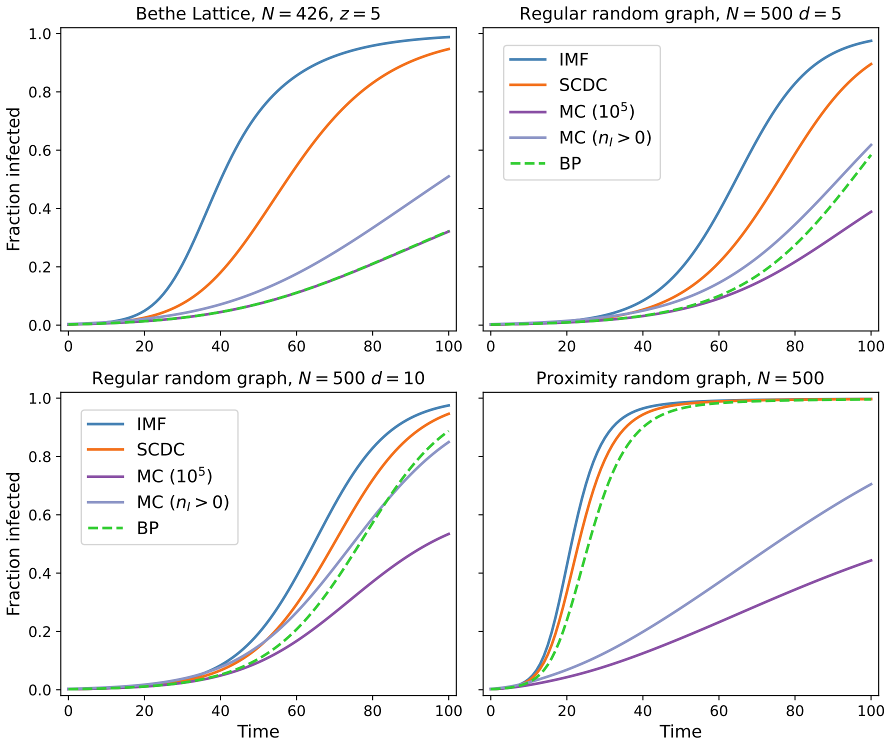

Figure 6.1: **Forward dynamics with SCDC and other mean field methods**. Here, the fraction of infected individuals is shown against the running time of the epidemic, with four different static contact graphs. All the mean field methods tend to overestimate this quantity, with respect to the Monte Carlo average. The probability of each individual being infected at time $t = 0$ is $\gamma_0 = {}^1/_N$ for all methods. The proximity graph is obtained by placing nodes at random in the unit square, and then connecting nodes that are at distance $r < r_{\text{max}} = \sqrt{\frac{2.8}{N}}$ with probability $p(r) = e^{-r}$. From top-left to bottom-right, each contact graph contains more and more loops: while in the RRG the loops are long-range, in the proximity graph there are many short-range loop and triangles. Very small values of the infectiousness λ are used: $\lambda_{\text{bethe}} = 0.06$, $\lambda_{\text{RRG}} = 0.02$ for $d = 5$ and $\lambda_{\text{RRG}} = 0.01$ for $d = 10$, $\lambda_{\text{proximity}} = 0.04$.

are made on the obtained epidemic cascade by randomly drawing a small number of individuals n_{obs} which are observed at the final time instant. To discover the accuracy of each method, we compute the Receiver Operating Characteristic curve (ROC) on the probability of unobserved individuals being infected at last time instant, with the simulated epidemic cascade as the ground truth.

In Figure 6.2 the results for different kinds of synthetic contact graphs are shown. First, there are two instances of static random graphs: Watts-Strogatz graphs (WS) [11] and soft random geometric graphs [53], which will be called *proximity* graphs in the following. For both kinds, we run epidemics with $N = 600$ and observe only a fraction of the nodes at random at the final time instant ($n_{\text{obs}} = 20$). Other parameters for the Watts-Strogatz graphs are the number of neighbours each node is connected to in the beginning, $d = 4$, and the probability for each edge to be rewired, $p_{rw} = 0.12$. In the proximity random graphs, the nodes are placed at random in the unit square, and then the nodes that are at distance $r < r_{\max} = \sqrt{\frac{2.8}{N}}$ are connected with a probability $p(r) = e^{-r}$. For each risk inference test instance that is run on these graphs, a new random graph is generated.

Results are shown for two other synthetic contact networks that are extracted from two agent-based models, the OpenABM-Covid19 [131] and Covasim [134] model, respectively. As already mentioned in the thesis, these models are able to generate realistic contact networks on large populations, that change during each day to reflect the dynamical nature of real-life interactions. For each model, one such network is extracted with a moderate time scale, that is then used to generate other epidemic cascades using the SI model. Also, in these models, to each generated contact between individuals i and j, a effective weight w_{ij} is assigned that reflects the duration and the intensity of the interaction between i and j. This weight is used to give an estimate of the probability of infection λ_{ij} used in the generation of epidemic cascades, as $\lambda_{ij} = 1-(1-\gamma)^{w_{ij}}$. In Figure 6.2 it can be seen that, when a relative small number of observations is given at last time instant, the SCDC method is able to outperform the Simple Mean Field heuristic and achieve accuracy on par with the BP method.

The same testing framework is also employed on two real contact networks, that have been collected with RFID tags in a *school* and in an *office* setting (respectively, the Thiers13 and InVS15 datasets from [61]), with different time periods (one week for the school, two weeks for the office dataset). These contacts are timestamped with a temporal resolution of 20 seconds, allowing for coarse-graining at different time resolutions. When performing this procedure, the number of contacts c^t_{ij} collected between i and j in time window t is used to compute the probability of infection between the two individuals at time t, $\lambda^t_{ij} = 1-(1-\gamma)^{c^t_{ij}}$, where γ is the parameter describing the infectiousness of a single contact. Time window sizes vary with both contact networks, ranging from 3 hours to a day. Four contact networks are obtained with this procedure, that have moderately high number of individuals ($N = 219$ for the office, $N = 328$ for the school) and span either $T = 12$ or $T = 24$ time instants for the office, and $T = 18$ or $T = 36$ time instants for the school. Risk inference results for real contact networks are shown in Figure 6.3: here it can be seen that the SCDC method has a performance very close to the BP method, and in general superior than the SMF heuristic, as in the case of random contact graphs.

6.5 Discussion

In this chapter, a new method for epidemic inference has been presented. This method, based on a small coupling expansion of the dynamic cavity equations for the SI model, belongs to the mean-field algorithms class, like the Simple Mean Field and Belief Propagation methods that have been previously discussed. While the new method is close to the mean field approximation, it incorporates the same Bayesian framework from which the BP method is derived, resulting in a far more accurate treatment of observations than the Simple Mean Field. This is made possible by the presence of the local fields μ, that carry the information on the observations on the neighbouring nodes backward in epidemic time. In fact, the message passing equations (see for example eq. (6.19) and (6.20), or (6.25) and (6.27)) support this interpretation, with the $m_{i\setminus j}$ messages carrying the probability of infection forward in epidemic time.

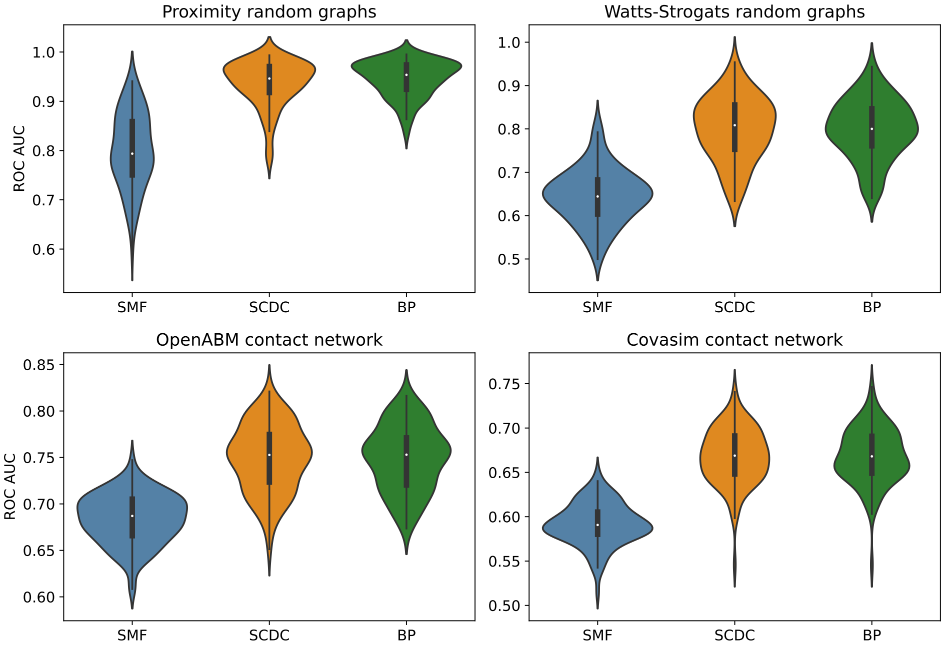

Figure 6.2: **Risk inference accuracy on random contact graphs**: in the top row, static random graphs are used (proximity and Watts-Strogatz) with $N = 600$ nodes, and $n_{obs} = 20$ random observations are made at the last time instant of the epidemic ($T = 28$ and $T = 25$, respectively). In the bottom row, the contact networks employed have been generated with the OpenABM ($N = 2000$, $T = 21$) and Covasim ($N = 1000$, $T = 24$), with $n_{obs} = 60$ and $n_{obs} = 40$ the number of individuals observed at random in each case, respectively. In all four cases, the distribution of the area under the curve (AUC) for the ROC curves for 100 epidemic realizations is shown as a violinplot. Epidemic parameters are $\lambda_{prox} = 0.08$ for the proximity graphs, $\lambda_{WS} = 0.16$ for Watts-Strogatz; $\gamma = 0.038$ for Covasim, and $\gamma = 0.026$ for the OpenABM contacts.

Results obtained on the risk inference problem show that the SCDC method is able to exploit the information given by the observations, achieving performance levels comparable with Belief Propagation. However, the implementation of the method suffers of convergence issues in some cases, requiring to perform an average of the marginals over the message passing in order to get good results. The method is under active development, and this could lead to a resolution of these issues. Moreover, while the infection times representation puts a limit on the possible epidemic models utilized to only those with non-recurring epidemic states, the method can be used without this representation, and it could be extended to recurring epidemic models such the SIS and SIRS, in which one individual can become Susceptible once again after being infected.

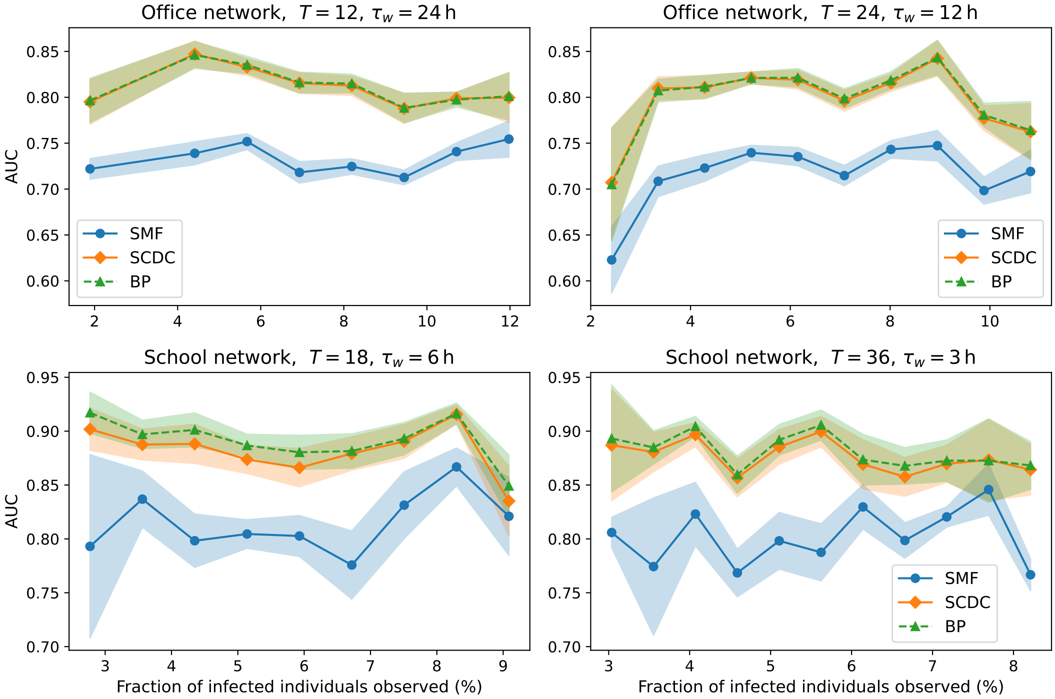

Figure 6.3: **Risk inference accuracy on real contact networks**: in all plots, the area under the curve (AUC) of the ROC curves is shown for all the methods, with the observed fraction of infected in the x axis. For each pane, 100 epidemic realizations are made with a variable number n_{obs} of individuals is observed at the last time instant ($n_{obs} = 15$ for the office network, and $n_{obs} = 20$ in the school network). Then, the AUCs are grouped together by the fraction of infected (of each epidemic cascade) that has been observed. The shaded area represents the error margin for the instances found in each infection fraction bin. The value of γ used for the generation is the same for each contact network: $\gamma = 6 \cdot 10^{-4}$ for the office and $\gamma = 7 \cdot 10^{-5}$ for the school. For reference, the number of individuals is $N = 219$ for the office and $N = 328$ for the school network. Each plot also indicates the width of the time window τ_w used for the coarse-graining of the contacts.

Chapter 7

Conclusions and future perspectives

This PhD thesis has focused on the subject of epidemic inference, that can be described as exploiting the information of both the observations and the model of the epidemic process in order to gain insight on an ongoing epidemic outbreak. This is an important subject in the context of public health, where predictions of possible epidemics are often needed. The COVID-19 pandemic in particular has highlighted the need for the study of these processes, and provided an important application for inference methods. However, epidemic inference at the individual scale is a highly non-trivial problem, even with simple stochastic models such as the Susceptible-Infected-Recovered model, as the computational cost required becomes prohibitive very fast with the scale of the considered system. For this reason, it is necessary to find methods providing good approximate solutions. The already developed Belief Propagation (BP) based method provides better performance then simpler mean field methods, but it suffers of convergence issues with very dense contact networks.

One of the objectives of this thesis, then, has been the development of new techniques to perform inference of the state of any individual at any time instant. The first of such methods is derived from the application of Autoregressive Neural Networks (ANNs, chapter 3). This approach leverages the generative power of ANNs to learn the posterior distribution, with a peculiar training approach in which the epidemic cascade samples created by the ANNs are employed. In this way, the neural networks encode the full distribution, which in the ideal case is very similar to the true posterior. Then, the method is able to estimate the marginal probabilities of any individual by the generation of new epidemic samples. This approach has been revealed to be very good in the approximation of the posterior distribution, with performance comparable to the BP method and better than other algorithms or heuristics, like Soft Margin and Simple Mean Field. In particular in real contact graphs, in the patient zero detection problem the ANN approach delivered better results than BP. Moreover, the neural network framework can be leveraged to perform inference of the epidemic parameters. In this task it is able to achieve on par or much lower error than Belief Propagation, too. The method's accuracy, however, comes at a high computational cost. This is due to the training procedure, which needs to generate a very large number of epidemic samples. Moreover, an annealing scheme has to be employed to help the gradient descent explore all the configurations compatible with the observations, which is time consuming. Overall, the ANN method is deemed suitable for inference with small size contact networks (of ~ 100 individuals). Future investigations of the approach could focus on making the training faster and lowering the memory footprint of method, so that it can be applied on larger systems. In particular, using different neural network architecture or different deep learning frameworks could bring improvements in this direction. The approach could also be applied to other epidemic spreading models.

Another approach investigated during the thesis is based on the dynamic cavity equations (chapter 6). This treatment of the epidemic inference problem results in the Small Coupling Dynamic Cavity (SCDC) method, whose set of local self-consistent equations can be solved using message passing. The distributed nature of the message passing solution makes the method quite efficient in practice, enabling its application on moderate-sized contact networks. Moreover, the set of messages and coupled fields in the equations (m and μ) lend themselves to the physical interpretation of carrying the infection state information forward and backward in epidemic time, starting from the observations. In fact, it has been proved that in the absence of observations the μ fields vanish, and the method's equations can be reduced to simpler mean field ones. In practice, the method retains mean-field like performance in the forward problem (without observations), and it is able to match Belief Propagation in the epidemic risk inference problem, surpassing other mean field methods. In order to improve the method, more work will be dedicated in the future to understand the properties of the convergence of the message passing equations. Moreover, the method could be extended to more complex epidemic models, like the SIR or to recurring epidemic models such as the Susceptible-Infected-Susceptible. In this latter case however it would not be possible to use the implementation derived in the thesis, as it is valid only for non-recurring epidemic models.

During the course of the COVID-19 pandemic, it has become clear that it is of utmost importance, when trying to containing the spread of a disease, to identify and isolate infectious individuals as quickly and efficiently as possible. This led to the deployment of digital contact tracing methods, in order to evaluate the risk of infection at the individual level. The other aim of this thesis, then, has been to investigate the efficacy of epidemic inference methods derived from statistical physics, such as Belief Propagation and Simple Mean Field (SMF), in enhancing contact tracing. In chapter 4, these methods have been applied for the containment of both simple SIR epidemic outbreaks and one COVID-19-like one, the OpenABM-Covid19 model. In both cases, the simulated system size has been very large, accounting for $N \sim 10^5$ individuals and $T = 100$ days, and the methods have been shown to contain the outbreaks much better than simple contact tracing. In the OpenABM model, in particular, containment experiments have been run with different number of daily tests, in order to show that the probabilistic ranking algorithms require less testing to completely stop an epidemic outbreak. Then, it has been shown how they are resistant to testing noise: the BP method in particular is able to completely contain an epidemic even with significant noise. However, reducing the fraction of contacts available to the methods has shown that the containment performance degrades significantly.

The effectiveness of epidemic inference methods has been investigated more deeply in chapter 5. Here, two new models for COVID-19 spreading have been introduced, the Spatiotemporal Epidemic Model (StEM) and Covasim. On these models, the containment performance of the algorithms has been investigated, with a containment setup in which a significant false negative error rate was applied to testing. Nevertheless, results showed that SMF and BP are able to improve on contact tracing methods, and even stop the propagation of the disease. Most importantly, the threshold testing variant of both methods has been introduced, showing the potential ability of containing the epidemic with lower testing requirements than their variant with fixed number of tests per day. Two different implementations of contact tracing were also used in this chapter of the thesis, the former being quite simple (DCT), and the latter more through (ICT). Then, each algorithm has been tested for the ability to detect super-spreader individuals during an epidemic outbreak. Finding and isolating super-spreaders is crucial for arresting the spreading of COVID-19, and both SMF and BP have shown to improve in their detection compared to contact tracing methods. Finally, the ranking algorithms have also been applied to the reconstruction of transmission chains, with BP and SMF showing superior results with respect to contact tracing methods in the detection of backward and forward propagations.

