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

Doctoral Dissertation
Doctoral Program in Physics (29^{th}cycle)

Belief Propagation approach to epidemics prediction on networks

By

Jacopo Bindi

******

Supervisor(s):
Dott. Luca Dall'Asta, Supervisor

Doctoral Examination Committee:
Prof. David Saad , Referee, Aston University
Prof. Alain Barrat, Referee, Aix-Marseille Université

Politecnico di Torino
2017

Declaration

I hereby declare that, the contents and organization 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.

Jacopo Bindi
2017

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

Accounting for all the people that offered their support, advice and help during the last three years, which have been so intense, is almost impossible. Therefore, I will focus on those who have a scientific background.

First of all, I would like to sincerely acknowledge my supervisor Luca Dall'Asta for having offered me the opportunity of working with him. He is a great and inspiring teacher and, furthermore, he has shown great patience and resilience dealing with a student like me. I also would thank Alfredo Braunstein who had great influence on my work. Their passion for science and their research expertise have been of major importance for my work and for my personal growth. Although I really tried their patience, they have been supportive even under the harshest conditions. I thank all the members of the CMP group of Politecnico for being challenging and thought-provoking scientists, whom I really appreciate sharing the last three years with.

A particular remark for Anna Paola, Carla, Gino, Davide, Laura and Michele for having been great friends, as well as great physicists, and for supporting freedom in academic research. Moreover, I thank my family for having made me possible to reach such an academic achievement and life goal, and for having been always supportive. Finally, I would thank all the people that have been close to me during these years, they have been of the utmost importance to me.

Abstract

In my thesis I study the problem of predicting the evolution of the epidemic spreading on networks when incomplete information, in form of a partial observation, is available. I focus on the irreversible process described by the discrete time version of the Susceptible-Infected-Recovered (SIR) model on networks. Because of its intrinsic stochasticity, forecasting the SIR process is very difficult, even if the structure of individuals contact pattern is known. In today's interconnected and interdependent society, infectious diseases pose the threat of a worldwide epidemic spreading, hence governments and public health systems maintain surveillance programs to report and control the emergence of new disease event ranging from the seasonal influenza to the more severe HIV or Ebola. When new infection cases are discovered in the population it is necessary to provide real-time forecasting of the epidemic evolution. However the incompleteness of accessible data and the intrinsic stochasticity of the contagion pose a major challenge.

The idea behind the work of my thesis is that the correct inference of the contagion process before the detection of the disease permits to use all the available information and, consequently, to obtain reliable predictions. I use the Belief Propagation approach for the prediction of SIR epidemics when a partial observation is available. In this case the reconstruction of the past dynamics can be efficiently performed by this method and exploited to analyze the evolution of the disease. Although the Belief Propagation provides exact results on trees, it turns out that is still a good approximation on general graphs. In this cases Belief Propagation may present convergence related issues, especially on dense networks. Moreover, since this approach is based on a very general principle, it can be adapted to study a wide range of issues, some of which I analyze in the thesis.

Table of contents

List of figures **8**

List of tables **16**

Introduction **1**

I Methods **5**

1 Epidemic models **7**

1.1 Compartmental models 7

1.2 Deterministic approximation in homogeneously mixed population 9

1.3 SIR epidemic model on networks 11

1.4 Individual and pair based approximation 12

1.5 Message Passing approach to SIR 15

1.5.1 Equivalence to the Dynamic Message Passing formulation 20

1.5.2 Equivalence to the pair-based approximation 24

1.5.3 Towards a new representation 28

1.6 Epidemic processes in temporal networks 30

1.7 Non Markovian epidemic model 31

2 Source inference for SIR on networks **32**

2.1 Rumor Centrality 32

2.2 Soft-Margin Estimator 35

2.3 Dynamic Message Passing – Naive Bayes 37

3 Belief Propagation **39**
3.1 Factor Graph Representation 39
3.2 The Belief Propagation equations 41
3.3 The Bethe Approximation 45
3.4 Correspondence between BP and Bethe Approximation 49
3.4.1 The Bethe free-energy as a function of BP messages . . . 50

4 SIR Prediction via Bayesian approach **52**
4.1 Bayesian Approach 52
4.2 The Belief Propagation approximation 53
4.2.1 Factor graph for the SIR model 54
4.2.2 Disentangled factor graph representation 58
4.2.3 Unknown initial time 60
4.2.4 BP updates for the SIR model 60
4.3 Temporal networks 64
4.4 The patient-zero problem via Belief Propagation 65
4.5 Extinction Time 66
4.6 Non-Markovian models 69

II Results **71**

5 Predicting from partial observations **73**
5.1 Sampling methods 74
5.2 Computational Cost 76
5.3 Epidemic Size and individual classification 77
5.3.1 A case study of real contact network 87
5.4 Extinction Time probability distribution 91

6 Generalizations **97**
6.1 Temporal Networks 97
6.2 Non-Markovian SIR model: fixed recovery delay 99
6.3 Non-Markovian SIR: non-infective period 101

Conclusions **102**

Bibliography **105**

Appendix A Networks **112**

References 113

Bibliography **114**

List of figures

1.1 Two main examples of compartment models. In the SIR model (upper row) individuals go from compartment *susceptible* to *infected* with probability λ. They move from *infected* to *recovered* with a probability μ. In the SIS model infected individuals go back to *susceptible* compartment with rate μ. 8

1.2 Two contributions make up the probability that node i is not infected by j before time $t+1$. The upper line represents the probability that j is infected and fails to transmit the disease before $t+1$. In the lower line, j is scheduled to transmit the disease at time $\tau < t+1$, but it is neither infected by its neighbors r, k, q on time (i.e before $t-\tau$) nor it was infected at $t=0$ with probability $1-P_j^S(0)$. 16

1.3 When the underlying network has a loopy structure the transmission events to one node from its neighbors are correlated. Node 1 is the seed and two different paths (blue and red) of infection lead to node 5 and 6. In the message passing approach they are considered as coming from different sources but actually only one contributes. 18

1.4 Temporal network: a set of nodes that at every time step is connected by a different set of edges. Considering the set of edges occurred before a given time, one gets a static projection of the network. 30

2.1 This example shows a typical situation in which even if DMP gives exact marginal probabilities the mean-field approximation introduces errors in the inference of the source. Red: node infected; Black: node recovered; White: node susceptible. 38

3.1 A factor graph. Circles correspond to variables $x_i \in (x_1, x_2, x_3, x_4, x_5)$. Black squares correspond to functions encoding dependencies among variables $F_a(x_2, x_3), F_b(x_3, x_4), F_c(x_1, x_2, x_4, x_5), F_d(x_3)$. . . 40

3.2 Factor graph for an Ising chain. Circles correspond to the Ising spins $x_i \in \{+1, -1\}$. Black squares correspond to the pairwise interaction terms $F_a(x_i, x_{i+1}) = e^{-x_i x_{i+1}}$. Grey squares correspond to the magnetic field terms $\phi_a(x_i) = e^{-Bx_i}$. 42

3.3 Left: the local portion of the factor graph involved in the computation of $m_{j \to F_a}(x_j)$. This message is a function of the incoming messages $p_{F_a \to j}(x_j)$ with $b \neq a$. Right: the portion of the factor graph involved in the computation of $p_{F_b \to j}(x_j)$ as a function of messages $m_{k \to F_a}(x_k)$ with $k \neq j$. 43

3.4 Each region in $\mathscr{R}_L$ contains one factor (for example F_a or F_b) and all its neighboring nodes (respectively blue and red ones). Each region in $\mathscr{R}_S$ contains a single variable node. In the example node x_k belongs to both the F_a and F_b regions and to the single variable region in $\mathscr{R}_S$. 47

4.1 (a): original graph with nodes (i, j, k). (b): the factor graph representation. Even in simple cases it shows a loopy structure. 57

4.2 The factor graph representation for the graph in fig. 4.1(a) when pairs of variables (t_i, t_j) are grouped together. 58

4.3 The factor graph representation for the graphical model associated to the distribution (4.18). 59

4.4 The factor graph representation for the graphical model associated to the distribution (4.18) with unknown initial time. 61

5.1 Each line represents a different realization for the SIR epidemic process given the (same) incomplete observation of the initial condition. Configurations in the leftmost column represent the observed state of the system, the other columns represent the time evolution of the epidemic process in that specific realization. Nodes colors: Green = Susceptible, Red = Infected, Black = Recovered, White = Unobserved. 75

5.2 a) Area under the ROC curve as function of the time $t > T_{obs} = 3$ on a random regular graph of $N = 1000$ nodes and average degree $k = 4$. The average is computed over $M_o = 10^3$ epidemic realizations (with homogeneous parameters $\lambda = 0.7$, $\mu = 0.5$); the vertical bars represent the standard error of the mean. The prediction is obtained after the observation at T_{obs} of a 10%-fraction of the nodes chosen randomly (*random observation*). b) Predicted average epidemic size on random regular graphs ($N = 1000$, $k = 4$, $\lambda_{ij} = 0.7$, $\mu_i = 0.5$) as function of time for a random observation of 10% of the nodes at $T_{obs} = 3$. The inference methods used are direct sampling with complete observation (black), random sampling (green), density sampling (blue), similarity sampling (magenta) and belief propagation (red). 79

5.3 a) The heatplots represent the average AUC as function of time and of the number of observed nodes that were infected before T_{obs}, computed by density sampling, similarity sampling, belief propagation. b) The average epidemic size predicted by density sampling, similarity sampling and belief propagation is also shown as function of the number of infected and recovered nodes in the observation. As a reference, in both panels, we plot results obtained, for the same realizations of the SIR process, by direct sampling with complete observation. The horizontal axis refers to the number of infected or recovered nodes present in the 10% observation (also in the case of complete observation). 81

5.4 Area under the ROC curve as function of the time $t > T_{obs} = 3$ on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ (with homogeneous epidemic parameters $\lambda = 0.5$, $\mu = 0.6$), in the case of observation of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The average is computed over $M = 201$ epidemic realizations. The inference methods used are direct sampling with complete observation (black), random sampling (green), density sampling (blue), similarity sampling (magenta) and belief propagation (red). 83

5.5 The heatplots represent the average AUC as function of time and of the number of observed nodes that were infected before $T_{obs} = 3$, computed by density sampling, similarity sampling, belief propagation, on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ with homogeneous parameters $\lambda = 0.5$, $\mu = 0.6$. As a reference, we also plot results obtained, for the same realizations of the SIR process, by direct sampling with complete observation. The prediction is obtained after the observation at T_{obs} of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The horizontal axis refers to the number of infected or recovered nodes present in the 30% observation (also in the case of complete observation). 84

5.6 Predicted average epidemic size as function of the time $t > T_{obs} = 3$ on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ (with homogeneous epidemic parameters $\lambda = 0.5$, $\mu = 0.6$), in the case of observation of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The average is computed over $M = 201$ epidemic realizations. The inference methods used are direct sampling with complete observation (black), density sampling (blue), similarity sampling (magenta) and belief propagation (red). 86

5.7 The heatplots represent the average epidemic size as function of time and of the number of observed nodes that were infected before $T_{obs} = 3$, computed by density sampling, similarity sampling, belief propagation, on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ with homogeneous parameters $\lambda = 0.5$, $\mu = 0.6$. As a reference, we also plot results obtained, for the same realizations of the SIR process, by direct sampling with complete observation. The prediction is obtained after the observation at T_{obs} of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The horizontal axis refers to the number of infected or recovered nodes present in the 30% observation (also in the case of complete observation). 86

5.8 Average area under the ROC curve (a,c) and average epidemic size (b,d) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.5$, $\mu = 0.4$) on the SC network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 4$ (a,b) and $T_{obs} = 8$ (c,d). In all plots direct sampling from a complete observation is shown for comparison (black). 89