As already mentioned, the results shown in chapter 5 are not final. More work will be devoted the application of epidemic inference methods to these epidemic models. Epidemic models like the Spatiotemporal Epidemic Model, in particular, provide an interesting challenge because of the highly detailed description from which epidemic features of COVID-19 spreading emerge. One research question that could be addressed in the future relates to the drop in performance efficiency (if any) that could arise because of testing delays, a factor which has not been treated so far in our research. However, it is important to remember that any risk inference method suffers from poor collection of contacts in the population. Therefore, it will be fundamental in future large scale epidemic outbreaks, or pandemics like COVID-19, that new ways are found to convince the general public of the benefits of digital contact tracing, so that the majority of the population participates in the contact collection, without renouncing to privacy.

Appendices

Appendix A

Simplified conditional probability on acyclic graph

In this appendix it will be shown that in the case of epidemic spreading in a acyclic interaction networks, it is possible to restrict the dependence of previous nodes in the conditional probabilities of Eq. (3.3), to the second neighbours with lower index. This proof is part of the work published in [158].

Let us consider the configuration $\underline{x}$ of N variables, a probability distribution $p(\underline{x}) = \frac{1}{Z}\prod_a \psi_a(x_a)$ factorized over a set of factors $A = \{a\}$. We first demonstrate the following statements:

Lemma 1 (Markov blanket)**.** *Let* $p(\underline{x}) = \frac{1}{Z}\prod_a \psi_a(x_a)$*, and* $I \cup J \cup K = \{1, \dots, N\}$ *be disjoint and assume no factor depends on* x_I *and* x_K *simultaneously. Then* $p(x_I, x_K|x_J) = p(x_I|x_J)p(x_K|x_J)$*.*

Proof. Considering proportionality $\propto$ with respect to x_I only, we have that

$$p\left(x_I|x_J, x_K\right) \propto p\left(x_I, x_J, x_K\right) \propto \prod_a \psi_a\left(x_a\right) \propto \prod_{a\in\partial I} \psi_a\left(x_a\right)$$

and

$$p\left(x_I|x_J\right) \propto p\left(x_I, x_J\right) \propto \sum_{x_K}\prod_a \psi_a\left(x_a\right) \propto \prod_{a\in\partial I} \psi_a\left(x_a\right)$$

As both distributions are normalized with respect to x_I, they must be equal. This implies that

$$\begin{aligned} p\left(x_I, x_K|x_J\right) &= p\left(x_I, x_J, x_K\right) p\left(x_J\right)^{-1} \\ &= p\left(x_I|x_J, x_K\right) p\left(x_J, x_K\right) p\left(x_J\right)^{-1} \\ &= p\left(x_I|x_J\right) p\left(x_K|x_J\right) \end{aligned}$$

where the last line follows from the derivation above.

Lemma 2 (Separated neighborhood)**.** *Let* $p(x) = \frac{1}{Z}\prod_a \psi_a(x_a)$*, and* $G = (V \cup A, E)$ *be the associated bipartite factor graph, with* $E = \{(i, a) \in V \times A : \psi_a$ *depends on variable* $x_i\}$*. Let* $I \cup J \cup K \subseteq \{1, \dots, N\}$ *be disjoint and assume that every path from I to K in G must pass through J (equivalently, removing vertices in J leave I and K separated). Then* $p(x_I, x_K|x_J) = p(x_I|x_J)p(x_K|x_J)$*, and* $p(x_I|x_{J\cup K}) = p(x_I|x_J)$

Proof. Let I' be the connected component of I in $G \setminus J$ and $K' = V \setminus (I' \cup J)$. As all paths from I to K pass through J, no factors can depend on variables both in I' and in K'. By Lemma 1, $p(x_{I'}, x_{K'}|x_J) = p(x_{I'}|x_J)p(x_{K'}|x_J)$. Then $p(x_I, x_K|x_J) = \sum_{x_{I' \setminus I}} \sum_{x_{K' \setminus K}} p(x_{I'}, x_{K'}|x_J) = \sum_{x_{I' \setminus I}} p(x_{I'}|x_J) \sum_{x_{K' \setminus K}} p(x_{K'}|x_J) = p(x_I|x_J)p(x_K|x_J)$. Then $p(x_I|x_K, x_J) = p(x_I, x_K|x_J)p(x_K|x_J)^{-1} = p(x_I|x_J)$.

Corollary 1 (Restricted autoregression)**.** *By calling* $I = \{i\}$, $J = \{j < i : j \in \partial i\}$ *and* $K = \{j < i : j \notin \partial i\}$, *we obtain that for an ordering of nodes such that J separates i from K we get* $p(x_i|\{x_j : j \in \partial i, j < i\}) = p(x_i|\{x_j : j < i\})$.

The last corollary is defined for a single node *i*. In considering the problem of approximating the posterior probability of an epidemic spreading process (as done in chapter 3) we distinguish two graphs: the first one, the *contact* graph, the nodes are the time trajectory of the states $\underline{x_i}$ of each individuals and the edges are the contacts between them. The second, the *factor* graph, has as nodes, again, the time trajectory of individuals and as factors those ($\{\Psi_i\}$) in the equation 6 in the main text. If the *contact* graph is acyclic and the nodes are topological ordered [10] then, thanks to the corollary 1, the following identity holds:

$$P(\underline{\mathbf{x}}|\mathcal{O}) = \prod_i P(\underline{x_i}|\mathbf{x}_{<i}) = \prod_i P(\underline{x_i}|\underline{\mathbf{x}}_{\partial^2 < i}) \tag{A.1}$$

where $\mathbf{x}_{\partial^2 < i}$ define the sets of nodes up to the next nearest neighbours (in the *contacts* graph) with index lower than *i*. The set of nodes $\mathbf{x}_{\partial^2 < i}$ in the *contact* graph correspond to the $\mathbf{x}_{\partial < i}$ in the *factor* graph. To better visualize the two graphs, in Fig.A.1 we show two cases of *contact* acyclic graphs. The first one represents a linear chain defined by the contacts among individuals. The corresponding *contact* graph is acyclic but the *factor* graph contains cycles. In this case, for instance, considering the conditional probability of node 7 we have that $P(\underline{x_7}|\underline{\mathbf{x}}_{<i}) = P(\underline{x_7}|\underline{\mathbf{x}}_{\partial^2 < i}) = P(\underline{x_7}|\underline{x_4}, \underline{x_6})$. The node 4,6 separated the nodes 7 from the previous nodes in the *factor* graph, as the corollary 1 require. Similarly, the same concept apply to the second case, where we consider a tree graph. In both cases the nodes are ordered topologically so the equation (A.1) holds for each conditional probability.

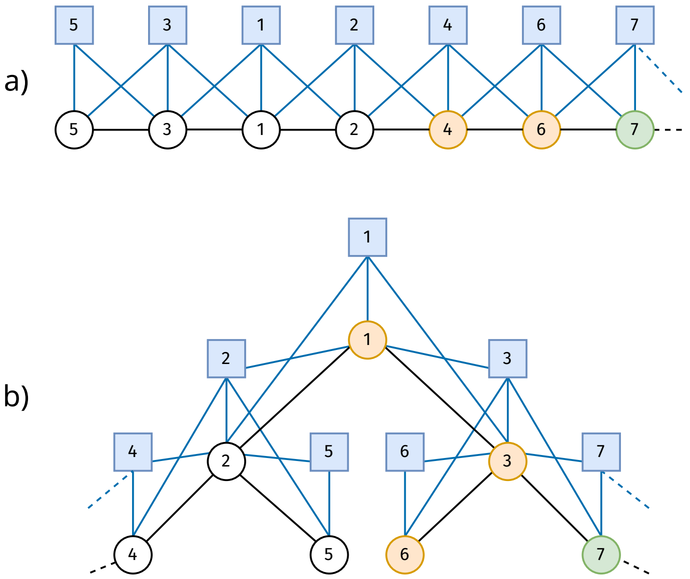

Figure A.1: Two example of acyclic contact graphs ordered topologically. In a) there is a linear chain, and in b) a tree. The squares represents the factor Ψ_i of the posterior probability distribution of the dynamical process we want approximate. The circle represents the time trajectory of the states of each individual i. The black lines are the contacts among individuals, and in both cases they generate a acyclic *contact* graph.

Appendix B

BP equations for the SIR model

The message passing equations are now going to be derived for the factor graph described in section 4.1.2 (we refer to section 2.2 for the theory of the Belief Propagation approach used). Before writing the equations for the messages, however, notice that the distribution R_i is continuous and needs to be discretized. All terms in (4.12) - (4.14), except R_i, are constant as functions of r_i in any interval $\left(\hat{r}_i, \hat{r}_i'\right)$ of consecutive times in $\mathcal{T}_i$: in other words, nothing can happen in this time interval except the recovery of a certain individuals. This fact is exploited by writing $r_i = \hat{r}_i + u_i$, where $\hat{r}_i = \max\left\{r \in \mathcal{T}_i \, : \, r < r_i\right\}$. Then, integrating away the u_i variables in (4.12), it's possible to obtain a fully discretized model for variables $\mathbf{t}, \mathbf{s}, \hat{\mathbf{r}}$, with an expression identical to (4.12) but in which r_i is replaced by $\hat{r}_i$, and $R_i\left(r_i - t_i\right)$ is replaced by

$$\hat{R}_i\left(\hat{r}_i - t_i\right) = \int_{\hat{r}_i - t_i}^{\hat{r}_i' - t_i} R_i\left(u\right) du$$

For simplicity of notation, the $\hat{}$ symbols will be dropped in the following. Now, the belief propagation messages for $\psi_i \to \left(s_{ij}, s_{ji}\right)$ and $\left(s_{ij}, s_{ji}\right) \to \psi_j$ can be condensed into a single message from ψ_i to ψ_j, since there are no other factors that are connected to $\left(s_{ij}, s_{ji}\right)$ (because of (2.4)). Then, the message $m_{ij}\left(s_{ij}, s_{ji}\right)$ can be computed as:

$$\begin{aligned} m_{ij}\left(s_{ij}, s_{ji}\right) \propto & \sum_{t_i} \sum_{r_i} p_{O,i}\left(\mathcal{O}_i | t_i, r_i\right) R_i\left(r_i - t_i\right) S_{ij}\left(s_{ij} | t_i, r_i\right) \times \\ & \times \sum_{\{s_{ki}, s_{ik}\}} A_i\left(\mathbf{s}_{i^*}\right) \delta\big(t_i, \min_{k \in \partial^* i} s_{ki}\big) \prod_{k \in \partial^* i \setminus j} S_{ik}\left(s_{ik} | t_i, r_i\right) m_{ki}\left(s_{ki}, s_{ik}\right) \end{aligned} \tag{B.1}$$

that shows its dependency on messages from other edges (k, i). The marginals for t_i are

$$\begin{aligned} b_i\left(t_i\right) \propto & \sum_{r_i} p_{O,i}\left(\mathcal{O}_i | t_i, r_i\right) R_i\left(r_i - t_i\right) S_{ij}\left(s_{ij} | t_i, r_i\right) \times \\ & \times \sum_{\{s_{ki}, s_{ik}\}} A_i\left(\mathbf{s}_{i^*}\right) \delta\big(t_i, \min_{k \in \partial^* i} s_{ki}\big) \prod_{k \in \partial^* i} S_{ik}\left(s_{ik} | t_i, r_i\right) m_{ki}\left(s_{ki}, s_{ik}\right) \end{aligned} \tag{B.2}$$

and similarly for r_i. A more efficient computation of the equations can be achieved by defining:

$$
\begin{aligned}
G_k^0\left(t_i, r_i\right) &= \sum_{\substack{s_{ki} \geq t_i \\ s_{ik} > t_i}} S_{ik}\left(s_{ik} | t_i, r_i\right) m_{ki}\left(s_{ki}, s_{ik}\right) \\
G_k^1\left(t_i, r_i\right) &= \sum_{\substack{s_{ki} > t_i \\ s_{ik} > t_i}} S_{ik}\left(s_{ik} | t_i, r_i\right) m_{ki}\left(s_{ki}, s_{ik}\right)
\end{aligned}
$$

and substituting the extra auto-infection neighbour messages

$$
m_{i_t^* i}\left(s_{i_t^* i}, s_{i i_t^*}\right) = \begin{cases} \gamma_i^t & s_{i_t^* i} = t, s_{i i_t^*} = \infty \\ 1 - \gamma_i^t & s_{i_t^* i} = \infty, s_{i i_t^*} = \infty \end{cases}
$$

Equations (B.1), (B.2) can be then rewritten as

$$
\begin{aligned}
m_{ij}\left(s_{ij}, s_{ji}\right) \propto & \sum_{t_i} \sum_{r_i} p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right) R_i\left(r_i - t_i\right) S_{ij}\left(s_{ij} \,\middle|\, t_i, r_i\right) \times \qquad \text{(B.3)} \\
& \times \sum_{\{s_{ki}, s_{ik}\}} A_i\left(\mathbf{s}_{i^*}\right) \left(\prod_{k \in \partial^* i} \mathbb{I}\left[s_{ki} \geq t_i\right] - \prod_{k \in \partial^* i} \mathbb{I}\left[s_{ki} > t_i\right] \right) \prod_{k \in \partial^* i \setminus j} S_{ik}\left(s_{ik} \,|\, t_i, r_i\right) m_{ki}\left(s_{ki}, s_{ik}\right) \\
\propto & \sum_{t_i < s_{ji}} \sum_{r_i \geq t_i} p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right) R_i\left(r_i - t_i\right) S_{ij}\left(s_{ij} \,\middle|\, t_i, r_i\right) \times \qquad \text{(B.4)} \\
& \times \prod_{t < t_i} \left(1 - \gamma_i^t\right) \left\{ \prod_{k \in \partial i \setminus j} G_k^0\left(t_i, r_i\right) - \left(1 - \gamma_i^{t_i}\right) \prod_{k \in \partial i \setminus j} G_k^1\left(t_i, r_i\right) \right\} + \\
& + \sum_{r_i \geq s_{ji}} p_{O,i}\left(\mathcal{O}_i \,\middle|\, s_{ji}, r_i\right) R_i\left(r_i - s_{ji}\right) \prod_{k \in \partial i \setminus j} G_k^0\left(s_{ji}, r_i\right) \\
b_{t_i}\left(t_i\right) \propto & \sum_{r_i} R_i\left(r_i - t_i\right) p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right) \times \qquad \text{(B.5)} \\
& \times \prod_{t < t_i} \left(1 - \gamma_i^t\right) \left\{ \prod_{k \in \partial i} G_k^0\left(t_i, r_i\right) - \left(1 - \gamma_i^{t_i}\right) \prod_{k \in \partial i} G_k^1\left(t_i, r_i\right) \right\} \\
b_{r_i}\left(r_i\right) \propto & \sum_{t_i} R_i\left(r_i - t_i\right) p_{O,i}\left(\mathcal{O}_i \,|\, t_i, r_i\right) \times \qquad \text{(B.6)} \\
& \times \prod_{t < t_i} \left(1 - \gamma_i^t\right) \left\{ \prod_{k \in \partial i} G_k^0\left(t_i, r_i\right) - \left(1 - \gamma_i^{t_i}\right) \prod_{k \in \partial i} G_k^1\left(t_i, r_i\right) \right\}
\end{aligned}
$$

Note that products $\prod_{k \in \partial i \setminus j} G_k$ can be computed simultaneously for $j \in \partial i$ in time $O(|\partial i|)$ (by first recursively computing $\prod_{\ell=1}^{\ell'} G_{k_\ell}$ and $\prod_{\ell=\ell'}^{|\partial i|} G_{k_\ell}$ for $\ell' = 1, \dots, |\partial i|$ and then $\prod_{\ell \neq \ell'} G_{k_\ell} = \prod_{\ell=1}^{\ell'-1} G_{k_\ell} \prod_{\ell=\ell'+1}^{|\partial i|} G_{k_\ell}$). The resulting implementation of the update of all messages in factor Ψ_i has complexity $O\left(|\mathcal{T}_i| \sum_{j \in \partial i} \left(|\mathcal{T}_i| + \left|\mathcal{T}_{ij}\right|^2\right)\right)$.

In practice, only a fixed (maximum) number of message passing rounds are done each time new observations are collected. Once this procedure has finished, the marginal probabilities for the SIR states can be computed. For an individual i, the probability to be in each state at time

$t \in \mathscr{T}_i$ can be computed as follows:

$$p\left(x_i^t = S\right) = \sum_{t' \geq t} b_{t_i}(t') \tag{B.7}$$

$$p\left(x_i^t = R\right) = \sum_{r' \leq t} b_{r_i}(r') \tag{B.8}$$

$$p\left(x_i^t = I\right) = 1 - \sum_{t' \geq t} b_{t_i}(t') - \sum_{r' \leq t} b_{r_i}(r') \tag{B.9}$$

where $t', r' \in \mathscr{T}_i$. The simplest estimation of the infection risk for the individual would then be the probability of infection at the current time, $p\left(x_i^t = I\right)$ given from the above equation. However, it has been found more convenient, given the characteristics of COVID-19, to use a criterion that prioritizes recent infections, so that the risk is given by the sum of the marginal probabilities $b_{t_i}(t_i)$ over $t_i \in [t - \delta_{rank}, t]$.

Appendix C

Derivation of the dynamic cavity equations for the SI model

In this appendix, a derivation of the dynamic cavity equations (6.9) is presented. For clarity of exposition, the calculations are carried out for the case of pure time-forward dynamics; the addition of observations is discussed afterwards. The derivation exploits a path-integral representation of the stochastic epidemic dynamics of the SI model, that is based on interpreting the (Markovian) update rule of the discrete-time stochastic process as a set of dynamical constraints for the degrees of freedom under study, i.e. the binary variables $\{x_i^t\}$, and on defining a dynamic partition function of the form

$$Z = \sum_X p(X) \tag{C.1a}$$

$$= \sum_X \prod_{i=1}^N \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t, h_i^t} P\left[r_i^t | h_i^t\right] \delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t\right) \delta\left(h_i^t - \sum_{j=1}^N v_{ji}^t x_j^t \right) \right] \right\} \tag{C.1b}$$

$$= \sum_{X,\mathbf{r}} \int d\mathbf{h} P[\mathbf{r}|\mathbf{h}] \prod_{i=1}^N \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\delta_{\mathrm{K}}\left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t\right) \delta\left(h_i^t - \sum_{j=1}^N v_{ji}^t x_j^t \right) \right] \right\} \tag{C.1c}$$

where $P[\mathbf{r}|\mathbf{h}] = \prod_{i=1}^N \prod_{t=0}^{T-1} \sum_{r_i^t, h_i^t} P\left[r_i^t | h_i^t\right]$ and the time-dependent matrix $\left\{v_{ij}^t\right\}$ already defined in Sec. 6.1.1 encodes the infection rates and possible interaction patterns over time. Using the integral representation of Dirac's and Kronecker's delta functions and substituting the explicit expression of the conditional probabilities $P\left[r_i^t|h_i^t\right]$ in Eq. (6.5), we obtain

$$Z = \sum_{X,\mathbf{r}} \int d\mathbf{h} P[\mathbf{r}|\mathbf{h}] \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\delta_{\mathrm{K}} \left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t \right) \delta \left(h_i^t - \sum_{j=1}^{N} v_{ji}^t x_j^t \right) \right] \right\} \tag{C.2a}$$

$$= \sum_{X} \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t=0,1} \int_{-\infty}^{+\infty} dh_i^t \int_0^{2\pi} \frac{d\hat{x}_i^t}{2\pi} e^{\mathrm{i}\hat{x}_i^t \left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t \right)} \left(1 - e^{h_i^t} \right)^{r_i^t} e^{\left(1 - r_i^t\right) h_i^t} \int_{-\infty}^{+\infty} \frac{d\hat{h}_i^t}{2\pi} e^{\mathrm{i}\hat{h}_i^t \left(h_i^t - \sum_{j=1}^{N} v_{ji}^t x_j^t \right)} \right] \right\} \tag{C.2b}$$

$$= \sum_{X} \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t=0,1} \int_0^{2\pi} \frac{d\hat{x}_i^t}{2\pi} \int_{-\infty}^{+\infty} dh_i^t e^{\mathrm{i}\hat{x}_i^t \left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t \right) + r_i^t \left(\log\left(1 - e^{h_i^t}\right) - h_i^t \right) + h_i^t} \int_{-\infty}^{+\infty} \frac{d\hat{h}_i^t}{2\pi} e^{\mathrm{i}\hat{h}_i^t \left(h_i^t - \sum_{j=1}^{N} v_{ji}^t x_j^t \right)} \right] \right\} \tag{C.2c}$$

$$= \sum_{X} \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t=0,1} \int_{-\infty}^{+\infty} \frac{d\hat{h}_i^t}{2\pi} \int_{-\infty}^{+\infty} dh_i^t e^{\mathrm{i}\hat{h}_i^t h_i^t + r_i^t \left(\log\left(1 - e^{h_i^t}\right) - h_i^t \right) + h_i^t} \int_0^{2\pi} \frac{d\hat{x}_i^t}{2\pi} e^{\mathrm{i}\hat{x}_i^t \left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t \right) - \mathrm{i}\hat{h}_i^t \sum_{j=1}^{N} v_{ji}^t x_j^t} \right] \right\}. \tag{C.2d}$$

It is convenient to proceed isolating the interaction terms and performing the sums over the random variables r_i^t,

$$Z = \sum_{X} \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\sum_{r_i^t=0,1} \int_{-\infty}^{+\infty} \frac{d\hat{h}_i^t}{2\pi} \int_{-\infty}^{+\infty} dh_i^t e^{\mathrm{i}\hat{h}_i^t h_i^t} \int_0^{2\pi} \frac{d\hat{x}_i^t}{2\pi} e^{\mathrm{i}\hat{x}_i^t \left(x_i^{t+1} - x_i^t - (1 - x_i^t) r_i^t \right) + r_i^t \left(\log\left(1 - e^{h_i^t}\right) - h_i^t \right) + h_i^t} \right. \right.$$
$$\left. \left. \times \prod_{j>i} e^{-\mathrm{i}\left(\hat{h}_i^t v_{ji}^t x_j^t + \hat{h}_j^t v_{ij}^t x_i^t \right)} \right] \right\} \tag{C.3a}$$

$$= \sum_{X} \prod_{i=1}^{N} \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\left[\int_{-\infty}^{+\infty} d\hat{h}_i^t \int_0^{2\pi} \frac{d\hat{x}_i^t}{2\pi} \int_{-\infty}^{+\infty} \frac{dh_i^t}{2\pi} e^{\mathrm{i}\hat{h}_i^t h_i^t + h_i^t} \left[1 + e^{-h_i^t - \mathrm{i}\hat{x}_i^t (1 - x_i^t)} - e^{-\mathrm{i}\hat{x}_i^t (1 - x_i^t)} \right] e^{+\mathrm{i}\hat{x}_i^t \left(x_i^{t+1} - x_i^t \right)} \right. \right. \right.$$
$$\left. \left. \left. \times \prod_{j>i} e^{-\mathrm{i}\left(\hat{h}_i^t v_{ji}^t x_j^t + \hat{h}_j^t v_{ij}^t x_i^t \right)} \right] \right] \right\} \tag{C.3b}$$

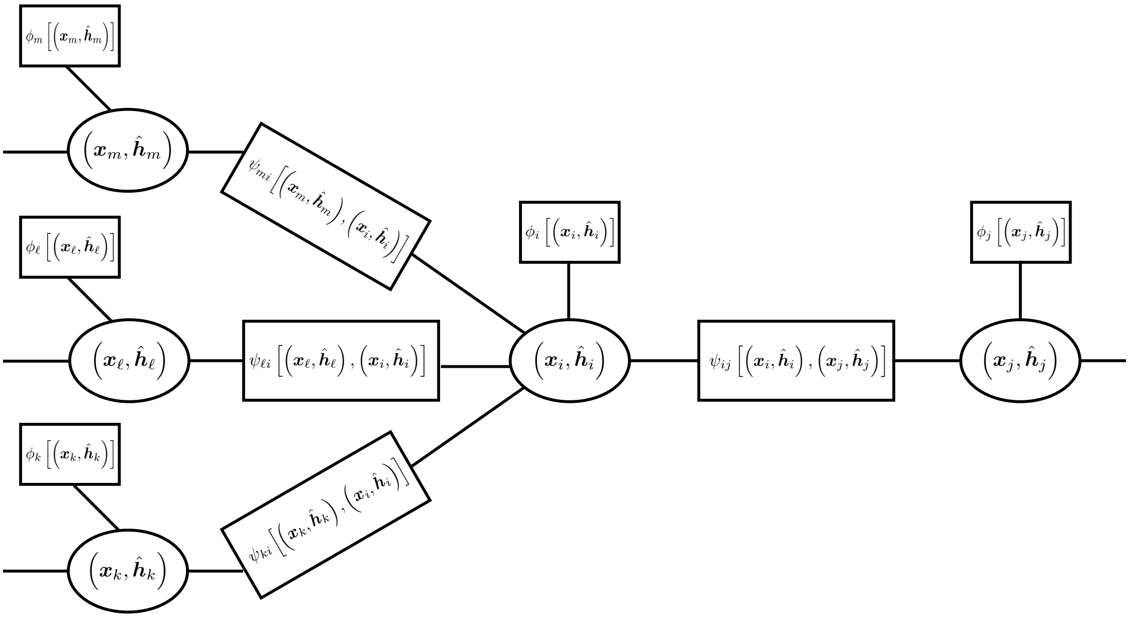

Figure C.1: Factor Graph for the graphical model interpretation of the dynamical partition function in Eq. (C.4) with factors as given in Eqs. (C.5)-(C.6).

then performing the integrals over h_i^t and $\hat{x}_i^t$

$$Z = \sum_X \prod_{i=1}^N \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\left[\int_{-\infty}^{+\infty} d\hat{h}_i^t \int_0^{2\pi} \frac{d\hat{x}_i^t}{2\pi} \left[\delta\left(\hat{h}_i^t - \mathrm{i}\right) \left(e^{\mathrm{i}\hat{x}_i^t\left(x_i^{t+1}-x_i^t\right)} - e^{\mathrm{i}\hat{x}_i^t\left(x_i^{t+1}-1\right)} \right) + \delta\left(\hat{h}_i^t\right) e^{\mathrm{i}\hat{x}_i^t\left(x_i^{t+1}-1\right)} \right] \right.\right.\right.$$
$$\left.\left.\left. \times \prod_{j>i} e^{-\mathrm{i}\left(\hat{h}_i^t v_{ji}^t x_j^t + \hat{h}_j^t v_{ij}^t x_i^t\right)} \right]\right]\right\} \tag{C.4a}$$
$$= \sum_X \prod_{i=1}^N \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\left[\int_{-\infty}^{+\infty} d\hat{h}_i^t \left[\delta\left(\hat{h}_i^t - \mathrm{i}\right) \left(\delta_\mathrm{K}\left(x_i^{t+1} - x_i^t\right) - \delta_\mathrm{K}\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right) \delta_\mathrm{K}\left(x_i^{t+1} - 1\right) \right] \right.\right.\right.$$
$$\left.\left.\left. \times \prod_{j>i} e^{-\mathrm{i}\left(\hat{h}_i^t v_{ji}^t x_j^t + \hat{h}_j^t v_{ij}^t x_i^t\right)} \right]\right]\right\} \tag{C.4b}$$
$$= \sum_X \int D\hat{h} \prod_{i=1}^N \left\{ p(x_i^0) \prod_{t=0}^{T-1} \left[\left[\delta\left(\hat{h}_i^t - \mathrm{i}\right) \left(\delta_\mathrm{K}\left(x_i^{t+1} - x_i^t\right) - \delta_\mathrm{K}\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right) \delta_\mathrm{K}\left(x_i^{t+1} - 1\right) \right] \right.\right.$$
$$\left.\left. \times \prod_{j>i} e^{-\mathrm{i}\left(\hat{h}_i^t v_{ji}^t x_j^t + \hat{h}_j^t v_{ij}^t x_i^t\right)} \right]\right]\right\} \tag{C.4c}$$

where $\int D\hat{h} = \prod_{i=1}^N \prod_{t=0}^{T-1} \left(\int_{-\infty}^{+\infty} d\hat{h}_i^t \right)$ for shortness of notation.

The probabilistic weight associated with the dynamic partition function Z is now in a form that can be represented as a graphical model, in which the variable nodes correspond to the spatio-temporal variables x_i and $\hat{h}_i$ and there are two types of factor nodes (see Fig. C.1): single-node factors

$$\phi_i\left[(x_i,\hat{h}_i)\right] = p(x_i^0)\prod_{t=0}^{T-1}\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta\left(x_i^{t+1}-x_i^t\right)-\delta\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1}-1\right)\right] \tag{C.5}$$

and factors involving pairs of variables on neighboring nodes at the same time

$$\psi_{ij}\left[(x_i,\hat{h}_i),(x_j,\hat{h}_j)\right] = \prod_{t=0}^{T-1} e^{-\mathrm{i}\left(\hat{h}_i^t v_{ji}^t x_j^t+\hat{h}_j^t v_{ij}^t x_i^t\right)}. \tag{C.6}$$

By grouping together single-node variables at all times, that is trajectories

$$(x_i,\hat{h}_i) = (\{x_i^0,\dots,x_i^T\},\{\hat{h}_i^0,\dots,\hat{h}_i^T\}),$$

the resulting factor graph reproduces the topology of the underlying interaction graph. Notice that, the fact that the locally-loopy structure of the factor graph associated to the space-time problem is disentangled by this choice of variable grouping is a consequence of the linear coupling between variables on neighboring nodes, obtained after introducing the auxiliary local fields $\hat{h}$. A different but equivalent formulation of the dynamic cavity considers factor graphs in which the variable nodes contain pairs of trajectories, e.g. (x_i, x_j), of sites that are neighbors on the underlying interaction graph. This is the formulation that leads to the BP method in Refs [37, 38].

According to this graphical model construction, the following *dynamic cavity equations* represent an ansatz for describing the stochastic dynamics associated with the dynamic partition function Z on a tree-like interaction graph,

$$\begin{aligned} c_{ij}\left[x_i,\hat{h}_i\right] = {} & \frac{1}{Z_{ij}}p(x_i^0)\prod_{t=0}^{T-1}\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta_{\mathrm{K}}\left(x_i^{t+1}-x_i^t\right)-\delta_{\mathrm{K}}\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta_{\mathrm{K}}\left(x_i^{t+1}-1\right)\right] \\ & \times\prod_{k\in\partial i\setminus j}\left\{\int D\hat{h}_k\sum_{x_k}c_{ki}\left[x_k,\hat{h}_k\right]e^{-\mathrm{i}\sum_t\left(\hat{h}_i^t v_{ki}^t x_k^t+\hat{h}_k^t v_{ik}^t x_i^t\right)}\right\} \end{aligned} \tag{C.7}$$

where $\int D\hat{h}_i = \prod_{t=0}^{T-1}\left[\int_{-\infty}^{+\infty} d\hat{h}_i^t\right]$. Then, using the Fourier transforms

$$c\left[x,s\right] = \prod_{t=0}^{T-1}\left[\int_{-\infty}^{+\infty} d\hat{h}^t e^{-\mathrm{i}s^t\hat{h}^t}\right]c\left[x,\hat{h}\right] \tag{C.8}$$

$$c\left[x,\hat{h}\right] = \prod_{t=0}^{T-1}\left[\int_{-\infty}^{+\infty}\frac{ds^t}{2\pi}e^{\mathrm{i}s^t\hat{h}^t}\right]c\left[x,s\right], \tag{C.9}$$

the dynamic cavity equations can be written as

$$\begin{aligned} c_{ij}\left[x_i,s_i\right] = {} & \frac{1}{Z_{ij}}p(x_i^0)\prod_{t=0}^{T-1}\left\{\int_{-\infty}^{+\infty} d\hat{h}_i^t e^{-\mathrm{i}s_i^t\hat{h}_i^t}\left[\delta\left(\hat{h}_i^t-\mathrm{i}\right)\left(\delta_{\mathrm{K}}\left(x_i^{t+1}-x_i^t\right)-\delta_{\mathrm{K}}\left(x_i^{t+1}-1\right)\right)+\delta\left(\hat{h}_i^t\right)\delta_{\mathrm{K}}\left(x_i^{t+1}-1\right)\right]\right. \\ & \left.\times\prod_{k\in\partial i\setminus j}\left[\sum_{x_k^t}\int_{-\infty}^{+\infty} d\hat{h}_k^t\int_{-\infty}^{+\infty}\frac{ds_k^t}{2\pi}e^{\mathrm{i}s_k^t\hat{h}_k^t}c_{ki}\left[x_k,s_k\right]e^{-\mathrm{i}\left(\hat{h}_i^t v_{ki}^t x_k^t+\hat{h}_k^t v_{ik}^t x_i^t\right)}\right]\right\}. \end{aligned} \tag{C.10}$$

The expression can be simplified by performing the integrals over the auxiliary variables $\{\hat{h}_k\}$ first,

$$\begin{aligned} c_{ij}\left[x_i, s_i\right] &= \frac{1}{Z_{ij}} p(x_i^0) \prod_{t=0}^{T-1} \Bigg\{ \int_{-\infty}^{+\infty} d\hat{h}_i^t e^{-\mathrm{i} s_i^t \hat{h}_i^t} \left[\delta\left(\hat{h}_i^t - \mathrm{i}\right)\left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1} - 1\right)\right] \\ &\qquad \times \prod_{k \in \partial i \setminus j} \left[\sum_{x_k^t} \int_{-\infty}^{+\infty} ds_k^t c_{ki}\left[x_k, s_k\right] e^{-\mathrm{i}\hat{h}_i^t v_{ki}^t x_k^t} \int_{-\infty}^{+\infty} \frac{d\hat{h}_k^t}{2\pi} e^{\mathrm{i}\hat{h}_k^t\left(s_k^t - v_{ik}^t x_i^t\right)} \right] \Bigg\} \qquad \text{(C.11a)} \\ &= \frac{1}{Z_{ij}} p(x_i^0) \prod_{t=0}^{T-1} \Bigg\{ \int_{-\infty}^{+\infty} d\hat{h}_i^t e^{-\mathrm{i} s_i^t \hat{h}_i^t} \left[\delta\left(\hat{h}_i^t - \mathrm{i}\right)\left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1} - 1\right)\right] \\ &\qquad \times \prod_{k \in \partial i \setminus j} \left[\sum_{x_k^t} \int_{-\infty}^{+\infty} ds_k^t c_{ki}\left[x_k, s_k\right] e^{-\mathrm{i}\hat{h}_i^t v_{ki}^t x_k^t} \delta\left(s_k^t - v_{ik}^t x_i^t\right) \right] \Bigg\} \qquad \text{(C.11b)} \end{aligned}$$

then over the variables $\{s_k\}$,

$$\begin{aligned} c_{ij}\left[x_i, s_i\right] &= \frac{1}{Z_{ij}} p(x_i^0) \prod_{t=0}^{T-1} \Bigg\{ \int_{-\infty}^{+\infty} d\hat{h}_i^t e^{-\mathrm{i} s_i^t \hat{h}_i^t} \left[\delta\left(\hat{h}_i^t - \mathrm{i}\right)\left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1} - 1\right)\right] \\ &\qquad \times \prod_{k \in \partial i \setminus j} \left[\sum_{x_k^t} c_{ki}\left[x_k, v_{ik} x_i\right] e^{-\mathrm{i}\hat{h}_i^t v_{ki}^t x_k^t} \right] \Bigg\} \qquad \text{(C.12a)} \\ &= \frac{1}{Z_{ij}} p(x_i^0) \sum_{\underline{x}_{\partial i \setminus j}} \Bigg\{ \left[\prod_{k \in \partial i \setminus j} c_{ki}\left[x_k, v_{ik} x_i\right] \right] \\ &\qquad \times \prod_{t=0}^{T-1} \left[\int_{-\infty}^{+\infty} d\hat{h}_i^t e^{-\mathrm{i}\hat{h}_i^t\left(s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t\right)} \left(\delta\left(\hat{h}_i^t - \mathrm{i}\right)\left(\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right) + \delta\left(\hat{h}_i^t\right)\delta\left(x_i^{t+1} - 1\right)\right) \right] \Bigg\} \\ &\qquad \qquad \text{(C.12b)} \end{aligned}$$

and finally over $\hat{h}_i$, to obtain a more natural form for the dynamic cavity equations

$$\begin{aligned} c_{ij}\left[x_i, s_i\right] &= \frac{1}{Z_{ij}} p(x_i^0) \sum_{\underline{x}_{\partial i \setminus j}} \Bigg\{ \left[\prod_{k \in \partial i \setminus j} c_{ki}\left[x_k, v_{ik} x_i\right] \right] \\ &\qquad \times \prod_{t=0}^{T-1} \left[\delta\left(x_i^{t+1} - x_i^t\right) e^{s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t} + \delta\left(x_i^{t+1} - 1\right)\left(1 - e^{s_i^t + \sum_{k \in \partial i \setminus j} v_{ki}^t x_k^t}\right) \right] \Bigg\}. \qquad \text{(C.13)} \end{aligned}$$

Due to the locality and independence of observations, the latter can then be included in the above equations as an additional single-node factor term, a local likelihood, to obtain the dynamic cavity equations in Eq. (6.9).

C.1 Reduction to the time-forward equations in the absence of observations

A major consequence of the introduction of time-forward messages $\rho_{t\rightarrow}^{i\setminus j}$ and time-backward messages $\rho_{t\leftarrow}^{i\setminus j}$ is that, in the absence of observations, it is possible to prove that the quantities $\mu_{i\setminus j}^t$ have to vanish for all edges $\forall(i,j)$ and $\forall t$ and then recover a purely time-forward dynamics.