5.9 Average area under the ROC curve (a,c) and average epidemic size (b,d) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.5$, $\mu = 0.4$) on the SC network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 4$ (a,b) and $T_{obs} = 8$ (c,d). For $T_{obs} = 8$, only instances with a number of observed infected and recovered nodes $N_{I+R} > 18$ is considered (75% of instances). For $T_{obs} = 4$, only instances with observed infected and recovered nodes $N_{I+R} > 6$ is considered (46% of instances). In all plots direct sampling from a complete observation is shown for comparison (black). 90

5.10 Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.5$, $\mu = 0.4$) on the WSC network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 8$. Direct sampling from a complete observation is shown for comparison (black). 91

5.11 Extinction time distributions for different complete observations: a) on trees with branching ratio $k = 3$ and $N = 1092$ (epidemic parameters $\lambda = 0.7$, $\mu = 0.5$ and observation time $T_{obs} = 5$); b) random regular graphs of degree 4 and $N = 1000$ nodes ($\lambda = 0.7$, $\mu = 0.5$ and $T_{obs} = 4$). (c)-(e): similar realizations of the epidemic process at T_{obs} on a tree graph corresponding to rather different predicted extinction time distributions with maximum value respectively at $T = 21$ (c), $T = 23$ (d), and $T = 16$ (e). Nodes color: Green= Susceptible, Red= Infected, Black= Recovered. 93

5.12 Absolute value of the difference between the extinction time distribution $P_{ext}(t)$ computed from direct sampling with complete information and those calculated with density sampling (blue), BP (red) and similarity sampling (magenta). a) On trees of $N = 1092$ nodes, with branching ratio 3 ($\langle k \rangle \approx 2$) and with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 10% of the nodes at $T_{obs} = 5$ and averaging over $M_o = 210$ such realizations. b) On random regular graphs of $N = 1000$ nodes and degree $k = 4$ with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 30% of the nodes at $T_{obs} = 4$ and averaging over $M_o = 150$ such realizations. 95

5.13 Absolute value of the difference between the extinction time probability distribution $P_{ext}(t)$ computed from direct sampling with complete information and those calculated with density sampling, BP and similarity sampling as a function of the number of infected and recovered nodes in the observed subset of nodes. a) On trees of $N = 1092$ nodes, with branching ratio 3 ($\langle k \rangle \approx 2$) and with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 10% of the nodes at $T_{obs} = 5$ and averaging over $M_o = 210$ such realizations. b) On random regular graphs of $N = 1000$ nodes and degree $k = 4$ and with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 30% of the nodes at $T_{obs} = 4$ and averaging over $M_o = 150$ such realizations. 96

6.1 Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.6$, $\mu = 0.02$) on the HN. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 15$. Direct sampling from a complete observation is shown for comparison (black). The average is computed on $M = 50$ epidemic realizations. 98

6.2 Number of contacts per time step in the High School network with a partition composed of $T = 25$ time steps. 99

6.3 Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.2$, the recovery occurs after $G = 10$ time steps) on a BA network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 10$. Direct sampling from a complete observation is shown for comparison (black). The average is computed on $M = 50$ epidemic realizations. 100

6.4 Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.6$, $\mu = .3$ and a node start infecting $L = 1$ time steps after the infection) on a BA network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 5$. Direct sampling from a complete observation is shown for comparison (black). The average is computed on $M = 50$ epidemic realizations. 102

List of tables

Introduction

Throughout history the spreading of infectious diseases has been a major threat to society, indeed one of the earliest attempt to model an epidemics (namely smallpox) dates back to mid 18th century and it is due to Daniel Bernoulli [1, 2]. In today's interconnected and interdependent society, infectious diseases pose the threat of a worldwide epidemic spreading, hence governments and public health systems maintain surveillance programs to report and control the emergence of new disease event ranging from the seasonal influenza to the more severe HIV or Ebola. Alongside surveillance programs, quantitative analysis and predictive tools gain importance to support policy-making [3]. On the one hand modeling efforts have been made to provide plausible scenarios and to evaluate containment procedures in the case of epidemic outbreaks [4–6]. On the other hand when new infection cases are discovered in the population it is necessary to provide real-time forecasting of the epidemic evolution [7–9]. However the incompleteness of accessible data and the intrinsic stochasticity of the contagion pose a major challenge.

In this thesis we study the problem of predicting the evolution of the epidemic spreading on networks when incomplete information, in form of a partial observation, is available. Different mathematical descriptions of contagion processes are possible, depending on the disease to be studied. We focus on the irreversible process (the reason will be clear later) described by the discrete time version of the Susceptible-Infected-Recovered (SIR) model on networks, a cornerstone in the modern mathematical modeling of infectious diseases originally formalized by W. O. Kermack and A. G. McKendrick in 1927 [10]. The underlying network structure approximates the strong dependencies naturally present in contagion processes, in fact whether an individual contracts the disease or not depends on the status of the other individuals he makes contact with. This is why social contact patterns have

recently become subjects of several studies [11–14], but, when different temporal and spatial scales are taken into consideration, it is necessary to assume different approximations of the underlying contact pattern [15]. Because of its intrinsic stochasticity, forecasting the SIR process is very difficult, even if the structure of individuals contact pattern is known, as recently shown by Petter Holme [16–18].

In its original formulation the SIR model is a simple Markov process, nevertheless it becomes rapidly intractable as the number of individuals involved grows. A well-known class of approximations yields exact results when the underlying network is a tree [19, 20], such as the Message Passing Approach firstly proposed by Karrer and Newman [21], and provides meaningful predictions in loopy graphs. This approaches require as set of initial conditions, but in a realistic case they usually cannot be provided by the empirical observations. In general the process is only partially observed: for example, it is usually not known by whom an sick individual was infected, nor the time of the infection.

The idea behind the work of this thesis is that the correct inference of the contagion process before the detection of the disease permits to use all the available information and, consequently, to obtain reliable predictions. The problem of reconstructing the history of an epidemics has indeed gained attention recently [22–26]. Many studies tackled the problem of identifying the origin of an epidemic outbreak (or *patient zero problem*), but, unfortunately, turned out to be a difficult problem to solve. In fact, when it is formulated as a maximum likelihood estimation problem, it corresponds to an optimization problem in the space of all possible epidemic realizations that are compatible with the available data. However, the number of possible epidemic realizations is exponentially large in the number of individuals, therefore every epidemic realizations appears, for any practical inferential purpose, a rare event. From this point of view, Shah and Zaman [25, 27, 28] proposed a maximum likelihood estimator, called *rumor centrality*, in the case of a single epidemic source and exact on regular trees. A Bayesian approach has been proposed by Lokhov *et al.* [29], that applied the Message Passing formulation by Karrer and Newman [21] to the inference of the source by a further mean-field approximation of the likelihood function (also called *naive Bayes method*). However the problem

is still open.

In the case of the irreversible dynamics, Altarelli *et al.* [30, 31] provided a method to study the statistical properties of rare events and find the initial conditions that give rise to a desired final configuration. They showed that a static representation of the dynamics allows to recast the problem into the computation of a partition function, in which the trajectories evaluated must meet the constraints on the dynamical rules and are weighted by an additional and opportunely defined energetic term. Then, on a network, the partition function can be computed by the Belief Propagation approximation. Following, the core idea of this approach has been exploited to study the dynamics of the SIR model [32] and the inference of the patient zero [33]. The constraint on the dynamical rules must be changed accordingly to the SIR dynamics and the energetic term weighting the trajectories becomes an hard constraint on the epidemic realizations that must be compatible with the observed configuration.

In this thesis, we use the Belief Propagation approach for the prediction of SIR epidemics when a partial observation is available. In this case the reconstruction of the past dynamics can be efficiently performed by this method and exploited to analyze the evolution of the disease. Although the Belief Propagation provides exact results on trees, it turns out that is still a good approximation on general graphs. In this cases Belief Propagation may present convergence related issues, especially on dense networks. Moreover, since this approach is based on a very general principle, it can be adapted to study a wide range of issues, some of which we analyze in this thesis.

- Chapter 1: we introduce the SIR model and the deterministic approximation for homogeneously mixed population. Then we discuss different approximations of the stochastic process and we show that the message passing approach is equivalent to the pair-based approximation. Lastly, we introduce non-Markovian epidemic models and temporal networks.

- Chapter 2: we discuss some approaches to the identification of the epidemic source, such as the rumors centrality, the Dynamic Message Passing + Naive Bayes and the Soft-margin estimator.

- Chapter 3: we introduce the factor graph representation for probability distributions of variable with local dependencies, then we discuss the Belief Propagation algorithm and the Bethe Approximation for the free-energy.
- Chapter 4: we introduce and analyze in details the Belief Propagation approach to SIR model, the corresponding factor graph representation, the update rules for the messages and computation of the extinction time probability distribution.
- Chapter 5: we discuss the results for the prediction of SIR evolution on different network architectures, for the computation of the extinction time probability distribution. Lastly, we discuss the results when a real-world network is considered.
- Chapter 6: we discuss the extension to non-Markovian models and temporal networks.

Part I

Methods

Chapter 1

Epidemic models

During its longstanding history, epidemic modeling has evolved spanning across several research areas, from biology to physics. Epidemic models describe the dynamical evolution of contagion processes within a population and compartmental models are the simplest and most common class of models concerned with epidemics spreading. In this chapter we will provide few basic notions about compartmental models and the deterministic approximation. Then we discuss the individual based mean-field approximation to SIR epidemic model on network, as well as a message passing approach. In conclusion we introduce cutting-hedge scenarios as the epidemic modeling on temporal networks and non-Markovian models.

1.1 Compartmental models

Compartmental models assume that the population can be divided in compartments depending on the stage of the disease. The most common compartments are: *susceptible*, i.e those who can contract the infection (denoted by S); *infected*, those who contracted the infection and are contagious (denoted by I); *recovered*, those who recovered from the disease (denoted by R). Depending on the specific disease considered it is possible to add different compartments to provide a better description. In order to study the evolution of the epidemic process as function of time, we have to define the individual-level processes that govern the transaction of individuals from one compartment to another. For example, let us consider the

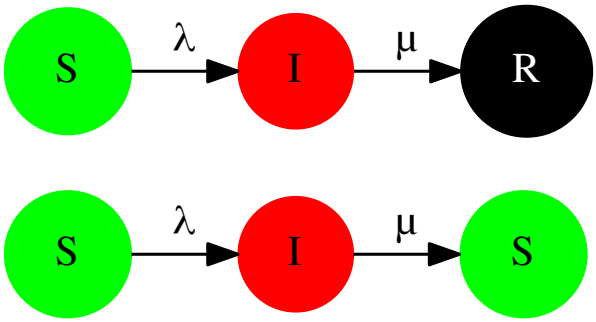


Figure 1.1: Two main examples of compartment models. In the SIR model (upper row) individuals go from compartment *susceptible* to *infected* with probability λ. They move from *infected* to *recovered* with a probability μ. In the SIS model infected individuals go back to *susceptible* compartment with rate μ.

SIS and SIR models that are among the simplest epidemic models. The SIS model is based on two compartments and it is defined by two transitions: transition $S \to I$ occurs when a susceptible individual interacts with an infected one and becomes infected; transition $I \to S$ occurs when an infected individual recovers and comes back to the susceptible population. The SIS model is suitable for diseases that do not confer immunity (or death), then the individuals can be infected over and over again. The cycle $S \to I \to S$ can be sustained forever and in a long time regime the SIS can exhibit an *endemic* state characterized by a constant (on average) fraction of the population infected. The SIR model is based on three compartments and two transitions: $S \to I$ as the SIS model; $I \to R$ (instead of the SIS $I \to S$) that occurs when an infected individual recovers from the disease and acquires permanent immunity or is removed (e.g has died). In the SIR model the number of infected individual tends to zero for long time. The parameter involved in the transitions described can be estimated, for example, from clinical data and it is usual to assume that transitions probability are constant in order to deal with simplified model (in Section 1.7 we will introduce different assumptions).

When considering epidemic spreading in a population, we face the problem of describing the contact pattern through which the contagion occurs. Most common approaches assume different time scales for the contagion dynamics and contacts between individuals (in Section 1.6 we will introduce a different assumption). When describing social contact pattern in term of networks every individual is represented by a vertex and contacts between pairs of individuals are represented by the edges of the graph. This description can be either a correct representation or an approximation, depending on the system under studied. Moreover one can consider a network whose nodes rewire on a time scale much faster than the contagion – for example, mean-field approaches rely upon this assumption, disregarding connections between individuals (Section 1.2). In the opposite limit, when we assume that the contact pattern evolves on a time scale much slower than the epidemic process, static networks provide a suitable description. In the following sections we will focus on approximated methods for the SIR model [15, 34].

1.2 Deterministic approximation in homogeneously mixed population

The most basic approach relies on the homogeneous mixing approximation which assumes that individuals interacts randomly with each other within each compartment $[S, I, R]$. This assumption is equivalent to say that the underlying contact pattern evolves at a time scale much faster than the epidemic – it is an appropriate approximation, for example, in case of large population or in the case of airborne disease. In this framework we are interested in the equations governing the evolution of density of individuals in each stage of the disease $i(t)$, $s(t)$ and $r(t)$ [34]. Now they represent fractions of the overall population that move from one compartment to another, but we describe the microscopic foundation to the population level approach in Section 1.3. Since the elementary transitions are qualitatively the same, it is possible to obtain the global infected, susceptible and recovered densities by summing over the single individual probabilities $P_I^i(t)$ and $P_S^i(t)$. We consider an approximation in which each individual in I interacts uniformly with an individuals in S with a probability proportional to $i(t)$, i.e we consider an effective interaction and determine the probability of infection in the same way for all individuals in the

system, then it is equivalent to a mean-field approximation. Thus instead of the probability of infection between two individual λ_{ij}, we will use a transition rate β to the infected compartment. The equations for the susceptible-infected-recovered model are

$$\begin{aligned} \frac{di}{dt} &= \beta i(t)s(t) - \mu i(t) \\ \frac{ds}{dt} &= -\beta i(t)s(t) \end{aligned} \tag{1.1}$$

The normalization conditions results in $r(t) = 1 - s(t) - i(t)$. In the early stage of the epidemics it is possible to assume that $i(t) \ll 1$ (consequently $s(t) \simeq 1$) and in this limit equation (1.1) can be linearized

$$\frac{di(t)}{dt} \simeq (\beta - \mu)i(t) \tag{1.2}$$

and its solution valid for early time

$$i(t) \simeq e^{(\beta-\mu)t} i(0). \tag{1.3}$$

In this framework the fraction of individual infected grows exponentially when

$$\beta - \mu > 0 \rightarrow R_0 = \frac{\beta}{\mu} > 1 \tag{1.4}$$

where R_0 is the *reproduction number* and it is defined as the average number of secondary infections by a primary case introduced in a fully susceptible population. This is one of the key concept of classical theoretical analysis of epidemic models and allows to introduce the concept – important as well – of epidemic threshold. If $R_0 > 1$, a single infected individual generates on average more than one secondary infection, then the process evolves causes an epidemic outbreak of finite relative size in SIR case or an endemic state, with a finite fraction of the population infected, in the SIS case. If $R_0 < 1$, a single infected individual generates on average less than one secondary infection, then the epidemic process involves only a negligible fraction of the population, vanishing in the limit of a large population in SIR model or leading to a steady state with all individuals susceptible in the SIS model.

1.3 SIR epidemic model on networks

The SIR model [10] applies to a the wide range of diseases that provides immunity subsequent the recovery as well as to knowledge and information diffusion [35, 36]. We introduce a discrete time version of the SIR model. In fact the most natural approach to describe the disease spreading relies on the definition of a *probability of infection* λ between two individual, the *probability of recover* μ and the reactions

$$S + I \xrightarrow{\lambda} 2I$$
$$I \xrightarrow{\mu} R.$$

We already pointed out that in a realistic case each individual can get infected or can infect by interacting with other individuals through a social contact network. In order to introduce the discrete time SIR model on networks, let us consider a graph $G = (V, E)$ that represents a contact network of $N = |V|$ individuals. The state of a node i at time t is represented by a variable $x_i^t \in \{S, I, R\}$ and the configuration of the population at time t is $\mathbf{x}^t = (x_1, \ldots, x_N)$. The process is irreversible, so once a node has recovered it does not get infected anymore. In a discrete time representation an infected node i can infect each of his neighbors $j \in \partial i$ (where ∂i is the set of neighbors of node i) with a given probability λ_{ij}, then recover with probability μ_i. If we assume that λ_{ij} and μ_i do not depend on time, we can describe a Markov chain by the use of the following transition probabilities

$$P(x_i^{t+1} = S|\mathbf{x}^t) = \mathbb{I}[x_i^t = S] \prod_{j \in \partial i} (1 - \lambda_{ij} \mathbb{I}[x_j^t = I]) \tag{1.5}$$

$$P(x_i^{t+1} = I|\mathbf{x}^t) = (1 - \mu_i)\mathbb{I}[x_i^t = I] + \mathbb{I}[x_i^t = S](1 - \prod_{j \in \partial i} (1 - \lambda_{ij} \mathbb{I}[x_j^t = I])) \tag{1.6}$$

$$P(x_i^{t+1} = R|\mathbf{x}^t) = \mathbb{I}[x_i^t = R] + \mu_i \mathbb{I}[x_i^t = I]. \tag{1.7}$$

In the case of SIR model it is possible to define the duration of the epidemic as the time between the first infection and the last recovery, thus providing the notion of *extinction time* as the time at which the epidemics dies out, i.e when no infected nodes are present in the population [37, 16].

1.4 Individual and pair based approximation

When considering a discrete time SIR model with constant infection $\lambda_{ij}\forall(i,j)\in E$ and recovery parameters $\mu_i\forall(i)\in V$, equations (1.7) provide transition probabilities for a master equation. Then we can obtain the probability that a node i is in state $X\in\{S,I,R\}$ through:

$$P(x_i^{t+1}=X)=\sum_{\mathbf{x}^t}P(x_i^{t+1}=X|\mathbf{x}^t)P(\mathbf{x}^t). \tag{1.8}$$

When the infected and susceptible state are considered the corresponding equations are

$$\begin{aligned}
P(x_i^{t+1}=S)&=\sum_{\mathbf{x}^t}\mathbb{I}[x_i^t=S]\prod_{j\in\partial i}(1-\lambda_{ij}\mathbb{I}[x_j^t=I])P(\mathbf{x}^t)\\
P(x_i^{t+1}=I)&=\sum_{\mathbf{x}^t}\left\{(1-\mu_i)\mathbb{I}[x_i^t=I]+\mathbb{I}[x_i^t=S]\left[1-\prod_{j\in\partial i}(1-\lambda_{ij}\mathbb{I}[x_j^t=I])\right]\right\}P(\mathbf{x}^t),
\end{aligned} \tag{1.9}$$

the probability for the recovered state is obtained by the normalization constraint. In general distribution probability $P(\mathbf{x}^t)$ shows non-trivial dependencies between individual states and we need approximations in order to get a tractable form. In the continuous time case the closure of equations (1.4) has been investigated by Sharkey [20, 38, 39], as well as the validity of different approximations. The simplest possibility is assuming that the state of a node i does not depend on the state of any other node j (*individual based approach*) [40], then we can assume a single site factorization $P(\mathbf{x}^t)=\prod_i P(x_i^t)$ (mean-field approximation). The resulting equations are:

$$\begin{aligned}
P(x_i^{t+1}=S)&=P(x_i^t=S)\prod_{j\in\partial i}\left[1-\lambda_{ij}P(x_j^t=I)\right]\\
P(x_i^{t+1}=I)&=(1-\mu_i)P(x_i^t=I)+P(x_i^t=S)\left[1-\prod_{j\in\partial i}(1-\lambda_{ij}P(x_j^t=I))\right]\\
P(x_i^{t+1}=R)&=P(x_i^t=R)+\mu_i P(x_i^t=I).
\end{aligned} \tag{1.10}$$

However this assumption is a drastic simplification, in fact the pairwise statistical independence approximation $P(x_i^t = S, x_j^t = I) = P(x_i^t = S)P(x_j^t = I)$ leads to anomalous terms that appear when we take into consideration the equation for the joint probability $P(x_i^{t+1} = S, x_j^{t+1} = I)$, that is

$$
\begin{aligned}
P(x_i^{t+1} = S, x_j^{t+1} = I) &= \sum_{\mathbf{x}^t} P(x_i^{t+1} = S, x_j^{t+1} = I|\mathbf{x}^t)P(\mathbf{x}^t) \simeq P(x_i^{t+1} = S)\sum_{\mathbf{x}^t} P(x_j^{t+1} = I|\mathbf{x}^t)P(\mathbf{x}^t) \\
&= P(x_i^{t+1} = S)\sum_{\mathbf{x}^t} P(\mathbf{x}^t)\left\{(1-\mu_j)\mathbb{I}[x_j^t = I] + \mathbb{I}[x_j^t = S]\left[1 - \prod_{z\in\partial j}(1-\lambda_{zj}\mathbb{I}[x_z^t = I])\right]\right\} \\
=&P(x_i^{t+1} = S)(1-\mu_j)P(x_i^t = I) + P(x_i^{t+1} = S)P(x_j^t = S) \\
&-P(x_i^{t+1} = S)P(x_j^t = S)\prod_{z\in\partial j}(1-\lambda_{zj}\mathbb{I}[x_z^t = I]) \\
=&P(x_i^{t+1} = S)(1-\mu_j)P(x_i^t = I) + P(x_i^{t+1} = S)P(x_j^t = S) \\
&- P(x_i^{t+1} = S)P(x_j^t = S)(1-\lambda_{ij}P(x_i^t = I))\prod_{k\in\partial j\setminus i}(1-\lambda_{kj}P(x_k^t = I)) \\
=&P(x_i^{t+1} = S)(1-\mu_j)P(x_i^t = I) + \\
&P(x_i^{t+1} = S)P(x_j^t = S)\left[1 - \prod_{k\in\partial j\setminus i}(1-\lambda_{kj}P(x_k^t = I))\right] \\
&+ P(x_i^{t+1} = S)P(x_j^t = S)\lambda_{ij}P(x_i^t = I)\prod_{k\in\partial j\setminus i}(1-\lambda_{kj}P(x_k^t = I)).
\end{aligned}
\tag{1.11}
$$

This expression involves terms inconsistent with the master equation and the premises of the compartmental model. For example last term in (1.11) accounts for the probability that node i infects node j (being infected at time t) and that the same node is susceptible at time $t+1$ that implies it is also susceptible at t): this is not possible in the SIR process. Considering the correlation between the state of node i and node j, the probability that i is susceptible and it infects j at the same time would be zero. In order to remove this anomalous term we can consider a *pair-based approximation* [19, 38], that means factorizing the probability distribution as follows:

$$
P(\mathbf{x}^t) = \frac{\prod_{(ij)\in E} P(x_i^t, x_j^t)}{\prod_{i\in V} P(x_i^t)}. \tag{1.12}
$$

This factorization reproduces the exact probability distribution when the underlying network is a tree [39]. Equations become

$$
\begin{aligned}
P(x_i^{t+1}=S) &= P(x_i^t=S)\prod_{j\in\partial i}\left[1-\lambda_{ji}\frac{P(x_i^t=S,x_j^t=I)}{P(x_i^t=S)}\right]\\
P(x_i^{t+1}=I) &= (1-\mu_i)P(x_i^t=I)+P(x_i^t=S)\left[1-\prod_{j\in\partial i}(1-\lambda_{ji}\frac{P(x_i^t=S,x_j^t=I)}{P(x_i^t=S)})\right]\\
P(x_i^{t+1}=R) &= P(x_i^t=R)+\mu_i P(x_i^t=I).
\end{aligned}
\tag{1.13}
$$