Using the definition of $m_{i\setminus j}^{t}$ by means of the quantities $\rho_{\to t}^{i\setminus j}\left(x_i^t\right)$ and $\rho_{t\leftarrow}^{i\setminus j}\left(x_i^t\right)$, but performing the slicing one time step later, we obtain

$$
\begin{aligned}
m_{i\setminus j}^{t} &= \frac{1}{Z_{ij}^{+}} \sum_{x_i^t, x_i^{t+1}} \rho_{\to t}^{i\setminus j}\left(x_i^t\right) x_i^t M_{x_i^t x_i^{t+1}}^{i\setminus j} \rho_{t+1\leftarrow}^{i\setminus j}\left(x_i^{t+1}\right) && \text{(C.14a)} \\
&= \frac{1}{Z_{ij}^{+}} \rho_{\to t}^{i\setminus j}(1)\, e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{t} \mu_{k\setminus i}^{t}} p\left(O_i^t \mid 1\right) \rho_{t+1\leftarrow}^{i\setminus j}(1) && \text{(C.14b)}
\end{aligned}
$$

with

$$
\begin{aligned}
Z_{ij}^{+} &= \sum_{x_i^t, x_i^{t+1}} \rho_{\to t}^{i\setminus j}\left(x_i^t\right) M_{x_i^t x_i^{t+1}}^{i\setminus j} \rho_{t+1\leftarrow}^{i\setminus j}\left(x_i^{t+1}\right) && \text{(C.15a)} \\
&= \rho_{\to t}^{i\setminus j}(0)\, e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t} \nu_{ki}^{t}} p\left(O_i^t \mid 0\right) \left(\rho_{t+1\leftarrow}^{i\setminus j}(0) - \rho_{t+1\leftarrow}^{i\setminus j}(1)\right) \\
&\quad + \left[\rho_{\to t}^{i\setminus j}(0)\, p\left(O_i^t \mid 0\right) + \rho_{\to t}^{i\setminus j}(1)\, e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{t} \mu_{k\setminus i}^{t}} p\left(O_i^t \mid 1\right)\right] \rho_{t+1\leftarrow}^{i\setminus j}(1) && \text{(C.15b)}
\end{aligned}
$$

or slicing a time step earlier,

$$
\begin{aligned}
m_{i\setminus j}^{t} &= \frac{1}{Z_{ij}^{-}} \sum_{x_i^{t-1}, x_i^{t}} \rho_{\to t-1}^{i\setminus j}\left(x_i^{t-1}\right) x_i^t M_{x_i^{t-1} x_i^{t}}^{i\setminus j} \rho_{t\leftarrow}^{i\setminus j}\left(x_i^{t}\right) && \text{(C.16a)} \\
&= \frac{1}{Z_{ij}^{-}} \sum_{x_i^{t-1}, x_i^{t}} \rho_{\to t-1}^{i\setminus j}\left(x_i^{t-1}\right) x_i^t \left\{\delta\left(x_i^t - x_i^{t-1}\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}} + \delta\left(x_i^t - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right]\right\} \\
&\qquad \times p\left(O_i^{t-1} \mid x_i^{t-1}\right) \rho_{t\leftarrow}^{i\setminus j}\left(x_i^{t}\right) && \text{(C.16b)} \\
&= \frac{1}{Z_{ij}^{-}} \left[\rho_{\to t-1}^{i\setminus j}(0) \left\{1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right\} p\left(O_i^{t-1} \mid 0\right) \rho_{t\leftarrow}^{i\setminus j}(1) + \rho_{\to t-1}^{i\setminus j}(1)\, e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{t-1} \mu_{k\setminus i}^{t-1}} p\left(O_i^{t-1} \mid 1\right) \rho_{t\leftarrow}^{i\setminus j}(1)\right] \\
& && \text{(C.16c)}
\end{aligned}
$$

with

$$
\begin{aligned}
Z_{ij}^{-} &= \sum_{x_i^{t-1}, x_i^{t}} \rho_{\to t-1}^{i\setminus j}\left(x_i^{t-1}\right) \left\{\delta\left(x_i^t - x_i^{t-1}\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}} + \delta\left(x_i^t - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right]\right\} \\
&\qquad \times p\left(O_i^{t-1} \mid x_i^{t-1}\right) \rho_{t\leftarrow}^{i\setminus j}\left(x_i^{t}\right) && \text{(C.17a)} \\
&= \rho_{\to t-1}^{i\setminus j}(0)\, p\left(O_i^{t-1} \mid 0\right) \left(e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}} \rho_{t\leftarrow}^{i\setminus j}(0) + \left\{1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t-1} \nu_{ki}^{t-1}}\right\} \rho_{t\leftarrow}^{i\setminus j}(1)\right) \\
&\quad + \rho_{\to t-1}^{i\setminus j}(1)\, e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{t-1} \mu_{k\setminus i}^{t-1}} p\left(O_i^{t-1} \mid 1\right) \rho_{t\leftarrow}^{i\setminus j}(1)\,. && \text{(C.17b)}
\end{aligned}
$$

Using the previous result it is possible to express $m_{i\setminus j}^{t}$ as function of $m_{i\setminus j}^{t-1}$, where

$$
m_{i\setminus j}^{t-1} = \frac{\rho_{\to t-1}^{i\setminus j}(1)\, e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{t-1} \mu_{k\setminus i}^{t-1}} p\left(O_i^{t-1} \mid 1\right) \rho_{t\leftarrow}^{i\setminus j}(1)}{\sum_{x_i^{t-1}, x_i^{t}} \rho_{\to t-1}^{i\setminus j}\left(x_i^{t-1}\right) M_{x_i^{t-1} x_i^{t}}^{i\setminus j} \rho_{t\leftarrow}^{i\setminus j}\left(x_i^{t}\right)} \tag{C.18}
$$

to obtain

$$m^{t}_{i\setminus j} = m^{t-1}_{i\setminus j} + \frac{\rho^{i\setminus j}_{\to t-1}(0)\, p\left(O^{t-1}_i \mid 0\right)\left\{1 - e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)}{Z^{-}_{ij}} \times \frac{1 - m^{t-1}_{i\setminus j}}{1 - m^{t-1}_{i\setminus j}} \tag{C.19a}$$

$$\begin{aligned} &= m^{t-1}_{i\setminus j} + \frac{1 - m^{t-1}_{i\setminus j}}{\frac{\rho^{i\setminus j}_{\to t-1}(0)p\left(O^{t-1}_i|0\right)e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\rho^{i\setminus j}_{t\leftarrow}(0)+\rho^{i\setminus j}_{\to t-1}(0)p\left(O^{t-1}_i|0\right)\left\{1-e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)}{Z^{-}_{ij}}} \\ &\quad \times \frac{\rho^{i\setminus j}_{\to t-1}(0)\, p\left(O^{t-1}_i \mid 0\right)\left\{1 - e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)}{Z^{-}_{ij}} \end{aligned} \tag{C.19b}$$

$$\begin{aligned} &= m^{t-1}_{i\setminus j} + \left(1 - m^{t-1}_{i\setminus j}\right) \\ &\times \frac{\rho^{i\setminus j}_{\to t-1}(0)\, p\left(O^{t-1}_i \mid 0\right)\left\{1 - e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)}{\rho^{i\setminus j}_{\to t-1}(0)\, p\left(O^{t-1}_i \mid 0\right) e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\rho^{i\setminus j}_{t\leftarrow}(0) + \rho^{i\setminus j}_{\to t-1}(0)\, p\left(O^{t-1}_i \mid 0\right)\left\{1 - e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)} \end{aligned} \tag{C.19c}$$

$$= m^{t-1}_{i\setminus j} + \left(1 - m^{t-1}_{i\setminus j}\right)\frac{\left\{1 - e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)}{e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\rho^{i\setminus j}_{t\leftarrow}(0) + \left\{1 - e^{\sum_{k\in\partial i\setminus j} m^{t-1}_{k\setminus i} v^{t-1}_{ki}}\right\}\rho^{i\setminus j}_{t\leftarrow}(1)}. \tag{C.19d}$$

As already stressed in the main text, the last expression does not represent a time-forward equation because the quantities $\rho^{i\setminus j}_{t\leftarrow}(x)$ are computed backward in time from T to step t. Time-forward dynamics is recovered if the two time-backward messages are equal, which is expected to occur in the absence of observations at later times. To prove this, one can first notice that $\mu^{t}_{i\setminus j} = 0$ if the time-backward messages are equal, i.e. if $\rho^{i\setminus j}_{t+1\leftarrow}(0) = \rho^{i\setminus j}_{t+1\leftarrow}(1)$. In the absence of observation also the inverse implication is true: when the set of messages $\mu^{t}_{i\setminus j}$ at time t are zero and there is no observation also at time t, then the corresponding time-backward messages $\rho^{i\setminus j}_{t\leftarrow}\left(x^t_i\right)$ are also uniform.

Let us start from the time $T-1$, because by construction $\mu^{T}_{i\setminus j} = 0$, then

$$\begin{aligned} \mu^{T-1}_{i\setminus j} &= \sum_{x_i} \frac{p\left(x^0_i\right)}{Z_{ij}}\left(\delta\left(x^T_i - x^{T-1}_i\right) - \delta\left(x^T_i - 1\right)\right) e^{\sum_{k\in\partial i\setminus j}\left(m^{T-1}_{k\setminus i} v^{T-1}_{ki} + x^{T-1}_i v^{T-1}_{ik}\mu^{T-1}_{k\setminus i}\right)} \prod_{\ell=1}^{N^i_O} p\left(O^{t_\ell}_i \mid x^{t_\ell}_i\right) \\ &\times \prod_{t'=0}^{T-1}\left\{\delta\left(x^{t'+1}_i - x^{t'}_i\right) e^{\sum_{k\in\partial i\setminus j} m^{t'}_{k\setminus i} v^{t'}_{ki}} + \delta\left(x^{t'+1}_i - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m^{t'}_{k\setminus i} v^{t'}_{ki}}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x^{t'}_i v^{t'}_{ik}\mu^{t'}_{k\setminus i}} \end{aligned} \tag{C.20a}$$

$$= \frac{\sum_{x^{T-1}_i x^T_i} \rho^{i\setminus j}_{\to T-1}\left(x^{T-1}_i\right)\left(\delta\left(x^T_i - x^{T-1}_i\right) - \delta\left(x^T_i - 1\right)\right) e^{\sum_{k\in\partial i\setminus j}\left(m^{T-1}_{k\setminus i} v^{T-1}_{ki} + x^{T-1}_i v^{T-1}_{ik}\mu^{T-1}_{k\setminus i}\right)} p\left(O^T_i \mid x^T_i\right)}{\rho^{i\setminus j}_{\to T-1}(0)\, q^{i\setminus j}_{\to T-1} + \rho^{i\setminus j}_{\to T-1}(1)\, e^{\sum_{k\in\partial i\setminus j} v^{T-1}_{ik}\mu^{T-1}_{k\setminus i}} p\left(O^T_i \mid 1\right)} \tag{C.20b}$$

$$= \frac{\rho^{i\setminus j}_{\to T-1}(0)\, e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}} p\left(O^T_i \mid 0\right) - \rho^{i\setminus j}_{\to T-1}(0)\, e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}} p\left(O^T_i \mid 1\right)}{\rho^{i\setminus j}_{\to T-1}(0)\, q^{i\setminus j}_{\to T-1} + \rho^{i\setminus j}_{\to T-1}(1)\, e^{\sum_{k\in\partial i\setminus j} v^{T-1}_{ik}\mu^{T-1}_{k\setminus i}} p\left(O^T_i \mid 1\right)}. \tag{C.20c}$$

where

$$q^{i\setminus j}_{\to T-1} = e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}}\, p\left(O_i^T \mid 0\right) + \left[1 - e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} \tilde{v}^{T-1}_{ki}}\right] p\left(O_i^T \mid 1\right) \tag{C.21}$$

If no observation is provided on the final time, then $p\left(O_i^T \mid x_i^T\right) = 1$ for $x_i^T = 0{,}1$ and the numerator vanishes, that is $\mu^{T-1}_{i\setminus j} = 0$. Moreover,

$$\begin{aligned} \rho^{i\setminus j}_{T-1\leftarrow}\left(x_i^{T-1}\right) &= \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} \sum_{x_i^T} M^{i\setminus j}_{x_i^{T-1} x_i^T} p\left(O_i^T \mid x_i^T\right) && \text{(C.22a)} \\ &= \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} \left\{ M^{i\setminus j}_{x_i^{T-1} 0} p\left(O_i^T \mid 0\right) + M^{i\setminus j}_{x_i^{T-1} 1} p\left(O_i^T \mid 1\right) \right\} && \text{(C.22b)} \end{aligned}$$

that is

$$\rho^{i\setminus j}_{T-1\leftarrow}(0) = \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} \left\{ e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}}\, p\left(O_i^{T-1} \mid 0\right) p\left(O_i^T \mid 0\right) + \left(1 - e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}}\right) p\left(O_i^{T-1} \mid 0\right) p\left(O_i^T \mid 1\right) \right\} \tag{C.23}$$

$$\rho^{i\setminus j}_{T-1\leftarrow}(1) = \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} \left\{ e^{\sum_{k\in\partial i\setminus j} v^{T-1}_{ik} \mu^{T-1}_{k\setminus i}}\, p\left(O_i^{T-1} \mid 1\right) p\left(O_i^T \mid 1\right) \right\}. \tag{C.24}$$

Assuming that there is no observation at time T, the two quantities become

$$\begin{aligned} \rho^{i\setminus j}_{T-1\leftarrow}(0) &= \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} \left\{ e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}}\, p\left(O_i^{T-1} \mid 0\right) + \left(1 - e^{\sum_{k\in\partial i\setminus j} m^{T-1}_{k\setminus i} v^{T-1}_{ki}}\right) p\left(O_i^{T-1} \mid 0\right) \right\} \\ &= \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} p\left(O_i^{T-1} \mid 0\right) && \text{(C.25)} \\ \rho^{i\setminus j}_{T-1\leftarrow}(1) &= \frac{1}{Z^{i\setminus j}_{T-1\leftarrow}} p\left(O_i^{T-1} \mid 1\right) && \text{(C.26)} \end{aligned}$$

meaning that $\rho^{i\setminus j}_{T-1\leftarrow}(0) = \rho^{i\setminus j}_{T-1\leftarrow}(1)$ if no observation is included at time $T-1$. In this way, the equality is guaranteed at time $T-1$ and one can proceed by induction. By assuming that, in the absence of observations at times larger than t, the equality is valid for time $t+1$, i.e. $\rho^{i\setminus j}_{t+1\leftarrow}(0) = \rho^{i\setminus j}_{t+1\leftarrow}(1) = \rho^{i\setminus j}_{t+1\leftarrow}$ for all directed edges (i,j), then one obtains that $\mu^{t}_{i\setminus j} = 0$.

Computing the time-backward messages at time t,

$$
\begin{aligned}
\rho^{i\setminus j}_{t\leftarrow}\left(x_i^t\right) &= \frac{\sum_{x_i^{t+1}} \rho^{i\setminus j}_{t+1\leftarrow}\left(x_i^{t+1}\right) M^{i\setminus j}_{x_i^t x_i^{t+1}}}{\sum_{x_i^t, x_i^{t+1}} \rho^{i\setminus j}_{t+1\leftarrow}\left(x_i^{t+1}\right) M^{i\setminus j}_{x_i^t x_i^{t+1}}} && \text{(C.27a)}\\
&= \frac{\rho^{i\setminus j}_{t+1\leftarrow}(0)\, M^{i\setminus j}_{x_i^t 0} + \rho^{i\setminus j}_{t+1\leftarrow}(1)\, M^{i\setminus j}_{x_i^t 1}}{\rho^{i\setminus j}_{t+1\leftarrow}(0)\, e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i} \nu^t_{ki}} + \rho^{i\setminus j}_{t+1\leftarrow}(1) \left[1 - e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i} \nu^t_{ki}}\right] + \rho^{i\setminus j}_{t+1\leftarrow}(1)\, e^{\sum_{k\in\partial i\setminus j} \nu^t_{ik} \mu^t_{k\setminus i}}} && \text{(C.27b)}\\
&= \frac{\rho^{i\setminus j}_{t+1\leftarrow}\left(M^{i\setminus j}_{x_i^t 0} + M^{i\setminus j}_{x_i^t 1}\right)}{\rho^{i\setminus j}_{t+1\leftarrow} e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i} \nu^t_{ki}} p\left(O_i^t \mid 0\right) + \rho^{i\setminus j}_{t+1\leftarrow}\left[1 - e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i} \nu^t_{ki}}\right] p\left(O_i^t \mid 0\right) + \rho^{i\setminus j}_{t+1\leftarrow} e^{\sum_{k\in\partial i\setminus j} \nu^t_{ik} \mu^t_{k\setminus i}} p\left(O_i^t \mid 1\right)} && \text{(C.27c)}\\
&= \frac{M^{i\setminus j}_{x_i^t 0} + M^{i\setminus j}_{x_i^t 1}}{e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i} \nu^t_{ki}} p\left(O_i^t \mid 0\right) + \left[1 - e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i} \nu^t_{ki}}\right] p\left(O_i^t \mid 0\right) + e^{\sum_{k\in\partial i\setminus j} \nu^t_{ik} \mu^t_{k\setminus i}} p\left(O_i^t \mid 1\right)} && \text{(C.27d)}\\
&= \frac{M^{i\setminus j}_{x_i^t 0} + M^{i\setminus j}_{x_i^t 1}}{p\left(O_i^t \mid 0\right) + p\left(O_i^t \mid 1\right) e^{\sum_{k\in\partial i\setminus j} \nu^t_{ik} \mu^t_{k\setminus i}}}, && \text{(C.27e)}
\end{aligned}
$$

and using that all $\mu^t_{k\setminus i}$ vanish,

$$
\rho^{i\setminus j}_{t\leftarrow}\left(x_i^t\right) = \begin{cases} \frac{p(O_i^t|0)}{p(O_i^t|0)+p(O_i^t|1)} & \text{if } x_i^t = 0,\\ \frac{p(O_i^t|1)}{p(O_i^t|0)+p(O_i^t|1)} & \text{if } x_i^t = 1, \end{cases} \tag{C.28}
$$

that is equal to $1/2$ and independent of the value of x_i^t if no observation occurs also at time t. By induction, this is true for every time, as long as no observation is included. Hence, it is possible to conclude that, in the absence of observations, the equations (C.19d) for the cavity marginals $m^t_{i\setminus j}$ reduce to the more standard time-forward mean-field equations in Eqs.(6.31a)-(6.31a).

Appendix D

Dynamic cavity equations implementation

In this appendix, the implementation for the infection times of the dynamic cavity mean field method is derived. Recalling that:

$$c_{ij}\left[x_i, s_i\right] = \frac{1}{Z_{ij}} P\left[x_i^0\right] \prod_{r=0}^{T-1} \left\{\delta\left(x_i^{r+1} - x_i^r\right) e^{s_i^r + \sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} + \delta\left(x_i^{r+1} - 1\right)\left[1 - e^{s_i^r + \sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^r v_{ik}^r \mu_{k\setminus i}^r}, \tag{D.1}$$

the expression for the cavity message $m_{i\setminus j}^t$ reads

$$m_{i\setminus j}^t = \frac{1}{Z_{ij}} \left[\sum_{x_i} x_i^t P\left[x_i^0\right] \prod_{s=0}^{T-1} \left\{\delta\left(x_i^{s+1} - x_i^s\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^s v_{ki}^s} + \delta\left(x_i^{s+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^s v_{ki}^s}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^s v_{ik}^s \mu_{k\setminus i}^s}\right] \tag{D.2}$$

with

$$Z_{ij} = \sum_{x_i} P\left[x_i^0\right] \prod_{t=0}^{T-1} \left\{\delta\left(x_i^{t+1} - x_i^t\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^t v_{ki}^t} + \delta\left(x_i^{t+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^t v_{ki}^t}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^t v_{ik}^t \mu_{k\setminus i}^t} \tag{D.3}$$

Then, the fields $\mu_{i\setminus j}^t$ are defined as:

$$\mu_{i\setminus j}^t = \sum_{x_i^t} \left.\frac{\partial c_{ij}}{\partial s_i^t}\right|_{s_i=0} \tag{D.4}$$

$$\begin{aligned} \mu_{i\setminus j}^t = \sum_{x_i} \frac{P\left[x_i^0\right]}{Z_{ij}} \left[\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right] e^{\sum_{k\in\partial i\setminus j}\left(m_{k\setminus i}^t v_{ki}^t + x_i^t v_{ik}^t \mu_{k\setminus i}^t\right)} \\ \times \prod_{r\neq t}^{T-1} \left\{\delta\left(x_i^{r+1} - x_i^r\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} + \delta\left(x_i^{r+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^r v_{ik}^r \mu_{k\setminus i}^r} \end{aligned} \tag{D.5}$$

Let us rewrite everything using the infection time representation. x_i^t is defined for $t \in \{0,1,\dots,T\}$, so it has $T+1$ values. Remember that by construction $\mu_{i\setminus j}^T = 0$.