Then the equation for the joint probability $P(x_i^t=S,x_j^t=I)$ is

$$
\begin{aligned}
P(x_i^{t+1}=S,x_j^{t+1}=I) &= \sum_{\mathbf{x}} P(x_j^{t+1}=I|x_i^{t+1}=S,\mathbf{x}^t)P(x_i^{t+1}=S|\mathbf{x}^t)P(\mathbf{x}^t)\\
&=\sum_{\mathbf{x}}\left[(1-\mu_j)\mathbb{I}[x_j^t=I]+\mathbb{I}[x_j^t=S](1-\prod_{z\in\partial j}(1-\lambda_{zj}\mathbb{I}[x_z^t=I]))\right]\\
&\quad\times\mathbb{I}[x_i^t=S]\prod_{k\in\partial i}(1-\lambda_{ki}\mathbb{I}[x_k^t=I])\times\frac{\prod_{(u,v)\in E}P(x_u^t,x_v^t)}{\prod_{u\in V}P(x_u^t)}\\
&=(1-\mu_j)P(x_i^t=S,x_j^t=I)\prod_{k\in\partial i}[1-\lambda_{ki}\frac{P(x_k^t=I,x_i^t=S)}{P(x_i^t)=S}]+\\
&+P(x_i^t=S,x_j^t=S)\left[1-\prod_{z\in\partial j\setminus i}(1-\lambda_{zj}\frac{P(x_z^t=I,x_j^t=S)}{P(x_j^t)=S})\right]\prod_{k\in\partial i\setminus j}[1-\lambda_{ki}P(x_k^t=I,x_i^t=S)]\\
&=(1-\mu_j)P(x_i^t=S,x_j^t=I)(1-\lambda_{ji})\prod_{k\in\partial i\setminus j}[1-\lambda_{ki}\frac{P(x_k^t=I,x_i^t=S)}{P(x_i^t)=S}]\\
&+P(x_i^t=S,x_j^t=S)\left[1-\prod_{z\in\partial j\setminus i}(1-\lambda_{zj}\frac{P(x_z^t=I,x_j^t=S)}{P(x_j^t)=S})\right]\prod_{k\in\partial i\setminus j}[1-\lambda_{ki}\frac{P(x_k^t=I,x_i^t=S)}{P(x_i^t)=S}].
\end{aligned}
\tag{1.14}
$$

For the joint probability $P(x_i^t = S, x_j^t = S)$ we get:

$$\begin{aligned}
P(x_i^{t+1} = S, x_j^{t+1} = S) &= \sum_{\mathbf{x}} P(x_j^{t+1} = S | x_i^{t+1} = S, \mathbf{x}^t) P(x_i^{t+1} = S | \mathbf{x}^t) P(\mathbf{x}^t) \\
&= \sum_{\mathbf{x}} \mathbb{I}[x_j^t = S] \prod_{z \in \partial j} (1 - \lambda_{zj} \mathbb{I}[x_z^t = I]) \\
&\quad \times \mathbb{I}[x_i^t = S] \prod_{k \in \partial i} (1 - \lambda_{ki} \mathbb{I}[x_k^t = I]) \times \frac{\prod_{(u,v) \in E} P(x_u^t, x_v^t)}{\prod_{u \in V} P(x_u^t)} \\
&= P(x_i^t = S, x_j^t = S) \prod_{z \in \partial j \setminus i} (1 - \lambda_{zj} \frac{P(x_z^t = I, x_j^t = S)}{P(x_j^t) = S}) \prod_{k \in \partial i \setminus j} [1 - \lambda_{ki} \frac{P(x_k^t = I, x_i^t = S)}{P(x_i^t) = S}].
\end{aligned} \tag{1.15}$$

Equations (1.13), (1.14), and (1.15) are a closed set of equations in the pair-based approximation for the time evolution of the probability distributions that describe a discrete time SIR model with constant epidemic parameters. In Section 1.5.2 we show that this formulation is equivalent to the Dynamic Message Passing approach firstly proposed by Lokhov, Mézard, Ohta and Zdeborová and inspired to the general formulation by Karrer and Newman.

1.5 Message Passing approach to SIR

The dynamics of epidemic models on networks can be formulated in terms of a message passing algorithm known as Belief Propagation. We will discuss more in detail Belief Propagation in Section 3, here instead we introduce the approach proposed by B. Karrer and M. E. J. Newman [21].

Let us consider a discrete time SIR model on a network. We call $s(\tau)$ the probability that an infected individual contacts a neighbor at after τ time steps, in other words $\tau + 1$ is the time needed to have a successful transmission. The probability that an infected individual recovers after τ is $r(\tau)$. We do not assume any explicit functional form. The probability that a node j infects without recovery one of its neighbors i a time τ from its infection is

$$f_{ji}(\tau) = s(\tau)(1 - \sum_{g=0}^{\tau-1} r(g)) \tag{1.16}$$

This formulation elects as fundamental quantity the probability $H^{ji}(t+1)$ that a node j has not infected neighboring node i until time $t+1$. In the simple case of a tree this quantity has only two contributions: the probability that j is infected but fails to transmit the disease to i before time t, i.e

$$1 - \sum_{\tau=0}^{t} f_{ji}(\tau); \tag{1.17}$$

and the probability that j contacts i in a time $\tau < t+1$ after infection but, since it has not been infected by its neighbors before, j does not infect i before $t+1$. That is

$$\sum_{\tau=0}^{t} f_{ji}(\tau) P(x_i^0 = S) \prod_{k \in \partial j \setminus i} H^{kj}(t-\tau) \tag{1.18}$$

where $P(x_i^0 = S)$ is the probability that j is not infected at $t = 0$. Combining the

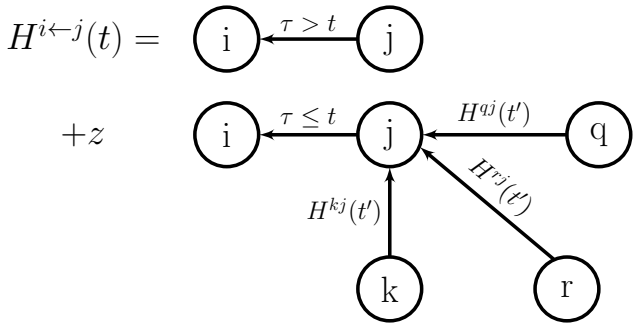


Figure 1.2: Two contributions make up the probability that node i is not infected by j before time $t+1$. The upper line represents the probability that j is infected and fails to transmit the disease before $t+1$. In the lower line, j is scheduled to transmit the disease at time $\tau < t+1$, but it is neither infected by its neighbors r, k, q on time (i.e before $t-\tau$) nor it was infected at $t = 0$ with probability $1 - P_j^S(0)$.

two contributions we get the following message passing equation

$$H^{ji}(t+1) = 1 - \sum_{\tau=0}^{t} f_{ji}(\tau) \left[1 - P(x_i^0 = S) \prod_{k \in \partial j \setminus i} H^{kj}(t-\tau) \right]. \tag{1.19}$$

This equation gives an exact solution when the underlying network is a tree. It is easy to get the probability that node i is in the susceptible compartment at time t

$$P(x_i^{t+1} = S) = P(x_i^0 = S) \prod_{j \in \partial i} H^{ji}(t+1). \tag{1.20}$$

Once $P(x_i^{t+1} = S)$ the probability that node i is infected at time t can be obtained by

$$\begin{aligned} P(x_i^{t+1} = I) = P(x_i^t = S) - P(x_i^{t+1} = S) - \left[1 - P(x_i^0 = S)\right] r(t) \\ + \sum_{\tau=1}^{t} r(t-\tau) \left[P(x_i^\tau = S) - P(x_i^{\tau-1} = S)\right] \end{aligned} \tag{1.21}$$

where the first term the contribution due to individual gone from S to I, the second term is the probability tha i does not recover at t being infected at $t = 0$ and the third term accounts for the probability that i gets infected at time $\tau < t+1$ and recovers at $t+1$. Finally the probability that node i is in the recovery state is $P(x_i^{t+1} = R) = 1 - P(x_i^{t+1} = I) - P(x_i^{t+1} = S)$.

This solution is exact on trees, because MP assumes that transmission processes from different branches to one node are not correlated. If the underlying network has loops this assumption is not true. For example (Fig. 1.3) in the case of a single seed i_0, a loop between i_0 and a node j leads to consider two possible paths of infection, but only one has actually transmitted the disease. The error induced by the correlation in loopy graphs decreases when the number of sources increases [30, 41].

However it is known that this kind of message passing method can give a good approximation in loopy graphs as well. In particular Karrer and Newman [21] pointed out that message passing provides a rigorous upper bound to the number of infected individuals on loopy networks. Let us consider the subgraph containing all the n_i nodes that can possibly infect a node i at time t, i.e all the nodes that are distant less then t from node i – where the distance is measured in terms of the sum of contact time w_{ji} drawn by the distribution $s(w_{ji})$. The time of recovery g_j is drawn by the distribution $r(g_j)$. If $w_{ji} > g_j$ no transmission takes place and we set $w_{ji} = \infty$. If any node in the subgraph is infected at time zero, then node i will be

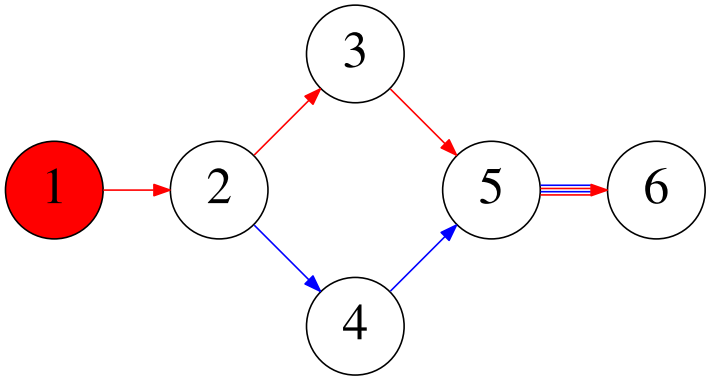


Figure 1.3: When the underlying network has a loopy structure the transmission events to one node from its neighbors are correlated. Node 1 is the seed and two different paths (blue and red) of infection lead to node 5 and 6. In the message passing approach they are considered as coming from different sources but actually only one contributes.

infected at time t. Averaging on the initial conditions $P(x_i^0 = S) = z$ for all i, the probability that i is susceptible

$$P(x_i^t = S) = z\langle z^{n_i}\rangle \tag{1.22}$$

where we averaged over the ensemble of values for w_{ij} and τ_i. This equation is valid in all cases. In order to obtain the relation with the message passing equation (1.19) we consider the number n_{ij} of nodes that are distant less than t from i, but restricted that j is the penultimate node in the path. Paths passing through i are forbidden, as well (they are not allowed epidemic spreading patterns). In this way a node k included in paths related to different j is counted more than once, thus it leads to the inequality: $n_i \leq \sum_{j\in\partial i} n_{ij}$. When the underlying network is a tree no over counting is possible, so it becomes an exact equality. Then

$$P(x_i^t = S) = z\langle z^{n_i}\rangle \geq z\langle z^{\sum_{j\in\partial i} n_{ij}}\rangle = z\langle \prod_{j\in\partial i} z^{n_{ij}}\rangle. \tag{1.23}$$

In the original formulation next steps use the *Chebyshev integral inequality* for non-negative functions $f_i(x_1,\ldots,x_k)$ monotone in every argument. That says

$$\langle\prod_{i=0}^{n} f_i(x_1,\ldots,x_k)\rangle \geq \prod_{i=0}^{n}\langle f_i(x_1,\ldots,x_k)\rangle. \tag{1.24}$$

In our formulation we use a straightforward discrete counterpart fo the Chebyshev integral inequality, also called *Chebyshev sum inequality*. Then equation (1.23) becomes

$$P(x_i^t = S) \geq z\prod_{j\in\partial i}\langle z^{n_{ij}}\rangle. \tag{1.25}$$

The quantity $H^{ji}(t) = \langle z^{n_{ij}}\rangle$ is the probability that at time t the infection has not reached node i through its connection with j. $H^{ji}(t)$ plays the same role than the quantity defined on a tree in (1.19) and it is calculated by the sum of two contributions. When we have to consider the probability that j doesn't get the infection on time to transmit it to i, we have to consider a *cavity configuration*, i.e the network as if node i and its outgoing edges are removed. This is required on a loopy network in order to exclude from the calculation the infection paths of j going through i itself. The cavity probability that j doesn't get the infection before time $t' = t-\tau$ is

$$P^C(S_j(t')) \geq z\prod_{k\in\partial j\setminus i}\langle z^{n_{jk}}\rangle = z\prod_{k\in\partial j\setminus i} H^{j\leftarrow k}(t'), \tag{1.26}$$

where the inequality holds because the cavity configuration does not take into account paths of infection passing through i as a middle node, instead of the term on the right. Combining all the contributions we get

$$H^{ji}(t) \geq 1 - \sum_{\tau=0}^{t} f(\tau)\left[1 - P(x_j^0 = S)\prod_{k\in\partial j\setminus i} H^{kj}(t-\tau)\right]. \tag{1.27}$$

This expression in the form of inequality is not suitable to calculate properties of epidemics. Karrer and Newman [21] showed that if one solves (1.27) with the equality, then a rigorous lower bound for the probability of being susceptible is

achieved

$$P(x_i^t = S) \geq z \prod_{j \in \partial i} H^{ji}(t), \tag{1.28}$$

and as a consequence also an upper bound for the probability $P_i^I(t)$ of being infected.