Consider the term

$$F_{i\setminus j}(x_i, r) = \left\{\delta\left(x_i^{r+1} - x_i^r\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} + \delta\left(x_i^{r+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^r v_{ik}^r \mu_{k\setminus i}^r} \tag{D.6}$$

Let us rewrite it in terms of the infection time, as $F_{i\setminus j}(r, t_i)$. At fixed t_i we have (assume $\theta[0] = 1$):

$$x_i^r = \theta[r - t_i]$$

Then,

$$\delta\left(x_i^{r+1} - 1\right) = \begin{cases} 0 & r < t_i - 1 \\ 1 & r \geq t_i - 1 \end{cases} = \theta[r - (t_i - 1)]$$

and

$$\begin{aligned}\delta\left(x_i^{r+1} - x_i^r\right) &= \begin{cases} 0 & r = t_i - 1 \\ 1 & r < t_i - 1 \cup r \geq t_i \end{cases} = 1 - \delta_{r,t_i-1} \\ &= \theta[r - t_i] + \theta[t_i - 2 - r]\end{aligned}$$

Therefore, the $F_{i\setminus j}$ can be written as:

$$\begin{aligned}F_{i\setminus j}(r, t_i) &= \left\{\left[1 - \delta_{r,t_i-1}\right] e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} + \theta[r - (t_i - 1)]\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r}\right]\right\} e^{\theta[r-t_i]\sum_{k\in\partial i\setminus j} v_{ik}^r \mu_{k\setminus i}^r} \\ &= \left\{(\theta[r - t_i] + \theta[t_i - 2 - r]) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} + \theta[r - (t_i - 1)]\left[1 - e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r}\right]\right\} e^{\theta[r-t_i]\sum_{k\in\partial i\setminus j} v_{ik}^r \mu_{k\setminus i}^r} \\ &= \left\{(\theta[r - t_i] + \theta[t_i - 2 - r] - \theta[r - (t_i - 1)]) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} + \theta[r - (t_i - 1)]\right\} e^{\theta[r-t_i]\sum_{k\in\partial i\setminus j} v_{ik}^r \mu_{k\setminus i}^r}\end{aligned} \tag{D.7}$$

Consider

$$\theta[r - t_i] + \theta[t_i - 2 - r] - \theta[r - (t_i - 1)] = \begin{cases} 1 & r \leq t_i - 2 \\ -1 & r = t_i - 1 \\ 0 & r \geq t_i \end{cases}$$

So that finally

$$F_{i\setminus j}(r, t_i) = \begin{cases} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^r v_{ki}^r} & r \leq t_i - 2 \\ 1 - \phi(t_i)\, e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i-1} v_{ki}^{t_i-1}} & r = t_i - 1 \\ e^{\sum_{k\in\partial i\setminus j} v_{ik}^r \mu_{k\setminus i}^r} & r \geq t_i \end{cases}$$

$$\phi(t_i) = \begin{cases} 1 & 1 \leq t_i \leq T \\ 0 & \text{else} \end{cases}$$

As a final remark, we rewrite the initial condition $P\left[x_i^0\right]$ as $\pi_i^0(t_i)$ where

$$\pi_i^0(r) = \begin{cases} \gamma_i^0 & r = 0 \\ 1 - \gamma_i^0 & r > 0 \end{cases}$$

D.1 The case of no observations

Let us start from the case with no observations, because of the simplicity. Let us rewrite the three terms Z_{ij}, $m^t_{i\setminus j}$ and $\mu^t_{i\setminus j}$ in the infection time representation.

D.1.1 Formulation for the messages and normalization

Messages

Let us start with the normalization of the messages. Using the definitions above, we get

$$\begin{aligned}
Z_{i\setminus j} &= \sum_{x_i} P\left[x_i^0\right] \prod_{r=0}^{T-1} \left\{\delta\left(x_i^{r+1} - x_i^r\right) e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i} \nu^r_{ki}} + \delta\left(x_i^{r+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^r \nu^r_{ik}\mu^r_{k\setminus i}} \\
&= \sum_{t_i=0}^{T+1} \pi_i^0\left(t_i\right) \prod_{r=0}^{T-1} F_{i\setminus j}\left(r, t_i\right) \\
&= \sum_{t_i=0}^{T+1} \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right]\left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m^{t_i-1}_{k\setminus i}\nu^{t_i-1}_{ki}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu^r_{ik}\mu^r_{k\setminus i}}
\end{aligned} \tag{D.8}$$

Let us now do the same for the messages $m_{i\setminus j}$. At each $t \in \{0, T\}$,$x_i^t = \theta\left[t - t_i\right]$:

$$\begin{aligned}
m^t_{i\setminus j} &= \frac{1}{Z_{i\setminus j}} \sum_{x_i} x_i^t P\left[x_i^0\right] \prod_{r=0}^{T-1} \left\{\delta\left(x_i^{r+1} - x_i^r\right) e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}} + \delta\left(x_i^{r+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^r \nu^r_{ik}\mu^r_{k\setminus i}} \\
&= \frac{1}{Z_{i\setminus j}} \sum_{t_i=0}^{T-1} \theta\left[t - t_i\right] \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right]\left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m^{t_i-1}_{k\setminus i}\nu^{t_i-1}_{ki}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu^r_{ik}\mu^r_{k\setminus i}} \\
&= \frac{1}{Z_{i\setminus j}} \sum_{t_i=0}^{t} \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right]\left[1 - \phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m^{t_i-1}_{k\setminus i}\nu^{t_i-1}_{ki}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu^r_{ik}\mu^r_{k\setminus i}}
\end{aligned} \tag{D.9}$$

Then for the fields $\mu^t_{i\setminus j}$:

$$\begin{aligned}
\mu^t_{i\setminus j} = \sum_{x_i} \frac{P\left[x_i^0\right]}{Z_{ij}} &\left[\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right)\right] e^{\sum_{k\in\partial i\setminus j}\left(m^t_{k\setminus i}\nu^t_{ki} + x_i^t \nu^t_{ik}\mu^t_{k\setminus i}\right)} \times \\
&\times \prod_{r\neq t}^{T-1} \left\{\delta\left(x_i^{r+1} - x_i^r\right) e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}} + \delta\left(x_i^{r+1} - 1\right)\left[1 - e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right]\right\} e^{\sum_{k\in\partial i\setminus j} x_i^r \nu^r_{ik}\mu^r_{k\setminus i}}
\end{aligned} \tag{D.10}$$

$$\delta\left(x_i^{t+1} - x_i^t\right) - \delta\left(x_i^{t+1} - 1\right) = \begin{cases} 1 & t \le t_i - 2 \\ -1 & t = t_i - 1 \\ 0 & t \ge t_i \end{cases}$$

So, t_i must be always $t_i \ge t + 1$, but we have two different terms depending on the weight.

$$
\begin{aligned}
\mu_{i\setminus j}^{t} = \frac{1}{Z_{i\setminus j}} \sum_{t_i=t+2}^{T+1} \pi_i^0\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t} \nu_{ki}^{t}} \left[\prod_{\substack{r=0\\ r\neq t}}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \left[1-\phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{r} \mu_{k\setminus i}^{r}} + \\
- \frac{1}{Z_{i\setminus j}} \mathbb{I}\left[0\leq t\leq T\right] e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t} \nu_{ki}^{t}} \pi_i^0\left(t+1\right) \left[\prod_{r=0}^{t-1} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \prod_{r=t+1}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{r} \mu_{k\setminus i}^{r}} \\
= \frac{1}{Z_{i\setminus j}} \sum_{t_i=t+2}^{T+1} \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \left[1-\phi\left(t_i\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{r} \mu_{k\setminus i}^{r}} + \\
- \frac{1}{Z_{i\setminus j}} \mathbb{I}\left[0\leq t\leq T\right] \pi_i^0\left(t+1\right) \left[\prod_{r=0}^{t} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \prod_{r=t+1}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{r} \mu_{k\setminus i}^{r}}
\end{aligned}
\tag{D.11}
$$

Shifting time indices in the sum and products, the result is

$$
\begin{aligned}
\mu_{i\setminus j}^{t} = \frac{1}{Z_{i\setminus j}} \sum_{t_i=t}^{T-1} \pi_i^0\left(t_i+2\right) \left[\prod_{r=0}^{t_i} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \left[1-\phi\left(t_i+2\right) e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{t_i+1} \nu_{ki}^{t_i+1}}\right] \prod_{r=t_i+2}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{r} \mu_{k\setminus i}^{r}} + \\
- \frac{1}{Z_{i\setminus j}} \mathbb{I}\left[0\leq t\leq T\right] \pi_i^0\left(t+1\right) \left[\prod_{r=0}^{t} e^{\sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \prod_{r=t+1}^{T-1} e^{\sum_{k\in\partial i\setminus j} \nu_{ik}^{r} \mu_{k\setminus i}^{r}}
\end{aligned}
\tag{D.12}
$$

The only non-zero configurations here are those with $t_i = T$ and $t_i = T+1$ with different weight.

Marginals

By analogy with the message $m_{i\setminus j}^{t}$, we can write

$$
\begin{aligned}
Z_i &= \sum_{t_i=0}^{T+1} \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \left[1-\phi\left(t_i\right) e^{\sum_{k\in\partial i} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i} \nu_{ik}^{r} \mu_{k\setminus i}^{r}} \\
m_i^t &= \frac{1}{Z_i} \sum_{t_i=0}^{t} \pi_i^0\left(t_i\right) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i} m_{k\setminus i}^{r} \nu_{ki}^{r}}\right] \left[1-\phi\left(t_i\right) e^{\sum_{k\in\partial i} m_{k\setminus i}^{t_i-1} \nu_{ki}^{t_i-1}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i} \nu_{ik}^{r} \mu_{k\setminus i}^{r}}
\end{aligned}
$$

D.2 Compact representation

Let us introduce now some definitions that are going to be used to rewrite the quantities in a cleaner way, in the case where there are no observations. Let us define M and N for the cavity messages:

$$
M_{\partial i\setminus j}\left(r\right) = \sum_{k\in\partial i\setminus j} m_{k\setminus i}^{r} \nu_{ki}^{r} \tag{D.13}
$$

$$
N_{\partial i\setminus j}\left(r\right) = \sum_{k\in\partial i\setminus j} \mu_{k\setminus i}^{r} \nu_{ik}^{r} \tag{D.14}
$$

and also a non-cavity version of the above quantities:

$$
\begin{aligned}
M_{\partial i}\left(r\right) &= \sum_{k\in\partial i} m_{k\setminus i}^{r} \nu_{ki}^{r} \\
&= M_{\partial i\setminus j}\left(r\right) + m_{j\setminus i}^{r} \nu_{ji}^{r}
\end{aligned}
\tag{D.15}
$$

$$\begin{aligned} N_{\partial i}(r) &= \sum_{k\in\partial i} \mu^r_{k\setminus i} \nu^r_{ik} \\ &= N_{\partial i\setminus j}(r) + \mu^r_{j\setminus i}\nu^r_{ij} \end{aligned} \tag{D.16}$$

D.2.1 Messages

When there are no observations, the denominator of the messages is:

$$Z_{i\setminus j} = \sum_{t_i=0}^{T+1} \pi_i^0(t_i) \left[\prod_{r=0}^{t_i-2} e^{\sum_{k\in\partial i\setminus j} m^r_{k\setminus i}\nu^r_{ki}}\right] \left[1 - \phi(t_i)\, e^{\sum_{k\in\partial i\setminus j} m^{t_i-1}_{k\setminus i}\nu^{t_i-1}_{ki}}\right] \prod_{r=t_i}^{T-1} e^{\sum_{k\in\partial i\setminus j}\nu^r_{ik}\mu^r_{k\setminus i}}$$

In order to streamline the calculation, let us define

$$G_{\partial i\setminus j}(s) = \sum_{r=0}^{s} M_{\partial i\setminus j}(r) \tag{D.17}$$

$$V_{\partial i\setminus j}(s) = \sum_{r=s}^{T-1} N_{\partial i\setminus j}(r) \tag{D.18}$$

Since $\phi(t_i) = 1$ only for $t_i \geq 1$, $\phi(t) = \theta(t-1)$. Then it's possible to define $L_{\partial i\setminus j}$:

$$L_{\partial i\setminus j}(t) = 1 - \phi(t+1)\, e^{\sum_{k\in\partial i\setminus j} m^t_{k\setminus i}\nu^t_{ki}} = 1 - \phi(t+1)\, e^{M_{\partial i\setminus j}(t)}, \tag{D.19}$$

which has the nice property:

$$\log L_{\partial i\setminus j}(t) = \phi(t+1)\log\left(1 - e^{M_{\partial i\setminus j}(t)}\right) = \theta(t)\log\left(1 - e^{M_{\partial i\setminus j}(t)}\right) \tag{D.20}$$

Now the calculation of the $Z_{i\setminus j}$ reads as:

$$Z_{i\setminus j} = \sum_{t_i=0}^{T+1} \exp\left\{H_{\partial i\setminus j}(t_i)\right\} \tag{D.21}$$

where

$$H_{\partial i\setminus j}(t) = \log \pi_i^0(t) + G_{\partial i\setminus j}(t-2) + V_{\partial i\setminus j}(t) + \log L_{\partial i\setminus j}(t-1) \tag{D.22}$$

Also, the messages can be computed as

$$m^t_{i\setminus j} = \frac{1}{Z_{i\setminus j}} \sum_{t_i=0}^{t} \exp\left\{H_{\partial i\setminus j}(t_i)\right\} \tag{D.23}$$

and the fields μ as:

$$\mu^t_{i\setminus j} = \frac{\mathbb{I}\left[0 \leq t \leq T-1\right]}{Z_{i\setminus j}} \left\{\sum_{t_i=t+2}^{T+1} \exp\left[H_{\partial i\setminus j}(t_i)\right] - \pi_i^0(t+1)\, e^{G_{\partial i\setminus j}(t) + V_{\partial i\setminus j}(t+1)}\right\} \tag{D.24}$$

D.2.2 Marginals

For the marginals, it would be ideal to rewrite the calculation in a similar way, with

$$H_{\partial i}(t) = \log \pi_i^0(t) + G_{\partial i}(t-2) + V_{\partial i}(t) + \log L_{\partial i}(t-1) \tag{D.25}$$

defining

$$\begin{aligned} G_{\partial i}(t) &= \sum_{r=0}^{t} M_{\partial i}(r) = \sum_{r=0}^{t} M_{\partial i \setminus j}(r) + \sum_{r=0}^{t} m_{j\setminus i}^{r} \nu_{ji}^{r} \\ G_{\partial i}(t) &= G_{\partial i \setminus j}(t) + K_{j\setminus i}^{F}(t) \end{aligned} \tag{D.26}$$

where $K_{j\setminus i}^{F} = \sum_{r=0}^{t} m_{j\setminus i}^{r} \nu_{ji}^{r}$. In the same way, $V_{\partial i}$ is defined as:

$$\begin{aligned} V_{\partial i}(s) &= \sum_{r=s}^{T-1} N_{\partial i}(r) = \sum_{r=s}^{T-1} N_{\partial i \setminus j}(r) + \sum_{r=s}^{T-1} \mu_{j\setminus i}^{r} \nu_{ij}^{r} \\ V_{\partial i}(s) &= V_{\partial i \setminus j}(s) + K_{j\setminus i}^{B}(s) \end{aligned} \tag{D.27}$$

with

$$K_{j\setminus i}^{B}(s) = \sum_{r=s}^{T-1} \mu_{j\setminus i}^{r} \nu_{ij}^{r} \tag{D.28}$$

$L_{\partial i}$ poses a problem now, as $L_{\partial i}(t) = 1 - \phi(t+1)\, e^{M_{\partial i}(t)}$. This can be solved by defining a $\tilde{H}_i$

$$\tilde{H}_i(t) = \log \pi_i^0(t) + G_{\partial i}(t-2) + V_{\partial i}(t) \tag{D.29}$$

redefining the Hs as:

$$H_{\partial i}(t) = \tilde{H}_i(t) + \log L_{\partial i}(t-1) \tag{D.30}$$

and

$$H_{\partial i\setminus j}(t) = \tilde{H}_i(t) - K_{j\setminus i}^{F}(t-2) - K_{j\setminus i}^{B}(t) + \log L_{\partial i\setminus j}(t-1) \tag{D.31}$$

which helps to reduce the computation required.

D.2.3 Summary

$$Z_{i\setminus j} = \sum_{t_i=0}^{T+1} e^{H_{\partial i\setminus j}(t_i)} \tag{D.32}$$

$$m_{i\setminus j}^{t} = \frac{1}{Z_{i\setminus j}} \sum_{t_i=0}^{t} e^{H_{\partial i\setminus j}(t_i)} \tag{D.33}$$

$$\mu_{i\setminus j}^{t} = \frac{\mathbb{I}\left[0 \le t \le T-1\right]}{Z_{i\setminus j}} \left\{ \sum_{t_i=t+2}^{T+1} e^{H_{\partial i\setminus j}(t_i)} - \pi_i^0(t+1)\, e^{G_{\partial i\setminus j}(t) + V_{\partial i\setminus j}(t+1)} \right\} \tag{D.34}$$

$$Z_i = \sum_{t_i=0}^{T+1} \exp\left\{H_{\partial i}(t_i)\right\} = \sum_{t_i=0}^{T+1} \exp\left\{\tilde{H}_i(t_i)\right\} \times L_{\partial i}(t_i - 1) \tag{D.35}$$

$$m_i(t) = \frac{1}{Z_i} \sum_{t_i=0}^{t} e^{\tilde{H}_i(t_i)} L_{\partial i}(t_i - 1) \tag{D.36}$$

D.3 Hard observations

Assume now observations are made on an epidemic cascade, and with each observation the state of an individual is observed with the possibility of making errors. Let' say that each node is observed in two subsequent times, first as susceptible at time τ_i^0 and then infected at time τ_i^1, such that $x_i^{\tau_i^s} = s$ with $s \in \{0,1\}$ and $\tau_i^1 > \tau_i^0$. If node i is observed only once in one of the two states or never observed, we can nominally set $\tau_i^0 = -1$ and/or $\tau_i^1 = T + 1$ if i is never observed susceptible (resp. infected). By construction $\tau_i^1 > \tau_i^0$. The general expression of the above formulas should be:

$$Z_{i\backslash j}\left(\tau_i^0, \tau_i^1\right) = \sum_{t_i=\tau_i^0+1}^{\tau_i^1} e^{\tilde{H}_i(t_i)-K^F_{j\backslash i}(t_i-2)-K^B_{j\backslash i}(t_i)} L_{\partial i\backslash j}\left(t_i - 1\right) \tag{D.37}$$

$$m^t_{i\backslash j}\left(\tau_i^0, \tau_i^1\right) = \frac{\theta\left[t - \left(\tau_i^0 + 1\right)\right]}{Z_{i\backslash j}\left(\tau_i^0, \tau_i^1\right)} \sum_{t_i=\tau_i^0+1}^{\min\left(t,\tau_i^1\right)} e^{\tilde{H}_i(t_i)-K^F_{j\backslash i}(t_i-2)-K^B_{j\backslash i}(t_i)} L_{\partial i\backslash j}\left(t_i - 1\right) \tag{D.38}$$

$$\begin{aligned}\mu^t_{i\backslash j}\left(\tau_i^0, \tau_i^1\right) = {} & \frac{1}{Z_{i\backslash j}\left(\tau_i^0, \tau_i^1\right)} \sum_{t_i=\max\left(t+2,\tau_i^0+1\right)}^{\tau_i^1} e^{\tilde{H}_i(t_i)-K^F_{j\backslash i}(t_i-2)-K^B_{j\backslash i}(t_i)} L_{\partial i\backslash j}\left(t_i - 1\right) + \\ & - \frac{1}{Z_{i\backslash j}\left(\tau_i^0, \tau_i^1\right)} \mathbb{I}\left[\max\left(\tau_i^0, 0\right) \leq t \leq \min\left(\tau_i^1, T\right) - 1\right] \pi_i^0\left(t + 1\right) e^{G_{\partial i}(t)-K^F_{j\backslash i}(t)+V_{\partial i}(t+1)-K^B_{j\backslash i}(t+1)}\end{aligned} \tag{D.39}$$

$$Z_i^O = \sum_{t_i=\tau_i^0+1}^{\tau_i^1} \exp\left\{\tilde{H}_i\left(t_i\right)\right\} \times L_{\partial i}\left(t_i - 1\right) \tag{D.40}$$

$$m_i\left(t \mid \tau_i^0, \tau_i^1\right) = \frac{1}{Z_i^O} \sum_{t_i=\tau_i^0+1}^{\min\left(t,\tau_i^1\right)} e^{\tilde{H}_i(t_i)} L_{\partial i}\left(t_i - 1\right) \tag{D.41}$$

$$\sum_{s=t+2}^{T+1} e^{H(s)} = \sum_{r=t}^{T-1} e^{H(r+2)}$$

D.3.1 Rewrite with the compact formalism

Compute

$$M_{\partial i}\left(r\right) = \sum_{k\in\partial i} m^r_{k\backslash i} v^r_{ki}$$

$$N_{\partial i}\left(r\right) = \sum_{k\in\partial i} \mu^r_{k\backslash i} v^r_{ik}$$

Then

$$G_{\partial i}\left(t\right) = \sum_{r=0}^{t} M_{\partial i}\left(r\right)$$

$$V_{\partial i}\left(s\right)=\sum_{r=s}^{T-1}N_{\partial i}\left(r\right)$$

Now we can calculate

$$\tilde{H}_i\left(t\right)=\log\pi_i^0\left(t\right)+G_{\partial i}\left(t-2\right)+V_{\partial i}\left(t\right)$$

Actually, since t ranges from 0 to $T+1$, and $G_{\partial i}$ and $V_{\partial i}$ range from 0 to $T-1$, we have to put:

$$\tilde{H}_i\left(t\right)=\log\pi_i^0\left(t\right)+\theta(t+2)\,G_{\partial i}\left(t-2\right)+\theta(T-1-t)\,V_{\partial i}\left(t\right)\tag{D.42}$$

These θ functions actually corresponds to the correct limits, as when $t_i<2$ the first term (corresponding to $G_{\partial i}$) in the product of equation D.8 (or equation D.9) is absent, and the same is true for $t_i>T-1$ for the last product in the equation, representing $V_{\partial i}$.

Considering the links, ij,

$$K^F_{j\setminus i}\left(t\right)=\sum_{r=0}^{t}m^r_{j\setminus i}v^r_{ji}$$

$$K^B_{j\setminus i}\left(t\right)=\sum_{r=t}^{T+1}\mu^r_{j\setminus i}v^r_{ij}$$

The same goes for K^F and K^B, so we now compute

$$H_{\partial i\setminus j}\left(t\right)=\tilde{H}_i\left(t\right)-\theta(t+2)\,K^F_{j\setminus i}\left(t-2\right)-\theta(T-1-t)\,K^B_{j\setminus i}\left(t\right)+\theta(t-1)\log\left[1-e^{M_{\partial i}(t-1)-m^t_{j\setminus i}v^t_{ji}}\right]\tag{D.43}$$

for each t in the sum in order to compute both

$$Z_{i\setminus j}\left(\tau_i^0,\tau_i^1\right)=\sum_{t_i=\tau_i^0+1}^{\tau_i^1}\exp\left\{H_{\partial i\setminus j}\left(t_i\right)\right\}\tag{D.44}$$

and

$$m^t_{i\setminus j}\left(\tau_i^0,\tau_i^1\right)=\frac{\theta\left[t-\left(\tau_i^0+1\right)\right]}{Z_{i\setminus j}\left(\tau_i^0,\tau_i^1\right)}\sum_{t_i=\tau_i^0+1}^{\min\left(t,\tau_i^1\right)}\exp\left\{H_{\partial i\setminus j}\left(t_i\right)\right\}\tag{D.45}$$

So for the equation of μ (eq D.39):

$$\tilde{H}_i\left(t_i\right)-K^F_{j\setminus i}\left(t_i-2\right)-K^B_{j\setminus i}\left(t_i\right)=\tilde{H}_i\left(t_i\right)-\theta(t_i+2)\,K^F_{j\setminus i}\left(t_i-2\right)-\theta(T-1-t_i)\,K^B_{j\setminus i}\left(t_i\right)$$

and also in the second part

$$V_{\partial i}\left(t+1\right)-K^B_{j\setminus i}\left(t+1\right)=\theta(T-2-t)\left[V_{\partial i}\left(t+1\right)-K^B_{j\setminus i}\left(t+1\right)\right]$$

So that the mu update is:

$$\begin{aligned}\mu^t_{i\setminus j}\left(\tau_i^0,\tau_i^1\right)=&\frac{1}{Z_{i\setminus j}\left(\tau_i^0,\tau_i^1\right)}\sum_{t_i=\max\left(t+2,\tau_i^0+1\right)}^{\tau_i^1}\exp\left\{H_{\partial i\setminus j}\left(t_i\right)\right\}+\\&-\frac{\mathbb{I}\left[\max\left(\tau_i^0,0\right)\leq t\leq\min\left(\tau_i^1,T\right)-1\right]}{Z_{i\setminus j}\left(\tau_i^0,\tau_i^1\right)}\pi_i^0\left(t+1\right)e^{G_{\partial i}(t)-K^F_{j\setminus i}(t)+\theta(T-2-t)\left[V_{\partial i}(t+1)-K^B_{j\setminus i}(t+1)\right]}\end{aligned}\tag{D.46}$$

This can be written with an indices shift into:

$$\mu^t_{i\setminus j}\left(\tau^0_i,\tau^1_i\right)=\frac{1}{Z_{i\setminus j}\left(\tau^0_i,\tau^1_i\right)}\sum_{s_i=\max(t,\tau^0_i-1)}^{\tau^1_i-2}\exp\left\{H_{\partial i\setminus j}\left(s_i+2\right)\right\}+ \\ -\frac{\mathbb{I}\left[\max\left(\tau^0_i,0\right)\leq t\leq\min\left(\tau^1_i,T\right)-1\right]}{Z_{i\setminus j}\left(\tau^0_i,\tau^1_i\right)}\pi^0_i\left(t+1\right)e^{G_{\partial i}(t)-K^F_{j\setminus i}(t)+\theta(T-2-t)\left[V_{\partial i}(t+1)-K^B_{j\setminus i}(t+1)\right]} \tag{D.47}$$

D.3.2 Numerical stability

The problem sometimes is that $e^{H_{\partial i\setminus j}(t_i)}$ gets to infinity. This can be solved by removing a constant factor C in the computation of each message.

$$Z_{i\setminus j}=\sum_{t_i=\tau^0_i+1}^{\tau^1_i}e^{H_{\partial i\setminus j}(t_i)}=\sum_{t_i=\tau^0_i+1}^{\tau^1_i}e^{H^c_{\partial i\setminus j}(t_i)+C}=e^C\sum_{t_i=\tau^0_i+1}^{\tau^1_i}e^{H^c_{\partial i\setminus j}(t_i)} \tag{D.48}$$

$$m^t_{i\setminus j}=\frac{1}{Z_{i\setminus j}}\sum_{t_i=t}^{z}e^{H_{\partial i\setminus j}(t_i)}=\frac{e^C}{e^C Z^c_{i\setminus j}}\sum_{t_i=t}^{z}e^{H^c_{\partial i\setminus j}(t_i)}=\frac{1}{Z^c_{i\setminus j}}\sum_{t_i=t}^{z}e^{H^c_{\partial i\setminus j}(t_i)} \tag{D.49}$$

Where Z^c and H^c are the "reduced" versions of Z and H. The time constrained version of the μ then becomes:

$$\mu^t_{i\setminus j}\left(\tau^0_i,\tau^1_i\right)=\frac{1}{Z^c_{i\setminus j}\left(\tau^0_i,\tau^1_i\right)}\sum_{s_i=\max(t,\tau^0_i-1)}^{\tau^1_i-2}\exp\left\{H^c_{\partial i\setminus j}\left(s_i+2\right)\right\}+ \\ -\frac{\mathbb{I}\left[\max\left(\tau^0_i,0\right)\leq t\leq\min\left(\tau^1_i,T\right)-1\right]}{Z^c_{i\setminus j}\left(\tau^0_i,\tau^1_i\right)e^C}\pi^0_i\left(t+1\right)e^{G_{\partial i}(t)-K^F_{j\setminus i}(t)+\theta(T-2-t)\left[V_{\partial i}(t+1)-K^B_{j\setminus i}(t+1)\right]} \tag{D.50}$$

It's easy to see that for the $\mu_{i\setminus j}$, the factor e^C can be simplified only in the first part of the sum, but not in the second. The value of C can be defined as the maximum value of $\tilde{H}_i$ over all times t_i that have to be considered.

D.4 Soft observations

Let us now say that with each observation that is made on the cascade, the probability of observing an individual i in state s is $P(o=s\mid t_i)$. Then, it's now necessary to perform the sums considering the whole time range of infection.

Let us define $\mathscr{L}(t)$ as the log-likelihood of infection time t, as the logarithm of the probability of a set of observations O_i that have been made on individual i:

$$\mathscr{L}_i\left(t_i\right)=\log p\left(O_i\,|\,t_i\right)=\sum_{\mathrm{obs}_i}\log p\left(x^{t_o}_i=o\mid t_i\right)$$

which, as written above, can be expressed as the sum of the logarithm of the probability, for each observation, of finding an individual in state $x_i^{t_o}$ given that he/she has been infected at time t_i. This is equivalently the probability of making a mistake in an observations, as in the noiseless case $x_i^{t_o} = 1 \iff t_o \geq t_i$. Then, remembering the definition of $\tilde{H}_i$,

$$\tilde{H}_i(t) = \log \pi_i^0(t) + \theta(t+2)\, G_{\partial i}(t-2) + \theta(T-1-t)\, V_{\partial i}(t) \tag{D.51}$$

notice that, since the term $p(O_i \,|\, t_i)$ is missing from eq. (D.8) and (D.9), $\mathscr{L}_i$ needs to be added to those equation:

$$Z_{i\setminus j} = \sum_{t_i=0}^{T+1} e^{\mathscr{L}_i(t_i)} e^{H_{\partial i\setminus j}(t_i)} \tag{D.52}$$

$$m_{i\setminus j}^{t} = \frac{1}{Z_{i\setminus j}} \sum_{t_i=0}^{t} e^{\mathscr{L}_i(t_i)} e^{H_{\partial i\setminus j}(t_i)} \tag{D.53}$$

$$\mu_{i\setminus j}^{t} = \frac{\mathbb{I}\left[0 \leq t \leq T-1\right]}{Z_{i\setminus j}} \left\{ \sum_{s=t}^{T-1} e^{\mathscr{L}_i(s+2)} e^{H_{\partial i\setminus j}(s+2)} - \pi_i^0(t+1)\, e^{\mathscr{L}_i(t+1)} e^{G_{\partial i\setminus j}(t) + V_{\partial i\setminus j}(t+1)} \right\} \tag{D.54}$$

$$Z_i = \sum_{t_i=0}^{T+1} e^{\mathscr{L}_i(t_i)} \exp\left\{H_{\partial i}(t_i)\right\} = \sum_{t_i=0}^{T+1} e^{\mathscr{L}_i(t_i)} \exp\left\{\tilde{H}_i(t_i)\right\} \times L_{\partial i}(t_i - 1) \tag{D.55}$$

$$m_i(t) = \frac{1}{Z_i} \sum_{t_i=0}^{t} e^{\mathscr{L}_i(t_i)} e^{\tilde{H}_i(t_i)} L_{\partial i}(t_i - 1) \tag{D.56}$$

Since the likelihood $\mathscr{L}_i$ needs to be included in both $H_{\partial i}$ and $H_{\partial i\setminus j}$, it can be incorporated right into the definition of $\tilde{H}_i(t)$:

$$\tilde{H}_i(t) = \log \pi_i^0(t) + G_{\partial i}(t-2) + V_{\partial i}(t) + \mathscr{L}_i(t) \tag{D.57}$$

In these way, the computation is streamlined, and a lot of boundaries calculations can be simplified. In practice this case can be used also for noiseless observations, by putting a very small probability of error, so that $p\left(x_i^{t_o} = 1 \mid t_i\right) = \varepsilon \ll 1$ when $t_o < t_i$. Also, the same normalization step can be applied here for numerical stability, thus improving the convergence.

Bibliography

[1] Daniel Bernoulli. "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir." In: *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem* (1760), pp. 1–45.

[2] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. "A contribution to the mathematical theory of epidemics." In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (Aug. 1, 1927), pp. 700–721. DOI: `10.1098/rspa.1927.0118`. URL: `https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118`.

[3] R. Gallager. "Low-density parity-check codes." In: *IRE Transactions on Information Theory* 8.1 (Jan. 1962), pp. 21–28. ISSN: 2168-2712. DOI: `10.1109/TIT.1962.1057683`.

[4] Daniel T. Gillespie. "Exact stochastic simulation of coupled chemical reactions." In: *The Journal of Physical Chemistry* 81.25 (Dec. 1, 1977). Publisher: American Chemical Society, pp. 2340–2361. ISSN: 0022-3654. DOI: `10.1021/j100540a008`. URL: `https://doi.org/10.1021/j100540a008`.

[5] Judea Pearl. "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach." In: 1982, pp. 133–136. ISBN: 0865760438. URL: `www.aaai.org`.

[6] T Plefka. "Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model." In: *Journal of Physics A: Mathematical and General* 15.6 (June 1982), p. 1971. DOI: `10.1088/0305-4470/15/6/035`. URL: `https://dx.doi.org/10.1088/0305-4470/15/6/035`.

[7] John S. Quarterman and Josiah C. Hoskins. "Notable computer networks." In: *Communications of the ACM* 29.10 (Oct. 1, 1986), pp. 932–971. ISSN: 0001-0782. DOI: `10.1145/6617.6618`. URL: `https://doi.org/10.1145/6617.6618`.

[8] Steffen L Lauritzen. *Graphical models.* Vol. 17. Clarendon Press, 1996.

[9] H. A. Bethe and William Lawrence Bragg. "Statistical theory of superlattices." In: *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences* 150.871 (Jan. 1997), pp. 552–575. DOI: `10.1098/rspa.1935.0122`. URL: `https://royalsocietypublishing.org/doi/10.1098/rspa.1935.0122`.

[10] Donald Knuth. *The art of computer programming*. Reading, Mass: Addison-Wesley, 1997. ISBN: 0-201-89683-4.

[11] Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." In: *Nature* 393.6684 (June 1998), pp. 440–442. ISSN: 1476-4687. DOI: `10.1038/30918`. URL: `https://www.nature.com/articles/30918.`.

[12] Steffen L Lauritzen. “Causal inference from graphical models.” In: *Monographs on Statistics and Applied Probability* 87 (2001), pp. 63–108.

[13] Yoshua Bengio et al. “A neural probabilistic language model.” In: *The journal of machine learning research* 3 (2003), pp. 1137–1155.

[14] K. T. D. Eames and M. J. Keeling. “Contact tracing and disease control.” In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.1533 (Dec. 22, 2003). Publisher: Royal Society, pp. 2565–2571. DOI: `10.1098/rspb.2003.2554`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2003.2554` (visited on 01/30/2023).

[15] Steven Riley et al. “Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions.” In: *Science* 300.5627 (June 20, 2003), pp. 1961–1966. DOI: `10.1126/science.1086478`. URL: `https://www.science.org/doi/abs/10.1126/science.1086478`.

[16] Thomas Schank and Dorothea Wagner. *Approximating clustering-coefficient and transitivity.* Vol. 2004. Interner Bericht. Fakultät für Informatik, Universität Karlsruhe 9. Universität Karlsruhe (TH), 2004.

[17] Feng Xie and David Levinson. “Measuring the Structure of Road Networks.” In: *Geographical Analysis* 39.3 (2007), pp. 336–356. ISSN: 1538-4632. DOI: `10.1111/j.1538-4632.2007.00707.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.2007.00707.x`.

[18] Fred Brauer, PV den Driessche, and Jianhong Wu. “Lecture notes in mathematical epidemiology.” In: *Berlin, Germany. Springer* 75.1 (2008), pp. 3–22.

[19] Guy Karlebach and Ron Shamir. “Modelling and analysis of gene regulatory networks.” In: *Nature Reviews Molecular Cell Biology* 9.10 (Oct. 2008), pp. 770–780. ISSN: 1471-0080. DOI: `10.1038/nrm2503`. URL: `https://www.nature.com/articles/nrm2503`.

[20] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. “Virus spread in networks.” In: *IEEE/ACM Transactions On Networking* 17.1 (2008), pp. 1–14.

[21] Martin J. Wainwright and Michael I. Jordan. “Graphical Models, Exponential Families, and Variational Inference.” In: *Foundations and Trends® in Machine Learning* 1.1 (Nov. 17, 2008). Publisher: Now Publishers, Inc., pp. 1–305. ISSN: 1935-8237, 1935-8245. DOI: `10.1561/2200000001`. URL: `https://www.nowpublishers.com/article/Details/MAL-001`.

[22] Duygu Balcan et al. “Multiscale mobility networks and the spatial spreading of infectious diseases.” In: *Proceedings of the National Academy of Sciences* 106.51 (Dec. 22, 2009). Publisher: Proceedings of the National Academy of Sciences, pp. 21484–21489. DOI: `10.1073/pnas.0906910106`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.0906910106`.