1.5.1 Equivalence to the Dynamic Message Passing formulation

In the following we will introduce the Dynamic Message Passing formulation for the SIR model proposed by Lokhov, Mézard, Ohta and Zdeborová [29, 42] that presents a set of closed equations to get probabilities at time t given the ones at time $t-1$ and we will show that it is equivalent to the more general approach by Karrer and Newman when it is assumed that the epidemic parameters do not depend on time. The quantity involved in the DMP formulation are: the probability $\theta^{ji}(t)$ that no infection signal passed from j to i until time t; the probability $\phi^{ji}(t)$ that j is infect at time t and the infection has not been transmitted to i up to t. The set of recursion rules is

$$\theta^{ji}(t+1) = \theta^{ji}(t) - \lambda_{ji}\phi^{ji}(t), \tag{1.29}$$

$$\begin{aligned}\phi^{ji}(t) =& (1-\lambda_{ji})(1-\mu_j)\phi^{ji}(t-1), \\ &- \left[P_S^{ji}(t) - P_S^{ji}(t-1)\right]\end{aligned} \tag{1.30}$$

$$P_S^{ji}(t+1) = P(x_i^0 = S) \prod_{j \in \partial k \setminus i} \theta^{kj}(t+1). \tag{1.31}$$

The last equation involves the quantity $P_S^{ji}(t+1)$ that is the probability that node k is in state S when i is fixed in the state S. The update rules start from a set of initial condition that in the case of an epidemic process at time $t = 0$ are

$$\theta^{ji}(0) = 1 \tag{1.32}$$

$$\phi^{ji}(0) = \delta_{q_j(0),I} \tag{1.33}$$

The probability that a node i is in a given state at time t are

$$P(x_i^{t+1} = S) = P(x_i^0 = S) \prod_{j \in \partial i} \theta^{ji}(t+1). \tag{1.34}$$

$$P(x_i^{t+1} = R) = P(x_i^t = R) + \mu_i P(x_i^t = I), \tag{1.35}$$

$$P(x_i^{t+1} = I) = 1 - P(x_i^{t+1} = S) - P(x_i^{t+1} = R). \tag{1.36}$$

In order to show the equivalence between equations (1.31) and equation (1.19), we reformulate equation (1.29) to get a relation in which $\theta^{ji}(t+1)$ is expressed in function of $\theta^{ji}(t)$. Firstly, we use the expression for $\phi^{ji}(t)$ and $\phi^{ji}(t-1)$ from Eq.(1.29) into Eq.(1.30) and we get

$$\frac{\theta^{ji}(t) - \theta^{ji}(t+1)}{\lambda_{ji}} = \frac{(1-\lambda_{ji})(1-\mu_i)}{\lambda_{ji}} \left(\theta^{ji}(t-1) - \theta^{ji}(t)\right) - P_S^{ji}(t) + P_S^{ji}(t-1) \tag{1.37}$$

and reordering the terms

$$\begin{aligned} &\theta^{ji}(t) - (1-\lambda_{ji})(1-\mu_i)\theta^{ji}(t-1) - \lambda_{ji} P_S^{ji}(t-1) = \\ &\quad \theta^{ji}(t+1) - (1-\lambda_{ji})(1-\mu_j)\theta^{ji}(t) - \lambda_{ij} P_S^{ji}(t). \end{aligned} \tag{1.38}$$

Since in this equation the two terms are equal except that they are evaluated at different time steps, it means that they must be equal to a constant C_{ji}. Then we can solve the equation

$$\theta^{ji}(t+1) - (1-\lambda_{ji})(1-\mu_j)\theta^{ji}(t) - \lambda_{ji} P_S^{ji}(t) = C_{ji}, \tag{1.39}$$

with the boundary condition $\theta^{ji}(0) = 1$ given by definition of this quantity (at time $t = 0$ no infection signal has been transmitted). In order to evaluate the constant C_{ji} we consider the case in which all nodes are initially susceptible, i.e $P_S^i(0) = 1$; in this case both $P_S^{ij}(t) = 1$ and $\theta^{ji}(t) = 1$ are valid at any time t, and the above

equation reduces to

$$1-(1-\lambda_{ji})(1-\mu_j)-\lambda_{ji}=C_{ji} \implies C_{ji}=(1-\lambda_{ji})\mu_j.$$

Using this result in (1.39) we find that the cavity quantities $\theta^{ji}(t)$ have to satisfy the dynamic equation

$$\theta^{ji}(t+1)=(1-\lambda_{ji})\left[\mu_j+(1-\mu_j)\theta^{ji}(t)\right]+\lambda_{ji}P_S^i(0)\prod_{k\in\partial j\setminus i}\theta^{kj}(t). \tag{1.40}$$

From this equation it is possible to recover the discrete time message passing equation (1.19) by Karrer and Newman when infection and recovery delays follow geometric distributions. If we write the terms in the right hand side depending on $\theta^{ji}(t)$ as functions of the $\theta^{ji}(t-1)$ we get

$$\begin{aligned}\theta^{ji}(t+1)&=(1-\lambda_{ji})\mu_i+(1-\lambda_{ji})(1-\mu_j)\theta^{ji}(t)+\lambda_{ji}P_S^{ji}(t)\\&=(1-\lambda_{ji})\mu_j+(1-\lambda_{ji})(1-\mu_j)\Big[(1-\lambda_{ji})\mu_j+(1-\lambda_{ji})(1-\mu_j)\theta^{ji}(t-1)+\lambda_{ji}P_S^{ji}(t-\\&\quad+\lambda_{ij}P_S^{ij}(t)\\&=(1-\lambda_{ji})\mu_i+(1-\lambda_{ji})^2(1-\mu_j)\mu_j+(1-\lambda_{ji})^2(1-\mu_j)^2\theta^{ji}(t-1)\\&\quad+\lambda_{ji}(1-\lambda_{ji})(1-\mu_j)P_S^{ji}(t-1)+\lambda_{ji}P_S^{ji}(t),\end{aligned} \tag{1.41}$$

we can repeat the procedure for $\theta^{ji}(t-2)$ and continue $t-2$ times until we reach $\theta^{ji}(0)=1$, therefore the resulting expression is

$$\begin{aligned}\theta^{ji}(t+1)=&\mu_i(1-\lambda_{ij})\sum_{s=0}^{t}(1-\lambda_{ij})^s(1-\mu_i)^s+(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}\\&+\lambda_{ij}\sum_{s=0}^{t}(1-\lambda_{ij})^s(1-\mu_i)^sP_S^i(0)\prod_{k\in\partial i\setminus j}\theta^{ki}(t-s).\end{aligned} \tag{1.42}$$

We can observe that the term

$$\lambda_{ij}\sum_{s=0}^{t}(1-\lambda_{ij})^s(1-\mu_i)^sP_S^i(0)\prod_{k\in\partial i\setminus j}\theta^{ki}(t-s)$$

is the second term in the Karrer-Newman's equation (1.19) when the delays follow geometric distributions. Now we can if we recall the definition in (1.16) $f_{ji}(s) = \lambda_{ji}(1-\lambda_{ji})^s(1-\mu_j)^s$, we get the exact expression for the first term of equation (1.19). Let us consider the first two terms

$$
\begin{aligned}
&\mu_i(1-\lambda_{ij})\sum_{s=0}^{t}(1-\lambda_{ij})^s(1-\mu_i)^s + (1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1} \\
&= \mu_i(1-\lambda_{ij})\left[\frac{1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}}{1-(1-\lambda_{ij})(1-\mu_i)}\right] + (1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1} \\
&= \big(\lambda_{ij}+\mu_i(1-\lambda_{ij})\big)\left[\frac{1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}}{1-(1-\lambda_{ij})(1-\mu_i)}\right] \\
&\quad + (1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1} - \lambda_{ij}\left[\frac{1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}}{1-(1-\lambda_{ij})(1-\mu_i)}\right] \\
&= 1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1} + (1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1} - \lambda_{ij}\left[\frac{1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}}{1-(1-\lambda_{ij})(1-\mu_i)}\right] \\
&= 1-\lambda_{ij}\left[\frac{1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}}{1-(1-\lambda_{ij})(1-\mu_i)}\right].
\end{aligned}
$$

The result is

$$
\begin{aligned}
\theta^{ji}(t+1) =& 1-\lambda_{ij}\left[\frac{1-(1-\lambda_{ij})^{t+1}(1-\mu_i)^{t+1}}{1-(1-\lambda_{ij})(1-\mu_i)}\right] + \\
&+\lambda_{ij}\sum_{s=0}^{t}(1-\lambda_{ij})^s(1-\mu_i)^s P_S^i(0)\prod_{k\in\partial i\setminus j}\theta^{ki}(t-s). \qquad (1.43)
\end{aligned}
$$

, we recover the first term in the equation by Karrer and Newman:

$$
1-\sum_{s=0}^{t} f_{ji}(s) \;=\; 1-\lambda_{ji}\frac{1-(1-\lambda_{ji})^{t+1}(1-\mu_j)^{t+1}}{1-(1-\lambda_{ji})(1-\mu_j)}. \qquad (1.44)
$$

Finally we get

$$
\theta^{ji}(t+1) = 1-\sum_{s=0}^{t} f_{ji}(s) + \sum_{s=0}^{t} f_{ji}(s)P_S^i(0)\prod_{k\in\partial i\setminus j}\theta^{kj}(t-s), \qquad (1.45)
$$

that is the same expression of equation (1.19). To complete the equivalence we show that (1.19) can naturally be written in the form (1.40), in fact

$$
\begin{aligned}
H^{ji}(t+1) =& 1-\lambda_{ji}+\sum_{\tau=1}^{t}\lambda_{ji}(1-\lambda_{ji})^{s}(1-\mu_j)^{s}+\lambda_{ji}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t)\\
&+\sum_{\tau=1}^{t}\lambda_{ji}(1-\lambda_{ji})^{s}(1-\mu_j)^{s}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t-\tau)\\
&=1-\lambda_{ji}+\sum_{\tau=0}^{t-1}(1-\lambda_{ji})^{s+1}(1-\mu_j)^{s+1}+\lambda_{ji}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t)\\
&+\sum_{\tau=0}^{t}\lambda(1-\lambda_{ji})^{s+1}(1-\mu_j)^{s+1}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t-\tau)\\
&=1-\lambda_{ji}+(1-\lambda_{ji})(1-\mu_j)\sum_{\tau=0}^{t-1}(1-\lambda_{ji})^{s+1}(1-\mu_j)^{s+1}+\lambda_{ji}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t)\\
&+(1-\lambda_{ji})(1-\mu_j)\sum_{\tau=0}^{t}\lambda_{ji}(1-\lambda_{ji})^{s}(1-\mu_i)^{s}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t-\tau-1)\\
&1-\lambda_{ji}-(1-\lambda_{ji})(1-\mu_j)+(1-\lambda_{ji})(1-\mu_j)H^{ji}(t)+\lambda_{ji}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t)\\
&(1-\lambda_{ji})\left[\mu_j+(1-\mu_j)H^{ji}(t)\right]+\lambda_{ji}P_S^j(0)\prod_{k\in\partial j\setminus i}H^{kj}(t).
\end{aligned}
$$