[23] Marc Mezard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[24] I. Neri and D. Bollé. “The cavity approach to parallel dynamics of Ising spins on a graph.” In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.8 (Aug. 2009), P08009. ISSN: 1742-5468. DOI: `10.1088/1742-5468/2009/08/P08009`. URL: `https://dx.doi.org/10.1088/1742-5468/2009/08/P08009`.

[25] Shawn Vasoo, Jane Stevens, and Kamaljit Singh. "Rapid Antigen Tests for Diagnosis of Pandemic (Swine) Influenza A/H1N1." In: *Clinical Infectious Diseases* 49.7 (Oct. 1, 2009), pp. 1090–1093. ISSN: 1058-4838. DOI: 10.1086/644743. URL: https://doi.org/10.1086/644743.

[26] Ciro Cattuto et al. "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks." In: *PLOS ONE* 5.7 (July 15, 2010), e11596. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0011596. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011596.

[27] Yashodhan Kanoria and Andrea Montanari. "Majority dynamics on trees and the dynamic cavity method." In: *The Annals of Applied Probability* 21.5 (Oct. 2011), pp. 1694–1748. ISSN: 1050-5164, 2168-8737. DOI: 10.1214/10-AAP729. URL: https://projecteuclid.org/journals/annals-of-applied-probability/volume-21/issue-5/Majority-dynamics-on-trees-and-the-dynamic-cavity-method/10.1214/10-AAP729.full.

[28] Hugo Larochelle and Iain Murray. "The neural autoregressive distribution estimator." In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings. 2011, pp. 29–37.

[29] Richard A. Stein. "Super-spreaders in infectious diseases." In: *International Journal of Infectious Diseases* 15.8 (Aug. 1, 2011), e510–e513. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2010.06.020. URL: https://www.sciencedirect.com/science/article/pii/S1201971211000245.

[30] Petter Holme and Jari Saramäki. "Temporal networks." In: *Physics Reports*. Temporal Networks 519.3 (Oct. 1, 2012), pp. 97–125. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157312000841 (visited on 12/29/2022).

[31] Pedro Larrañaga et al. "A review on probabilistic graphical models in evolutionary computation." In: *Journal of Heuristics* 18.5 (Oct. 1, 2012), pp. 795–819. ISSN: 1572-9397. DOI: 10.1007/s10732-012-9208-4. URL: https://doi.org/10.1007/s10732-012-9208-4.

[32] Federico Ricci-Tersenghi. "The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods." In: *Journal of Statistical Mechanics: Theory and Experiment* 2012.8 (Aug. 2012). Publisher: IOP Publishing and SISSA, P08015. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2012/08/P08015. URL: https://dx.doi.org/10.1088/1742-5468/2012/08/P08015.

[33] Michele Tizzoni et al. "Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm." In: *BMC Medicine* 10.1 (Dec. 13, 2012), p. 165. ISSN: 1741-7015. DOI: 10.1186/1741-7015-10-165. URL: https://doi.org/10.1186/1741-7015-10-165.

[34] Laijun Zhao et al. "SIHR rumor spreading model in social networks." In: *Physica A: Statistical Mechanics and its Applications* 391.7 (Apr. 1, 2012), pp. 2444–2453. ISSN: 0378-4371. DOI: 10.1016/j.physa.2011.12.008. URL: https://www.sciencedirect.com/science/article/pii/S0378437111009058.

[35] F. Altarelli et al. "Large deviations of cascade processes on graphs." In: *Physical Review E* 87.6 (June 11, 2013), p. 062115. DOI: 10.1103/PhysRevE.87.062115. URL: https://link.aps.org/doi/10.1103/PhysRevE.87.062115.

[36] Sheng-Jun Xu et al. "Image segmentation using adaptive loopy belief propagation." In: *Optik* 124.22 (Nov. 1, 2013), pp. 5732–5738. ISSN: 0030-4026. DOI: `10.1016/j.ijleo.2013.04.013`. URL: `https://www.sciencedirect.com/science/article/pii/S0030402613004981`.

[37] Fabrizio Altarelli et al. "Bayesian Inference of Epidemics on Networks via Belief Propagation." In: *Physical Review Letters* 112.11 (Mar. 17, 2014), p. 118701. DOI: `10.1103/PhysRevLett.112.118701`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.112.118701` (visited on 12/29/2022).

[38] Fabrizio Altarelli et al. "The patient-zero problem with noisy observations." In: *Journal of Statistical Mechanics: Theory and Experiment* 2014.10 (Oct. 2014), P10016. ISSN: 1742-5468. DOI: `10.1088/1742-5468/2014/10/P10016`. URL: `https://dx.doi.org/10.1088/1742-5468/2014/10/P10016`.

[39] Andrey Y. Lokhov et al. "Inferring the origin of an epidemic with a dynamic message-passing algorithm." In: *Physical Review E* 90.1 (July 1, 2014), p. 012801. DOI: `10.1103/PhysRevE.90.012801`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.90.012801`.

[40] Arkadiusz Stopczynski et al. "Measuring Large-Scale Social Networks with High Resolution." In: *PLOS ONE* 9.4 (Apr. 25, 2014), e95978. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0095978`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095978`.

[41] Nino Antulov-Fantulin et al. "Identification of Patient Zero in Static and Temporal Networks: Robustness and Limitations." In: *Physical Review Letters* 114.24 (June 2015). Publisher: American Physical Society, p. 248701. DOI: `10.1103/PhysRevLett.114.248701`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.114.248701` (visited on 02/02/2021).

[42] Jeffrey W. Eaton et al. "Assessment of epidemic projections using recent HIV survey data in South Africa: a validation analysis of ten mathematical models of HIV epidemiology in the antiretroviral therapy era." In: *The Lancet Global Health* 3.10 (Oct. 1, 2015), e598–e608. ISSN: 2214-109X. DOI: `10.1016/S2214-109X(15)00080-7`. URL: `https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(15)00080-7/fulltext`.

[43] Mathieu Germain et al. "Made: Masked autoencoder for distribution estimation." In: *International Conference on Machine Learning*. PMLR. 2015, pp. 881–889.

[44] Andrey Y. Lokhov, Marc Mézard, and Lenka Zdeborová. "Dynamic message-passing equations for models with unidirectional dynamics." In: *Physical Review E* 91.1 (Jan. 13, 2015), p. 012811. DOI: `10.1103/PhysRevE.91.012811`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.91.012811`.

[45] Thomas Obadia et al. "Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals." In: *PLoS Computational Biology* 11 (3 Mar. 2015). Ed. by Marcel Salathé, e1004170. ISSN: 15537358. DOI: `10.1371/journal.pcbi.1004170`. URL: `https://dx.plos.org/10.1371/journal.pcbi.1004170`.

[46] Munik Shrestha, Samuel V. Scarpino, and Cristopher Moore. "Message-passing approach for recurrent-state epidemic models on networks." In: *Physical Review E* 92.2 (Aug. 27, 2015). Publisher: American Physical Society, p. 022821. DOI: 10.1103/PhysRevE.92.022821. URL: https://link.aps.org/doi/10.1103/PhysRevE.92.022821.

[47] Alfredo Braunstein and Alessandro Ingrosso. "Inference of causality in epidemics on temporal contact networks." In: *Scientific Reports* 6.1 (June 10, 2016), p. 27538. ISSN: 2045-2322. DOI: 10.1038/srep27538. (Visited on 11/18/2020).

[48] B Bravi, P Sollich, and M Opper. "Extended Plefka expansion for stochastic dynamics." In: *Journal of Physics A: Mathematical and Theoretical* 49.19 (Apr. 2016), p. 194003. DOI: 10.1088/1751-8113/49/19/194003. URL: https://dx.doi.org/10.1088/1751-8113/49/19/194003.

[49] Zhongxiang Dai et al. "Temporal efficiency evaluation and small-worldness characterization in temporal networks." In: *Scientific Reports* 6.1 (Sept. 29, 2016), p. 34291. ISSN: 2045-2322. DOI: 10.1038/srep34291. URL: https://www.nature.com/articles/srep34291.

[50] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[51] Aäron van den Oord et al. "WaveNet: A Generative Model for Raw Audio." In: *CoRR* abs/1609.03499 (2016). arXiv: 1609.03499. URL: http://arxiv.org/abs/1609.03499.

[52] Aäron van den Oord et al. "Conditional Image Generation with PixelCNN Decoders." In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16. Barcelona, Spain: Curran Associates Inc., 2016, pp. 4797–4805. ISBN: 9781510838819.

[53] Mathew D. Penrose. "Connectivity of soft random geometric graphs." In: *The Annals of Applied Probability* 26.2 (Apr. 2016), pp. 986–1028. ISSN: 1050-5164, 2168-8737. DOI: 10.1214/15-AAP1110. URL: https://projecteuclid.org/journals/annals-of-applied-probability/volume-26/issue-2/Connectivity-of-soft-random-geometric-graphs/10.1214/15-AAP1110.full.

[54] Benigno Uria et al. "Neural autoregressive distribution estimation." In: *The Journal of Machine Learning Research* 17.1 (2016), pp. 7184–7220.

[55] Aaron Van Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks." In: *International Conference on Machine Learning*. PMLR. 2016, pp. 1747–1756.

[56] T. Berge et al. "A simple mathematical model for Ebola in Africa." In: *Journal of Biological Dynamics* 11.1 (Jan. 1, 2017), pp. 42–74. ISSN: 1751-3758. DOI: 10.1080/17513758.2016.1229817. URL: https://doi.org/10.1080/17513758.2016.1229817.

[57] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization." In: *arXiv:1412.6980 [cs]* (Jan. 2017). arXiv: 1412.6980. URL: http://arxiv.org/abs/1412.6980 (visited on 10/19/2021).

[58] Ross C. Phillips and Denise Gorse. "Predicting cryptocurrency price bubbles using social media data and epidemic modelling." In: *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2017 IEEE Symposium Series on Computational Intelligence (SSCI). Nov. 2017, pp. 1–7. DOI: 10.1109/SSCI.2017.8280809.

[59] Muneki Yasuda and Shun Kataoka. "Solving Non-parametric Inverse Problem in Continuous Markov Random Field Using Loopy Belief Propagation." In: *Journal of the Physical Society of Japan* 86.8 (Aug. 15, 2017). Publisher: The Physical Society of Japan, p. 084806. ISSN: 0031-9015. DOI: `10.7566/JPSJ.86.084806`. URL: `https://journals.jps.jp/doi/10.7566/JPSJ.86.084806`.

[60] Florian Dörfler, John W. Simpson-Porco, and Francesco Bullo. "Electrical Networks and Algebraic Graph Theory: Models, Properties, and Applications." In: *Proceedings of the IEEE* 106.5 (May 2018), pp. 977–1005. ISSN: 1558-2256. DOI: `10.1109/JPROC.2018.2821924`.

[61] Mathieu Génois and Alain Barrat. "Can co-location be used as a proxy for face-to-face contacts?" In: *EPJ Data Science* 7.1 (May 2018), p. 11. ISSN: 2193-1127. DOI: `10.1140/epjds/s13688-018-0140-1`. URL: `https://doi.org/10.1140/epjds/s13688-018-0140-1`.

[62] Mark Newman. *Networks / Mark Newman.* eng. 2nd ed. Oxford: Oxford University Press, 2018. ISBN: 9780198805090.

[63] Alfredo Braunstein, Alessandro Ingrosso, and Anna Paola Muntoni. "Network reconstruction from infection cascades." In: *Journal of The Royal Society Interface* 16.151 (Feb. 2019), p. 20180844. DOI: `10.1098/rsif.2018.0844`. URL: `royalsocietypublishing.org/doi/full/10.1098/rsif.2018.0844`.

[64] Nitish Shirish Keskar et al. "Ctrl: A conditional transformer language model for controllable generation." In: *arXiv preprint arXiv:1909.05858* (2019).

[65] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In: *Advances in Neural Information Processing Systems 32.* Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[66] Manliura Datilo Philemon, Zuhaimy Ismail, and Jayeola Dare. "A review of epidemic forecasting using artificial neural networks." In: *International Journal of Epidemiologic Research* 6.3 (2019), pp. 132–143.

[67] Piotr Sapiezynski et al. "Interaction data from the Copenhagen Networks Study." In: *Scientific Data* 6.1 (Dec. 11, 2019), p. 315. ISSN: 2052-4463. DOI: `10.1038/s41597-019-0325-x`. URL: `https://www.nature.com/articles/s41597-019-0325-x`.

[68] Dian Wu, Lei Wang, and Pan Zhang. "Solving Statistical Mechanics Using Variational Autoregressive Networks." In: *Phys. Rev. Lett.* 122 (8 Feb. 2019), p. 080602. DOI: `10.1103/PhysRevLett.122.080602`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.122.080602`.

[69] *Aarogya setu app.* `https://www.aarogyasetu.gov.in/`. 2020.

[70] Alberto Aleta et al. "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19." In: *Nature Human Behaviour* 4.9 (Aug. 5, 2020), pp. 964–971. ISSN: 2397-3374. DOI: `10.1038/s41562-020-0931-9`. URL: `https://www.nature.com/articles/s41562-020-0931-9/`.

[71] Hannah Alsdurf et al. *COVI White Paper.* July 27, 2020. DOI: `10.48550/arXiv.2005.08502`. arXiv: `2005.08502[cs]`. URL: `http://arxiv.org/abs/2005.08502`.

[72] Benjamin M. Althouse et al. "Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control." In: *PLOS Biology* 18.11 (Nov. 12, 2020), e3000897. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.3000897`. URL: `https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000897`.

[73] Apple and Google. "Privacy-preserving contact tracing." https://covid19.apple.com/contacttracing. 2020.

[74] Jason Bay et al. *BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders.* Government Technology Agency-Singapore, Tech. Rep. 2020. URL: `%5Curl%7Bbluetrace.io/static/bluetrace%5C_whitepaper-938063656596c104632def383eb33b3c.pdf%7D`.

[75] Ryan Bergstrom and Ashkan Soltani Calo {and} Carl. *Contact-tracing apps are not a solution to the COVID-19 crisis.* Brookings. Apr. 27, 2020. URL: `https://www.brookings.edu/techstream/inaccurate-and-insecure-why-contact-tracing-apps-could-be-a-disaster/`.

[76] Alfredo Braunstein, Luca Dall'Asta, and Alessandro Ingrosso. "Casualità, causalità e Machine Learning nel contenimento epidemico." In: *Ithaca* (2020). URL: `https://hdl.handle.net/11583/2868632`.

[77] Justin Chan et al. "Pact: Privacy sensitive protocols and mechanisms for mobile contact tracing." arXiv:2004.03544. 2020.

[78] Vinay Kumar Reddy Chimmula and Lei Zhang. "Time series forecasting of COVID-19 transmission in Canada using LSTM networks." In: *Chaos, Solitons & Fractals* 135 (2020), p. 109864. ISSN: 0960-0779. DOI: `https://doi.org/10.1016/j.chaos.2020.109864`. URL: `https://www.sciencedirect.com/science/article/pii/S0960077920302642`.

[79] Hyunghoon Cho, Daphne Ippolito, and Yun William Yu. "Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs." arXiv:2003.11511. 2020.

[80] "Coronavirus confirmed as pandemic by World Health Organization." In: *BBC News* (Mar. 11, 2020). URL: `https://www.bbc.com/news/world-51839944`.

[81] *Coronavirus spread now a global emergency declares World Health Organization | UN News.* Jan. 30, 2020. URL: `https://news.un.org/en/story/2020/01/1056372`.

[82] *CoronaWarn contact tracing mobile application.* Aug. 2020. URL: `https://www.coronawarn.app/en/`.

[83] *COVIDSafe app.* `https://www.health.gov.au/resources/apps-and-tools/covidsafe-app`. 2020.

[84] Akira Endo et al. "Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China." In: *Wellcome Open Research* 5 (July 10, 2020), p. 67. ISSN: 2398-502X. DOI: `10.12688/wellcomeopenres.15842.3`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7338915/`.

[85] Luca Ferretti et al. "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing." In: *Science* 368.6491 (May 8, 2020), eabb6936. DOI: `10.1126/science.abb6936`. URL: `https://www.science.org/doi/10.1126/science.abb6936`.

[86] Seth Flaxman et al. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe." In: *Nature* 584.7820 (2020), pp. 257–261.

[87] Valeria Gelardi et al. “Measuring social networks in primates: wearable sensors versus direct observations.” In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 476.2236 (Apr. 29, 2020), p. 20190737. DOI: `10.1098/rspa.2019.0737`. URL: `https://royalsocietypublishing.org/doi/10.1098/rspa.2019.0737`.

[88] *GitHub : Epidemic Mitigation Framework.* github.com/sibyl-team/epidemic_mitigation. Sept. 2020.

[89] Kyra H Grantz et al. “The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology.” In: *Nature communications* 11.1 (2020), pp. 1–8.

[90] Prateek Gupta et al. “COVI-AgentSim: an agent-based model for evaluating methods of digital contact tracing.” In: *arXiv preprint arXiv:2010.16004* (Oct. 2020).

[91] Agus Hasan et al. “Superspreading in early transmissions of COVID-19 in Indonesia.” In: *Scientific Reports* 10.1 (Dec. 28, 2020), p. 22386. ISSN: 2045-2322. DOI: `10.1038/s41598-020-79352-5`. URL: `https://www.nature.com/articles/s41598-020-79352-5`.

[92] *Immuni app.* `https://www.immuni.italia.it/`. 2020.

[93] Stephen M. Kissler et al. *Sparking “The BBC Four Pandemic”: Leveraging citizen science and mobile phones to model the spread of disease.* May 12, 2020. DOI: `10.1101/479154`. URL: `https://www.biorxiv.org/content/10.1101/479154v2`.

[94] Adam J Kucharski et al. “Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study.” In: *The Lancet Infectious Diseases* 20.10 (2020), pp. 1151–1160.

[95] Max S. Y. Lau et al. “Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA.” In: *Proceedings of the National Academy of Sciences* 117.36 (Sept. 8, 2020), pp. 22430–22435. DOI: `10.1073/pnas.2011802117`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.2011802117`.

[96] Stephen A. Lauer et al. “The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application.” In: *Annals of Internal Medicine* 172.9 (May 5, 2020), pp. 577–582. ISSN: 1539-3704. DOI: `10.7326/M20-0504`.

[97] Seung Won Lee et al. “Nationwide Results of COVID-19 Contact Tracing in South Korea: Individual Participant Data From an Epidemiological Survey.” In: *JMIR Medical Informatics* 8.8 (Aug. 25, 2020), e20992. DOI: `10.2196/20992`. URL: `https://medinform.jmir.org/2020/8/e20992`.

[98] B. McNaughton et al. “Boosting Monte Carlo simulations of spin glasses using autoregressive neural networks.” In: *Phys. Rev. E* 101 (5 May 2020), p. 053312. DOI: `10.1103/PhysRevE.101.053312`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.101.053312`.

[99] *NHS COVID-19 app.* `https://covid19.nhs.uk/`. 2020.

[100] Hitoshi Oshitani, Labour The Expert Members of The National COVID-19 Cluster Taskforce at The Ministry of Health, and Japan Welfare. “Cluster-Based Approach to Coronavirus Disease 2019 (COVID-19) Response in Japan, from February to April 2020.” In: *Japanese Journal of Infectious Diseases* 73.6 (Nov. 30, 2020), pp. 491–493. DOI: `10.7883/yoken.JJID.2020.363`.