1.5.2 Equivalence to the pair-based approximation

In Section 1.4 we introduced a pair-based approximation that leads to a set of closed equation for the time evolution of the probabilities of the states of the individuals. Here we show that the pair-based approximation can be recovered from the Dynamic Message Passing approach. We use the formalism by Lokhov, Mézard, Ohta and Zdeborová to express the relevant quantities in equations (1.13) in function of cavity marginals. Then the marginal probability that node i is susceptible at time t is:

$$P(x_i^t = S) = P(x_i^0 = S)\prod_{j\in\partial i}\theta^{ji}(t). \tag{1.46}$$

The joint probability $P(x_i^t = S, x_j^t = S)$ is

$$P(x_i^t = S, x_j^t = S) = P(x_i^0 = S)P(x_j^0 = S) \prod_{k \in \partial i \setminus j} \theta^{ki}(t) \prod_{l \in \partial j \setminus i} \theta^{lj}(t), \tag{1.47}$$

which follows from the fact that both i and j cannot infect each other. For the joint probability $P(x_i^t = S, x_j^t = I)$ we must consider the probability that i has never get the disease multiplied by the probability that j is infected at time t but it has not transmitted the infection to i, that is

$$P(x_i^t = S, x_j^t = I) = P(x_i^0 = S) \left[\prod_{k \in \partial i \setminus j} \theta^{ki}(t) \right] \phi^{ji}(t) \tag{1.48}$$

$$= P(x_i^0 = S) \prod_{k \in \partial i \setminus j} \theta^{ki}(t) \left[\theta^{ji}(t) - \frac{1 - \lambda_{ji}}{\lambda_{ji}} \mu_j \left(1 - \theta^{ji}(t)\right) - P(x_j^0 = S) \prod_{h \in \partial j \setminus i} \theta^{hj}(t) \right], \tag{1.49}$$

where the last expression is obtained using (1.40) in equation (1.29). Now we can derive equations (1.13) starting from the Dynamic Message Passing formalism. The probability that a node i is susceptible at time t is

$$\begin{aligned} P(x_i^{t+1} = S) =& P(x_i^0 = S) \prod_{k \in \partial i} \theta^{ki}(t+1) = P(x_i^0 = S) \prod_{k \in \partial i} \left[\theta^{ki}(t) - \lambda_{ki} \phi^{ki}(t) \right] \\ &= P(x_i^0 = S) \prod_{k \in \partial i} \theta^{ki}(t) \left[1 - \frac{\lambda_{ki} \phi^{ki}(t)}{\theta^{ki}(t)} \right], \end{aligned} \tag{1.50}$$

where we use (1.49) obtaining

$$\begin{aligned} P(x_i^{t+1} = S) =& P(x_i^0 = S) \prod_{k \in \partial i} \theta^{ki}(t) \left[1 - \frac{\lambda_{ki} P(x_i^t = S, x_k^t = I)}{\theta^{ki}(t) P(x_i^0 = S) \left[\prod_{v \in \partial i \setminus k} \theta^{vi}(t) \right]} \right] \\ &= P(x_i^t = S) \prod_{k \in \partial i} \left[1 - \frac{\lambda_{ki} P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)} \right], \end{aligned} \tag{1.51}$$

that is the same expression found in (1.13) for the probability that a node is susceptible at time $t+1$. The probability that a node is infected is given by

equation (1.21) (where we used a constant probability of recovery)

$$\begin{aligned} P(x_i^{t+1} = I) &= (1-\mu_i)P(x_i^t = I) + P(x_i^t = S) - P(x_i^{t+1} = S) \\ &= (1-\mu_i)P(x_i^t = I) + P(x_i^t = S)\left[1 - \prod_{k\in\partial i}\left(1 - \frac{\lambda_{ki}P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right)\right]. \end{aligned} \tag{1.52}$$

Then the equation that describes the evolution of the probability that a node i is in the recovered state is

$$P(x_i^{t+1} = R) = P(x_i^t = R) + \mu_i P(x_i^t = I). \tag{1.53}$$

Now we need the time evolution for the joint probability $P(x_i^t = S, x_k^t = I)$ present in (1.52) and it is given by

$$
\begin{aligned}
P(x_i^{t+1} = S, x_j^{t+1} = I) &= P(x_i^0 = S)\left[\prod_{k\in\partial i\setminus j}\theta^{ki}(t+1)\right]\phi^{ji}(t+1)\\
&= P(x_i^0 = S)\prod_{k\in\partial i\setminus j}\left[\theta^{ki}(t) - \lambda_{ki}\phi^{ki}(t)\right]\times\\
&\times\left[(1-\lambda_{ji})(1-\mu_j)\phi^{ji}(t) - P(x_j^0 = S)\prod_{h\in\partial j\setminus i}\theta^{hj}(t+1) + P(x_j^0 = S)\prod_{v\in\partial j\setminus i}\theta^{vj}(t)\right]\\
&= P(x_i^0 = S)\prod_{k\in\partial i\setminus j}\theta^{ki}(t)\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right](1-\lambda_{ji})(1-\mu_j)\phi^{ji}(t)\\
&- P(x_i^0 = S)\prod_{k\in\partial i\setminus j}\theta^{ki}(t)\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]P(x_j^0 = S)\prod_{h\in\partial j\setminus i}\theta^{hj}(t+1)\\
&+ P(x_j^0 = S)\prod_{v\in\partial j\setminus i}\theta^{vj}(t)P(x_i^0 = S)\prod_{k\in\partial i\setminus j}\theta^{ki}(t)\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]\\
&= (1-\lambda_{ji})(1-\mu_j)P(x_i^t = S, x_j^t = I)\prod_{k\in\partial i\setminus j}\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]\\
&- P(x_i^t = S, x_j^t = S)\prod_{k\in\partial i\setminus j}\theta^{ki}(t)\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]\times\\
&\times\prod_{h\in\partial j\setminus i}\theta^{hj}(t)\left[1 - \lambda_{hj}\frac{P(x_j^t = S, x_h^t = I)}{P(x_j^t = S)}\right]\\
&+ P(x_i^t = S, x_j^t = S)\prod_{k\in\partial i\setminus j}\theta^{ki}(t)\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]\\
&= (1-\lambda_{ji})(1-\mu_j)P(x_i^t = S, x_j^t = I)\prod_{k\in\partial i\setminus j}\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]\\
&+ P(x_i^t = S, x_j^t = S)\left[1 - \prod_{h\in\partial j\setminus i}\theta^{hj}(t)\left[1 - \lambda_{hj}\frac{P(x_j^t = S, x_h^t = I)}{P(x_j^t = S)}\right]\right]\times\\
&\times\prod_{k\in\partial i\setminus j}\theta^{ki}(t)\left[1 - \lambda_{ki}\frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)}\right]. \qquad (1.54)
\end{aligned}
$$

Then we write the expression for the time evolution of the joint probability $P(x_i^t = S, x_j^t = S)$, that is given by

$$\begin{aligned} P(x_i^{t+1} = S, x_j^{t+1} = S) &= P(x_i^0 = S) \left[\prod_{k \in \partial i \setminus j} \theta^{ki}(t+1) \right] P(x_j^0 = S) \left[\prod_{h \in \partial j \setminus i} \theta^{hj}(t+1) \right] \\ &= P(x_i^0 = S) \prod_{k \in \partial i \setminus j} \left[\theta^{ki}(t) - \lambda_{ki} \phi^{ki}(t) \right] \\ &\times P(x_j^0 = S) \prod_{h \in \partial j \setminus i} \left[\theta^{hi}(t) - \lambda_{hi} \phi^{hi}(t) \right], \end{aligned} \tag{1.55}$$

using the expression (1.49) and reordering the terms

$$\begin{aligned} P(x_i^{t+1} = S, x_j^{t+1} = S) &= P(x_i^0 = S) \prod_{k \in \partial i \setminus j} \theta^{ki}(t) \left[1 - \lambda_{ki} \frac{P(x_i^t = S, x_k^t = I)}{\prod_{v \in \partial i \setminus k} \theta^{vi}(t) \theta^{ki}(t)} \right] \\ &\times P(x_j^0 = S) \prod_{h \in \partial j \setminus i} \theta^{hi}(t) \left[1 - \lambda_{hi} \frac{P(x_j^t = S, x_h^t = I)}{\prod_{u \in \partial j \setminus h} \theta^{uj}(t) \theta^{hj}(t)} \right] \\ = P(x_i^t = S, x_j^t = S) &\prod_{k \in \partial i \setminus j} \left[1 - \lambda_{ki} \frac{P(x_i^t = S, x_k^t = I)}{P(x_i^t = S)} \right] \prod_{h \in \partial j \setminus i} \left[1 - \lambda_{hi} \frac{P(x_i^t = S, x_h^t = I)}{P(x_j^t = S)} \right]. \end{aligned} \tag{1.56}$$

We expressed the probability distribution evolution over time in term of cavity messages. Then, by the use of their evolution equations, we have recovered equations (1.13), (1.14) and (1.15) obtained by considering the master equation of the process in a pair-based approximation. Thus we proved that the two approaches are indeed two equivalent way to arrange the terms that contribute to the probability distributions.

1.5.3 Towards a new representation

In the message passing formalisms that we presented, each message is function of a single cavity time variable t_{ij}, i.e the time t_{ij} at which i get infected when j is fixed in the susceptible state. They correctly deal with the direct problem of the epidemic spreading: given the initial conditions they allows to evaluate the marginal probability of being susceptible, infected or recovery for every node.

When considering an inverse problem, such as the patient zero identification, these approaches need further approximations, such as a mean-field approximation (see Section 2), that may lead to errors in the inference of the source. Instead, it is possible to tackle inverse problems by a different approach. We compute a partition function in which we trace over the cavity infection times t_{ij}, on which constraints are imposed by the dynamics of the process Φ and by the final observed configuration. However the two constraints depend both on t_{ij} and t_{ji}, that means that the forward dynamics as well as the backward propagation of the information has to be taken into account. Therefore, we need a representation of the messages $m_{ji}(t_{ji}, g_j, t_{ij}, g_i)$ that depends both on the cavity infection time t_{ij} and t_{ji}. The constraint on the dynamics will be a function of t_{ij} and of the infection time of the nodes in $\partial i \setminus j$ and it will be enforced on every edge of the network:

$$\prod_{(i,j)\in E} \Psi_{ij}(t_{ij}, \{t_{ki}, s_{ki}\}_{k\in\partial i\setminus j}) = \prod_{(i,j)\in E} \delta\left[t_{ij} = \min_{k\in\partial i\setminus j}\{t_{ki} + s_{ki}\}\right]. \tag{1.57}$$

When the constraint on the dynamics Φ is considered, the partition function is

$$Z \propto \sum_{\{g_i\}_{i\in N}} \sum_{\{t_{ij}, s_{ij}\}_{(i,j)\in E}} \prod_{i\in N} r_i(g_i) \prod_{(i,j)\in E} \omega(s_{ij}|g_j)\delta\left[t_{ij} = \min_{\ell\in\partial i\setminus j}\{t_{\ell' i} + s_{\ell' i}\}\right]. \tag{1.58}$$

The Belief Propagation equation for the messages is:

$$m_{ij}(t_{ij}, g_i, t_{ji}, g_j) \propto \sum_{\{t_{ki}, g_k, s_{ki}\}_{k\in\partial i\setminus j}} r_i(g_i) \prod_{k\in\partial i\setminus j} \omega(s_{ki}|g_k) \prod_{\ell\in\partial i} \delta\left[t_{i\ell} = \min_{\ell'\in\partial i\setminus\ell}\{t_{\ell' i} + s_{\ell' i}\}\right] \prod_{k\in\partial i\setminus j} m_{ki}(t_{ki}, g_k, t_{ik}, g_i), \tag{1.59}$$

where $\omega(s_{ki}|g_k)$ is the probability distribution of the infection delay s_{ki} conditioned on the recovery delay g_k.

However, the computation of quantities involving this term becomes infeasible as the nodes degree increases. Therefore, we will introduce a computationally tractable representation in which the constraint Ψ_{ij} depending on the cavity infection time is replaced by a function on the total infection time t_i and on the neighbors infection time $\{t_k\}_{k\in\partial i\setminus j}$: this function is evaluated on each node rather than on each edge. Moreover, the new representation allows to conveniently treat

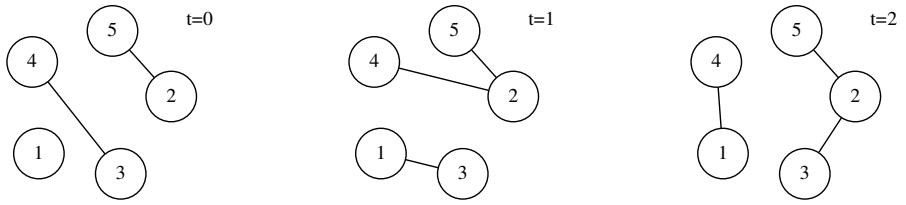


Figure 1.4: Temporal network: a set of nodes that at every time step is connected by a different set of edges. Considering the set of edges occurred before a given time, one gets a static projection of the network.

the constraint on the final observed configuration. We will present this approach in Chapter 4.

1.6 Epidemic processes in temporal networks

An intermediate scenario exists, in fact in many realistic cases contacts and the epidemics evolve on a comparable time scale. Thus we define a *temporal network* [43, 44] as a set of nodes that at every instant are connected by a different set of edges. A *contact sequence* is the collection of the edges existing at every time step taken under consideration. We can recover a static projection of the network until time T by aggregating the contact sequence for $t < T$, thus obtaining a network containing every edge existed before T. A more informative representation can be achieved by assigning a weight to every edge between nodes i and j proportional to the total amount of time they have been in contact. On a temporal networks the time ordering of connections affects epidemic processes. In fact an infected node can propagate the disease only through edges existing during its infectious period and the propagation into the population must follow paths respecting the time ordering. In most cases the consequence is an epidemic size smaller than in static networks. Data about dynamical human contacts can be collected in many ways. For example, socio-technological networks can provide large datasets that cover a large time interval (as sexual encounters websites [45]). Instead networks of wireless sensors [11, 12] correspond to high resolution data about spatial proximity

between individuals, but they involve smaller populations and, often, confined space.

1.7 Non Markovian epidemic model

In the framework of epidemic modeling it is usually assumed that the spread of a disease can be described as a discrete time Markov process, i.e that the probability per unit of time of infection or recovery has a constant value. The time that an individual spends in the infected state, as well as the time to successfully transmit infection, has an exponential probability distribution. Since the rate of infection and recovery do not depend on the history of individuals, it is possible to describe the epidemics as a Markov process with no memory and take advantage of an (possibly) increased mathematical tractability [46]. However empirical evidence leads to consider different epidemic models that often cannot be described as a Markov process (or, in continuous time, by a Poisson process). For example, in the case of HIV the distribution of infective periods is poorly approximated by an exponential function [47]. For the smallpox the residence period in the various stages of the disease is rather approximated by a gamma distribution [48, 49]

$$f(t,\alpha,\beta) = \frac{\beta^{\alpha} t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}. \tag{1.60}$$

A SIR model with a fixed length for the infective period well approximates an outbreak of measles [50]. Conversely in the case of the virus Ebola epidemiologists often assume the infective period distributed following a Weibull distribution [51]

$$g(t,k,b) = bkt^{k-1} e^{-bt^{k}}. \tag{1.61}$$

In general few analytical results are available for non-Markovian epidemic models [15], because the infection process does not depend only on the infected (or not) neighbors of a node but also on their history. It is worth noting that the message passing approach to SIR by Karrer and Newman [21] (described in Section 1.5) assumes generic probability distribution for the infection and recovery time, then it can be used to study non-Markovian epidemic models.

Chapter 2

Source inference for SIR on networks

In this chapter we introduce different methods to infer the origin of an epidemic process given an observation of the system. They are based on three different approaches: *rumor centrality* exploits the topological properties of spreading patterns on networks; the *Soft-Margin* estimator is based on extensive Monte Carlo simulations; the *Dynamic Message Passing – Naive Bayes* computes exact marginal probability on trees and relies on mean-field approximation to infer the source. Nevertheless the inverse problem of finding the origin of a stochastic process such as an SIR epidemic is not trivial and these methods have drawbacks that will be discussed in this chapter.

2.1 Rumor Centrality

In order to propose a maximum likelihood estimator for the epidemic source D. Shah and T. Zaman [25, 27, 28] began from the heuristic assumption that the source of an epidemic spreading is the *center* of a network composed by the nodes reached by the disease. Then one can define an infection path as a sequence of infected nodes ordered according to the constraints set by the underlying contact network and by the epidemic process. The problem of finding the patient-zero is set as follows. Given a graph $G(V,E)$, a subgraph $G_N(V,E)$ composed of the nodes reached by the epidemic is provided. Assuming that the a priori probability

of being the origin i_0 is the same for each node i, the Maximum-Likelihood (ML) estimator is

$$\tilde{i}_0 = \arg\max_{i \in G_N} \mathscr{P}(G_N | i_0 = i), \tag{2.1}$$

where $\mathscr{P}(G_N|i_0 = i)$ is the likelihood of an infection graph $G_N(V,E)$ given a candidate source i. It can be calculated by the sum of the probability of every permitted infection sequence $\overline{\sigma} \in G_N$ originated from i

$$\mathscr{P}(G_N|i_0 = i) = \sum_{\overline{\sigma} \in G_N} \mathscr{P}(\overline{\sigma}|i_0 = i) \tag{2.2}$$

In general the probability of each permitted infection sequence is different, but in the case of regular tree they are all equal due to the constant degree of the nodes and the memoryless Markovian nature of the spreading. Therefore $\mathscr{P}(G_N|i_0 = i)$ is proportional to the number of permitted infection sequence originated in i and resulting in G_N. It is possible to define the notion of *rumor centrality* $R(i,T)$ [25, 27, 28] as the number of infection paths that begin with node i and result in a regular tree T. Thus the problem can be recast in the computation of the rumor centrality [25]

$$\tilde{i}_0 = \arg\max_{i \in G_N} \mathscr{P}(G_N|i_0 = i) = \arg\max_{i \in G_N} R(i, G_N). \tag{2.3}$$

In general this quantity involves an exponential number of terms, but on regular tree it can be computed in linear time as follows. We call T_i^k the subtree that belongs to G_N with k as source rooted in i and $|T_i^k|$ the number of its nodes. The reasoning put forward clarifies how we can count the number of infection paths in G_N given the epidemic origin i_0. The contagion first spreads to the origin's neighbors, then to the subtrees $T_k^{i_0}$ rooted in each $k \in \partial i_0$. The size of G_N (i.e the number of infected individuals) is N and sets the number of node in the infection sequences whose first element is i_0. From the remaining $N-1$ we choose $|T_{k_1}^{i_0}|$ elements for the first neighbor $k_1 \in \partial i_0$ and they can be ordered in $R(k_1, T_{k_1}^{i_0})$ ways. Then from the remaining $N-1-|T_{k_1}^{i_0}|$ elements we choose $|T_{k_2}^{i_0}|$ that can be ordered in $R(k_2, T_{k_2}^{i_0})$ ways. It is worth noting that the ordering is constrained only for nodes in the same

subtree, since the spreading in each branch is independent. Then we obtain

$$R(i_0, G_N) = \binom{N-1}{|T_{k_1}^{i_0}|} \binom{N-1-|T_{k_1}^{i_0}|}{|T_{k_2}^{i_0}|} \cdots \binom{N-1-\sum_{i=1}^{|\partial i_0|-1}|T_{k_i}^{i_0}|}{T_{k_{|\partial i_0|}}^{i_0}} \prod_{k\in\partial i_0} R(k, T_k^{i_0})$$
$$= (N-1)! \prod_{k\in\partial i_0} \frac{R(k, T_k^{i_0})}{|T_k^{i_0}|!} \tag{2.4}$$

By expanding this calculation in terms of subtrees $T_l^{i_0}$ rooted at nearest neighbor children l of $k \in \partial i_0$ we get

$$\begin{aligned} R(i_0, G_N) =& (N-1)! \prod_{k\in\partial i_0} \frac{R(k, T_k^{i_0})}{|T_k^{i_0}|!} = \\ &= (N-1)! \prod_{k\in\partial i_0} \frac{(|T_k^{i_0}|-1)!}{|T_k^{i_0}|!} \prod_{l\in\partial k} \frac{R(l, T_l^{i_0})}{|T_l^{i_0}|!} = \\ &= (N-1)! \prod_{k\in\partial i_0} \frac{1}{|T_k^{i_0}|} \prod_{l\in\partial k} \frac{R(l, T_l^{i_0})}{|T_l^{i_0}|!}. \end{aligned} \tag{2.5}$$

This procedure is repeated until the leaves of G_N are reached [25] and exploiting the fact that the tree rooted in i_0 has $|T_{i_0}^{i_0}| = N$ we get

$$R(i_0, G_N) = N! \prod_{i\in G_N} \frac{1}{|T_i^{i_0}|}. \tag{2.6}$$

The rumor center is found by computing $R(i_0, G_N)$ for every node in G_N in order to find its likelihood to be the origin of the epidemic. However to compute this quantity we need the size $|T_i^{i_0}|$ of every subtree rooted in every node i for every source candidate i_0, that means N^2 subtrees. Shah and Zaman proposed an algorithm to calculate rumor centrality for every node with $O(N)$ computations [25] that we present in the following. As a first step consider two neighboring nodes i and j, all their subtrees have the same size except for the ones rooted in j and i respectively, i.e T_j^i and T_i^j. Between this two subtrees the following relation is valid for the size

$$T_j^i = N - T_i^j. \tag{2.7}$$

If one inserts this expression in equation (2.6) then finds an equivalent relation between the rumor centralities of two neighboring nodes i and j

$$R(i,G_N) = N! \prod_{k\in G_N} \frac{1}{|T_k^i|} = N! \prod_{k\in G_N\setminus j} \frac{1}{|T_k^i|}\frac{1}{|T_j^i|}$$
$$= R(j,G_N)|T_i^j|\frac{1}{|T_j^i|} = R(j,G_N)\frac{|T_i^j|}{N-|T_i^j|}. \tag{2.8}$$

In this scheme the rumor centrality $R(i_0,G_N)$ of a source candidate i_0 is obtained by multiplying the cumulative products of its children subtrees size. The latter quantity is computed starting at the leaves (where the subtree size is one) and recursively transmitting upward the information about cumulative products and size of the subtree to parents nodes. Once we obtain the rumor centrality for i_0, we can calculate the rumor centrality for each node starting from $k \in \partial i_0$ using equation (2.8).