[101] Ratnabali Pal et al. “Neural Network Based Country Wise Risk Prediction of COVID-19.” In: *Applied Sciences* 10.18 (2020). ISSN: 2076-3417. DOI: `10.3390/app10186448`. URL: `https://www.mdpi.com/2076-3417/10/18/6448`.

[102] Jasmina Panovska-Griffiths. *Can mathematical modelling solve the current Covid-19 crisis?* 2020.

[103] Sangchul Park, Gina Jeehyun Choi, and Haksoo Ko. “Information technology–based tracing strategy in response to COVID-19 in South Korea—privacy controversies.” In: *Jama* 323.21 (2020), pp. 2129–2130.

[104] Ramesh Raskar et al. “Apps gone rogue: Maintaining personal privacy in an epidemic.” arXiv:2003.08567. 2020.

[105] Marcel Salathé et al. “Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland.” In: *medRxiv* (2020).

[106] Or Sharir et al. “Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems.” In: *Phys. Rev. Lett.* 124 (2 Jan. 2020), p. 020503. DOI: `10.1103/PhysRevLett.124.020503`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.124.020503`.

[107] Luke Taylor. “Uruguay is winning against covid-19. This is how.” In: *BMJ* 370 (Sept. 18, 2020), p. m3575. ISSN: 1756-1833. DOI: `10.1136/bmj.m3575`. URL: `https://www.bmj.com/content/370/bmj.m3575`.

[108] *TousAntiCovid app*. `https://bonjour.tousanticovid.gouv.fr/`. 2020.

[109] *Tracetogether app*. `https://www.tracetogether.gov.sg/`. 2020.

[110] Thomas A. Treibel et al. “COVID-19: PCR screening of asymptomatic health-care workers at London hospital.” In: *The Lancet* 395.10237 (May 23, 2020), pp. 1608–1610. ISSN: 0140-6736, 1474-547X. DOI: `10.1016/S0140-6736(20)31100-4`. URL: `https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31100-4/fulltext?utm_campaign=tlcoronavirus20&utm_source=twitter&utm_medium=social`.

[111] Carmela Troncoso et al. *Decentralized Privacy-Preserving Proximity Tracing*. May 25, 2020. DOI: `10.48550/arXiv.2005.12273`. arXiv: `2005.12273[cs]`. URL: `http://arxiv.org/abs/2005.12273`.

[112] Zeynep Tufekci. *This Overlooked Variable Is the Key to the Pandemic*. Sept. 2020. URL: `https://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/`.

[113] Michel Walrave, Cato Waeterloos, and Koen Ponnet. “Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach.” In: *JMIR Public Health and Surveillance* 6.3 (Sept. 1, 2020), e20572. DOI: `10.2196/20572`. URL: `https://publichealth.jmir.org/2020/3/e20572`.

[114] Chen Wang et al. “A novel coronavirus outbreak of global health concern.” In: *The Lancet* 395.10223 (Feb. 15, 2020), pp. 470–473. ISSN: 0140-6736, 1474-547X. DOI: `10.1016/S0140-6736(20)30185-9`. URL: `https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30185-9/fulltext#tbl1` (visited on 01/23/2023).

[115] Felix Wong and James J. Collins. “Evidence that coronavirus superspreading is fat-tailed.” In: *Proceedings of the National Academy of Sciences* 117.47 (Nov. 2020), pp. 29416–29418. DOI: `10.1073/pnas.2018490117`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.2018490117`.

[116] Yunjun Zhang et al. “Evaluating Transmission Heterogeneity and Super-Spreading Event of COVID-19 in a Metropolis of China.” In: *International Journal of Environmental Research and Public Health* 17.10 (May 24, 2020), p. 3705. ISSN: 1660-4601. DOI: `10.3390/ijerph17103705`. URL: `https://www.mdpi.com/1660-4601/17/10/3705`.

[117] Peter Ashcroft et al. “Quantifying the impact of quarantine duration on COVID-19 transmission.” In: *Elife* 10 (2021), e63704.

[118] Antoine Baker et al. “Epidemic mitigation by statistical inference from contact tracing data.” In: *Proceedings of the National Academy of Sciences* 118.32 (Aug. 10, 2021). ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.2106548118`. URL: `https://www.pnas.org/content/118/32/e2106548118`.

[119] A. Barrat et al. “Effect of manual and digital contact tracing on COVID-19 outbreaks: a study on empirical contact data.” In: *Journal of The Royal Society Interface* 18.178 (May 5, 2021), p. 20201000. DOI: `10.1098/rsif.2020.1000`. URL: `https://royalsocietypublishing.org/doi/10.1098/rsif.2020.1000`.

[120] Christian Bjørnskov. “Did Lockdown Work? An Economist’s Cross-Country Comparison.” In: *CESifo Economic Studies* 67.3 (Sept. 1, 2021), pp. 318–331. ISSN: 1610-241X. DOI: `10.1093/cesifo/ifab003`. URL: `https://doi.org/10.1093/cesifo/ifab003` (visited on 01/24/2023).

[121] Annelies G. Blom et al. “Barriers to the Large-Scale Adoption of a COVID-19 Contact Tracing App in Germany: Survey Study.” In: *Journal of Medical Internet Research* 23.3 (Mar. 2, 2021), e23362. DOI: `10.2196/23362`. URL: `https://www.jmir.org/2021/3/e23362`.

[122] William J. Bradshaw et al. “Bidirectional contact tracing could dramatically improve COVID-19 control.” In: *Nature Communications* 12.1 (Jan. 11, 2021), p. 232. ISSN: 2041-1723. DOI: `10.1038/s41467-020-20325-7`. URL: `https://www.nature.com/articles/s41467-020-20325-7`.

[123] G. Cencetti et al. “Digital proximity tracing on empirical contact networks for pandemic control.” In: *Nature Communications* 12.1 (Mar. 12, 2021), p. 1655. ISSN: 2041-1723. DOI: `10.1038/s41467-021-21809-w`. URL: `https://www.nature.com/articles/s41467-021-21809-w`.

[124] Serina Chang et al. “Mobility network models of COVID-19 explain inequities and inform reopening.” In: *Nature* 589.7840 (2021), pp. 82–87.

[125] Sebastian Contreras et al. “The challenges of containing SARS-CoV-2 via test-trace-and-isolate.” In: *Nature Communications* 12.1 (Jan. 15, 2021), p. 378. ISSN: 2041-1723. DOI: `10.1038/s41467-020-20699-8`. URL: `https://www.nature.com/articles/s41467-020-20699-8`.

[126] Jacqueline Dinnes et al. “Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2 infection.” en. In: *Cochrane Database of Systematic Reviews* 3 (Mar. 24, 2021). ISSN: 1465-1858. DOI: `10.1002/14651858.CD013705.pub2`. URL: `https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013705.pub2/full`.

[127] Akira Endo et al. “Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks.” In: *Wellcome Open Research* 5 (Mar. 31, 2021), p. 239. ISSN: 2398-502X. DOI: `10.12688/wellcomeopenres.16344.3`.

[128] Norman E. Fenton et al. *A Bayesian network model for personalised COVID19 risk assessment and contact tracing*. Jan. 5, 2021. DOI: `10.1101/2020.07.15.20154286`. URL: `https://www.medrxiv.org/content/10.1101/2020.07.15.20154286v2`.

[129] Alexander Harmon et al. “Validation of an At-Home Direct Antigen Rapid Test for COVID-19.” In: *JAMA Network Open* 4.8 (Aug. 27, 2021), e2126931. ISSN: 2574-3805. DOI: `10.1001/jamanetworkopen.2021.26931`. URL: `https://doi.org/10.1001/jamanetworkopen.2021.26931`.

[130] Mohamed Hibat-Allah et al. “Variational neural annealing.” In: *Nature Machine Intelligence* 3.11 (Nov. 2021), pp. 952–961. ISSN: 2522-5839. DOI: `10.1038/s42256-021-00401-3`. URL: `https://www.nature.com/articles/s42256-021-00401-3`.

[131] Robert Hinch et al. “OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing.” In: *PLOS Computational Biology* 17.7 (July 12, 2021), e1009146. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1009146`. URL: `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009146`.

[132] Lee Kennedy-Shaffer, Michael Baym, and William P Hanage. “Perfect as the enemy of good: tracing transmissions with low-sensitivity tests to mitigate SARS-CoV-2 outbreaks.” In: *The Lancet Microbe* 2.5 (May 1, 2021), e219–e224. ISSN: 2666-5247. DOI: `10.1016/S2666-5247(21)00004-5`. URL: `https://www.sciencedirect.com/science/article/pii/S2666524721000045`.

[133] Cliff C. Kerr et al. “Controlling COVID-19 via test-trace-quarantine.” In: *Nature Communications* 12.1 (May 20, 2021), p. 2993. ISSN: 2041-1723. DOI: `10.1038/s41467-021-23276-9`. URL: `https://www.nature.com/articles/s41467-021-23276-9`.

[134] Cliff C. Kerr et al. “Covasim: An agent-based model of COVID-19 dynamics and interventions.” In: *PLOS Computational Biology* 17.7 (July 26, 2021), e1009149. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1009149`. URL: `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009149`.

[135] Sadamori Kojaku et al. “The effectiveness of backward contact tracing in networks.” In: *Nature Physics* 17.5 (May 2021), pp. 652–658. ISSN: 1745-2481. DOI: `10.1038/s41567-021-01187-2`. URL: `https://www.nature.com/articles/s41567-021-01187-2`.

[136] Jacob E. Lemieux et al. “Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events.” In: *Science* 371.6529 (Feb. 5, 2021), eabe3261. DOI: `10.1126/science.abe3261`. URL: `https://www.science.org/doi/10.1126/science.abe3261`.

[137] Tiantian Liu, Zijian Chen, and Jin Xu. “Epidemiological characteristics and incubation period of SARS-CoV-2 during the 2020–2021 winter pandemic wave in North China: An observational study.” In: *Journal of Medical Virology* 93.12 (2021), pp. 6628–6633. ISSN: 1096-9071. DOI: 10.1002/jmv.27226. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.27226 (visited on 12/30/2022).

[138] Davin Lunz, Gregory Batt, and Jakob Ruess. “To quarantine, or not to quarantine: A theoretical framework for disease control via contact tracing.” In: *Epidemics* 34 (2021), p. 100428.

[139] Hinta Meijerink et al. “The First GAEN-Based COVID-19 Contact Tracing App in Norway Identifies 80% of Close Contacts in “Real Life” Scenarios.” In: *Frontiers in Digital Health* 3 (2021). ISSN: 2673-253X. URL: https://www.frontiersin.org/articles/10.3389/fdgth.2021.731098.

[140] Dominik Menges et al. “A Data-Driven Simulation of the Exposure Notification Cascade for Digital Contact Tracing of SARS-CoV-2 in Zurich, Switzerland.” In: *JAMA Network Open* 4.4 (Apr. 30, 2021), e218184. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2021.8184. URL: https://doi.org/10.1001/jamanetworkopen.2021.8184.

[141] Tim R. Mercer and Marc Salit. “Testing at scale during the COVID-19 pandemic.” In: *Nature Reviews Genetics* 22.7 (July 2021), pp. 415–426. ISSN: 1471-0064. DOI: 10.1038/s41576-021-00360-w. URL: https://www.nature.com/articles/s41576-021-00360-w.

[142] Kevin Murphy, Abhishek Kumar, and Stylianos Serghiou. “Risk score learning for COVID-19 contact tracing apps.” In: *Proceedings of the 6th Machine Learning for Healthcare Conference*. Machine Learning for Healthcare Conference. ISSN: 2640-3498. PMLR, Oct. 21, 2021, pp. 373–390. URL: https://proceedings.mlr.press/v149/murphy21a.html.

[143] Laura Ozella et al. “Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi.” In: *EPJ Data Science* 10.1 (Dec. 2021), pp. 1–17. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-021-00302-w. URL: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00302-w.

[144] Feng Pan et al. “Solving statistical mechanics on sparse graphs with feedback-set variational autoregressive networks.” In: *Physical Review E* 103.1 (2021), p. 012103.

[145] Nicola Perra. “Non-pharmaceutical interventions during the COVID-19 pandemic: A review.” In: *Physics Reports*. Non-pharmaceutical interventions during the COVID-19 pandemic: a review 913 (May 23, 2021), pp. 1–52. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2021.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157321000624.

[146] Giulia Pullano et al. “Underdetection of cases of COVID-19 in France threatens epidemic control.” In: *Nature* 590.7844 (2021), pp. 134–139.

[147] Robert C. Reiner et al. “Modeling COVID-19 scenarios for the United States.” In: *Nature Medicine* 27.1 (Jan. 2021), pp. 94–105. ISSN: 1546-170X. DOI: 10.1038/s41591-020-1132-9. URL: https://www.nature.com/articles/s41591-020-1132-9.

[148] Pablo Rodríguez et al. "A population-based controlled experiment assessing the epidemiological impact of digital contact tracing." In: *Nature communications* 12.1 (2021), pp. 1–6.

[149] Emily Seto, Priyanka Challa, and Patrick Ware. "Adoption of COVID-19 Contact Tracing Apps: A Balance Between Privacy and Effectiveness." In: *Journal of Medical Internet Research* 23.3 (Mar. 4, 2021), e25726. DOI: `10.2196/25726`. URL: `https://www.jmir.org/2021/3/e25726`.

[150] Mrinank Sharma et al. "Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe." In: *Nature Communications* 12 (Oct. 5, 2021), p. 5820. ISSN: 2041-1723. DOI: `10.1038/s41467-021-26013-4`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8492703/` (visited on 01/24/2023).

[151] Sibyl-Team. "SIB: [S]tatistical [I]nference in Epidemics via [B]elief Propagation." In: *code-repository* (2021). URL: `https://github.com/sibyl-team/sib`.

[152] Kaiyuan Sun et al. "Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2." In: *Science* 371.6526 (Jan. 15, 2021), eabe2424. DOI: `10.1126/science.abe2424`. URL: `https://www.science.org/doi/10.1126/science.abe2424`.

[153] Jeanne Trinquier et al. "Efficient generative modeling of protein sequences using simple autoregressive models." In: *Nature communications* 12.1 (2021), pp. 1–11.

[154] Viktor von Wyl et al. "Drivers of Acceptance of COVID-19 Proximity Tracing Apps in Switzerland: Panel Survey Analysis." In: *JMIR Public Health and Surveillance* 7.1 (Jan. 6, 2021), e25701. DOI: `10.2196/25701`. URL: `https://publichealth.jmir.org/2021/1/e25701`.

[155] Chris Wymant et al. "The epidemiological impact of the NHS COVID-19 app." In: *Nature* (2021). DOI: `10.1038/s41586-021-03606-z`. URL: `https://doi.org/10.1038/s41586-021-03606-z`.

[156] Indaco Biazzo and Fabio Mazza. "annfore: [a]utoregressive [n]eural [n]etworks [for] [e]pidemics inference problems." In: *code-repository* (2022). DOI: `https://zenodo.org/badge/latestdoi/405138309`. URL: `https://github.com/ocadni/annfore`.

[157] Indaco Biazzo and Fabio Mazza. "Repository of results for Autoregressive Neural Network for Epidemics." In: *code-repository* (2022). DOI: `https://zenodo.org/badge/latestdoi/405138833`. URL: `https://github.com/ocadni/annfore-results`.

[158] Indaco Biazzo et al. "A Bayesian generative neural network framework for epidemic inference problems." In: *Scientific Reports* 12.1 (Nov. 16, 2022), p. 19673. ISSN: 2045-2322. DOI: `10.1038/s41598-022-20898-x`. URL: `https://www.nature.com/articles/s41598-022-20898-x` (visited on 11/22/2022).

[159] *GitHub fork of the Spatiotemporal model used in this work.* `https://github.com/sibyl-team/simulator`. 2022.

[160] Muhammad Fazli Khalid et al. "Performance of Rapid Antigen Tests for COVID-19 Diagnosis: A Systematic Review and Meta-Analysis." In: *Diagnostics* 12.1 (Jan. 2022), p. 110. ISSN: 2075-4418. DOI: `10.3390/diagnostics12010110`. URL: `https://www.mdpi.com/2075-4418/12/1/110`.

[161] Jana Lasser et al. “Assessing the impact of SARS-CoV-2 prevention measures in Austrian schools using agent-based simulations and cluster tracing data.” In: *Nature Communications* 13.1 (Jan. 27, 2022), p. 554. ISSN: 2041-1723. DOI: `10.1038/s41467-022-28170-6`. URL: `https://www.nature.com/articles/s41467-022-28170-6`.

[162] Lars Lorch et al. “Quantifying the Effects of Contact Tracing, Testing, and Containment Measures in the Presence of Infection Hotspots.” In: *ACM Transactions on Spatial Algorithms and Systems* 8.4 (Nov. 2, 2022), 25:1–25:28. ISSN: 2374-0353. DOI: `10.1145/3530774`. URL: `https://doi.org/10.1145/3530774`.

[163] Dana Naous et al. “Learning From the Past to Improve the Future.” In: *Business & Information Systems Engineering* 64.5 (Oct. 1, 2022), pp. 597–614. ISSN: 1867-0202. DOI: `10.1007/s12599-022-00742-2`. URL: `https://doi.org/10.1007/s12599-022-00742-2`.

[164] Samuel Okyere and Joseph Ackora-Prah. “A Mathematical Model of Transmission Dynamics of SARS-CoV-2 (COVID-19) with an Underlying Condition of Diabetes.” In: *International Journal of Mathematics and Mathematical Sciences* 2022 (May 18, 2022), e7984818. ISSN: 0161-1712. DOI: `10.1155/2022/7984818`. URL: `https://www.hindawi.com/journals/ijmms/2022/7984818/`.

[165] *OpenStreetMap project.* 2022. URL: `https://www.openstreetmap.org/`.

[166] Ernesto Ortega, David Machado, and Alejandro Lage-Castellanos. “Dynamics of epidemics from cavity master equations: Susceptible-infectious-susceptible models.” In: *Physical Review E* 105.2 (Feb. 17, 2022). Publisher: American Physical Society, p. 024308. DOI: `10.1103/PhysRevE.105.024308`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.105.024308`.

[167] Michael J Plank et al. “Potential reduction in transmission of COVID-19 by digital contact tracing systems: a modelling study.” In: *Mathematical Medicine and Biology: A Journal of the IMA* (2022).

[168] Florian Vogt et al. “Effectiveness evaluation of digital contact tracing for COVID-19 in New South Wales, Australia.” eng. In: *The Lancet. Public Health* 7.3 (Mar. 2022), e250–e258. ISSN: 2468-2667. DOI: `10.1016/S2468-2667(22)00010-X`.

[169] *COVID-19 lockdowns*. In: *Wikipedia*. Page Version ID: 1133317800. Jan. 13, 2023. URL: `https://en.wikipedia.org/w/index.php?title=COVID-19_lockdowns&oldid=1133317800` (visited on 01/24/2023).

[170] *Operation Outbreak*. Operation Outbreak. URL: `https://operationoutbreak.org` (visited on 01/23/2023).

This Ph.D. thesis has been typeset by means of the TeX-system facilities. The typesetting engine was LuaLaTeX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TeX-system installation.