Shah and Zaman [25, 27, 28] showed that on a regular tree with degree $d > 2$ the probability of a correct detection of the source is always greater than zero. Instead in the case of a line (a regular tree with $d = 2$) this probability scales with time t as $O(t^{1/2})$. This calculation is exact on regular trees, but it is not on general trees. In fact if the node degrees vary then the infection sequences are not equally probable. In this case it is only possible to have $P(i_0 = i|G_N) \propto R(i,G_N)$. In the case of general graphs a simplified approximated approach based on heuristic assumptions is possible. The basic assumption is that given a source i_0 the epidemic spreads on a breadth first search tree. So we can use this trees $T_{bfs}^{i_0}$ instead of the general graph G_N and find the source using the modified maximum likelihood estimator

$$\tilde{i}_0 = \arg\max_{i\in G_N} R(i, T_{bfs}(i)). \tag{2.9}$$

2.2 Soft-Margin Estimator

The problem of the identification of the patient zero can be addressed by the use of Monte Carlo methods. Let us consider the case in which we observe the configuration $\mathbf{x}_*^t$ of a realization of the SIR model at time t, we can represent

this configuration by the graph $G_N(V,E)$ of the infected and recovery nodes at time t (equivalent to the nodes that got infected until time t). Given an observed realization G_N* an immediate approach to compute $P(x_i^0 = I|G_N)$ is a *direct Monte Carlo approach*. We can simulate a large number n of realizations of the process for each potential source i (i.e. that has to be among the nodes that has been infected before t). From the number M_i of simulations that provide a configuration at time t coincident with G_N, we compute the posterior probability

$$P(x_i^0 = I|G_N) = \frac{M_i}{\sum_j M_j}. \tag{2.10}$$

However the number of possible configurations $\mathbf{x}^t$ and G_N is exponential with the size of the network because it is based on the sampling of rare events (a realization at time t), thus the direct Monte Carlo method becomes computationally demanding for large networks. An approximate probability distribution $P(x_i^0 = I|G_N)$ can be computed by using the *Soft-Margin estimator* [24]. This is a generalization of the direct Monte Carlo method that weights the simulated realizations on the basis of their similarity to the observed one G_N. The similarity between a generic realization G_N^i with source node i and the real one G_N is measured by computing the Jaccard similarity function

$$\phi(G_N, G_N^i) = \frac{|G_N \cap G_N^i|}{|G_N \cup G_N^i|}. \tag{2.11}$$

This function takes value in $[0,1]$ and increases with the size of the infected subpopulation shared between the realizations G_N and G_N^i. The Gaussian weight function is

$$w_a(\phi(G_N, G_N^i)) = \exp\left(\frac{-(\phi(G_N, G_N^i) - 1)^2}{a^2}\right). \tag{2.12}$$

By simulating M realizations G_N^{il} with $l \in (1,\ldots,M)$ it is possible to compute the likelihood of G_N given that the source is i (see [24])

$$P(G_N|\mathbf{x}_i^0 = I) = \frac{1}{M}\sum_{l=1}^{M} \exp\left(\frac{-(\phi(G_N, G_N^{il}) - 1)^2}{a^2}\right). \tag{2.13}$$

If we assume that the prior probability $P(\mathbf{x}_i^0 = I)$ to be the origin of the epidemic is the same for every node, then by the use of the Bayes' theorem the posterior probability for a node to be the source is

$$P(\mathbf{x}_i^0 = I|G_N) = \frac{P(G_N|\mathbf{x}_i^0 = I)}{\sum_j P(G_N|\mathbf{x}_j^0 = I)}. \tag{2.14}$$

This approximation computes the posterior probability accepting contributions from realizations that do not yield the observed configuration $\mathbf{x}_*^t$, but they are within a certain distance from $\mathbf{x}_*^t$. A larger value for the standard deviation a will produce a poorer approximation, instead when $a \to 0$ the direct Monte Carlo method is recovered.

2.3 Dynamic Message Passing – Naive Bayes

In their work Lokhov, Mézard, Ohta and Zdeborová [29] proposed to infer the origin of the epidemic by a mean-field approximation of the conditional probability $P(\mathbf{x}^{T_{obs}}|i_0)$ of finding the observed configuration given the epidemic source, that leads to the following factorized form for the conditional probability

$$P(\mathbf{x}^{T_{obs}}|i_0) = \prod_{k/x_k^{T_{obs}}=S} P_S^k(T_{obs}, i_0) \prod_{l/x_l^{T_{obs}}=I} P_S^l(T_{obs}, i_0) \prod_{m/x_m^{T_{obs}}=R} P_S^m(T_{obs}, i_0). \tag{2.15}$$

Then by the use of Bayes' theorem it is possible to find the probability that a node i is the source given the observation $P(i_0 = i|\mathbf{x}^{T_{obs}}) \propto P(\mathbf{x}^{T_{obs}}|i_0 = i)$. For each candidate to be the origin an energy function $E(i) = -\log(P(\mathbf{x}^{T_{obs}}|i_0))$ is defined and the most probable seed is the node with the lowest energy. In order to compute marginal probabilities that a node i is in each of the three states given an epidemic origin i_0 they used the Dynamic Message Passing algorithm [29, 42] described in Section 1.5. The mean-field approximation underlying equation (2.15) leads to significant inaccuracies even in the case of a tree, when the message passing gives the exact value for the marginals. In fact, the factorization over single nodes disregards the information that the observed state $\mathbf{x}^{Tobs}$ gives about correlations. A clarifying example is given by Fig. 2.1 [32]. Given the observed configuration where node 5 and leaves 6-10 are in the susceptible state,

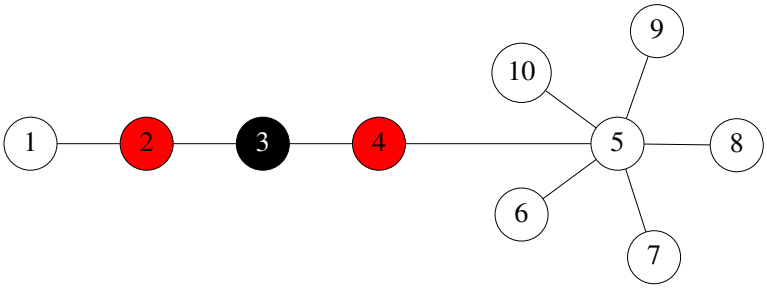


Figure 2.1: This example shows a typical situation in which even if DMP gives exact marginal probabilities the mean-field approximation introduces errors in the inference of the source. Red: node infected; Black: node recovered; White: node susceptible.

the probability that the leaves have not been infected is one. In fact the states of these nodes are correlated, thus if node 5 is susceptible it forces leaves 6-10 to be the same because the infection cannot reach the leaves. Nevertheless in the mean-field approximation (2.15) the leaves 6-10 contributes to the energy function, because their marginals $P_S(T_{obs}, i_0)$ to be susceptible is not identically one. In fact disregarding their correlations with the state of node 5 the resultant probability that the leaves get infected is not zero and it is larger as the candidate seed i_0 is closer to the leaves. Therefore minimizing the mean-field free-energy the inferred seed will be positioned at larger distance from the leaves than the real one. As a consequence if the number of leaves is large enough their contributes dominate the energy function leading to a most probable seed as far as possible from them.

Chapter 3

Belief Propagation

In this chapter we first introduce the factor graph representation for dependencies among random variables. Then we introduce the Belief Propagation, a message passing algorithm that can be used to compute exact marginal probabilities on trees. We will finally show that the Belief Propagation equations fixed points can be cast into the Bethe free energy stationary points.

3.1 Factor Graph Representation

When dealing with random variables with mutual interaction, we can express their dependencies in a factorized form, i.e as product over terms depending only on subsets of variables. In these cases, the factor graph language is a powerful tool in order to represent graphically the structure of dependencies among random variables and we speak of *graphical models*. Let us consider a set $[N]$ of N variables $(x_1,\dots,x_N)$ taking values in a set χ, we assume that their joint probability distribution reads

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{a=1}^{M} F_a(\mathbf{x}_{\partial a}), \tag{3.1}$$

where $\mathbf{x} \equiv \{x_1,\dots,x_N\}$, $\mathbf{x}_{\partial a} \equiv \{x_i \mid i \in \partial a\}$ and $\partial a \subseteq [N]$. The set of indexes ∂a, with $a \in [M]$, has size $k_a \equiv |\partial a|$. The compatibility functions $F_a : \chi^{k_a} \to \mathbb{R}$ are non negative and Z is a positive normalization constant. The factor graph representation of a probability distribution is made up of a bipartite graph composed of factor

nodes and variable nodes [52]: each of the M functions F_a in 3.1 is represented by a factor node and each of the N variable x_i of the problem is represented by a variable node. An edge links a factor node a and a variable node i if $F_a(x_{\partial a})$ depends on x_i.

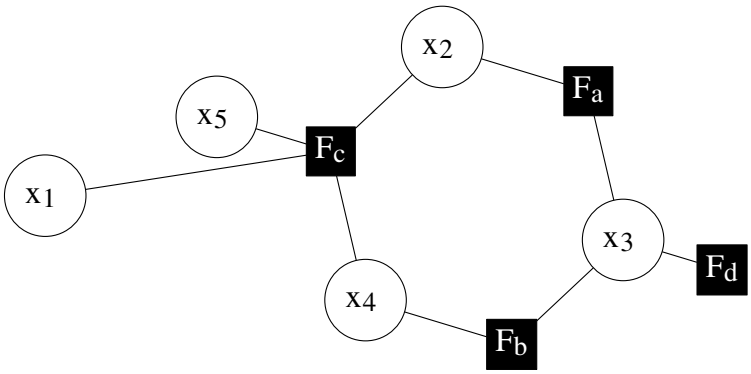


Figure 3.1: A factor graph. Circles correspond to variables $x_i \in (x_1, x_2, x_3, x_4, x_5)$. Black squares correspond to functions encoding dependencies among variables $F_a(x_2, x_3), F_b(x_3, x_4), F_c(x_1, x_2, x_4, x_5), F_d(x_3)$.

A probability distribution written in the factorized form (3.1) respects a conditional independence assumption – also called *global Markov property* [52]:

Proposition 1. *Consider three disjoint subset of variable nodes $A, B, S \subseteq [N]$ and denote by $\mathbf{x}_A, \mathbf{x}_B$ and $\mathbf{x}_S$ the corresponding set of variables. The variables $\mathbf{x}_A, \mathbf{x}_B$ are said conditionally independent given S if*

$$P(\mathbf{x}_A, \mathbf{x}_B|\mathbf{x}_S) = P(\mathbf{x}_A|\mathbf{x}_S)\, P(\mathbf{x}_B|\mathbf{x}_S)\,. \tag{3.2}$$

In the factor graph language the conditional independence sets a notion of locality: in fact if the variables sets A and B are conditionally independent given S, it means that there is no path on the factor graph joining a node in A to a node in B without passing through S. The inverse relation between factor graph representation and conditional independence is encoded in the *Hammersley-Clifford Theorem*:

Theorem 1 (Hammersley-Clifford)**.** *Let P be a strictly positive probability distribution over the variables* $\mathbf{x} = (x_1 \dots x_N) \in \chi^N$ *that satisfies the global Markov property. Then P can be written in the factorized form* (3.1) *with respect to a factor graph F.*

It is worth noting that we do not need conditions on the form of functions $F_a(\mathbf{x}_{\partial a})$ other than the respect of mutual dependencies between variables, so they are not defined in a unique way. It simply leads to different factorized expressions for the joint probability and different factor graphs. It is always possible to choose a convenient form for the factor nodes in order to get a factorized representation as in (3.1). The only assumption behind the factorized form in (3.1) is the notion of locality that comes with the global Markov property.

Example: the Ising chain. As a simple, but explicative, example of factor graph representation for a probability distribution, we consider the ferromagnetic one dimensional Ising model. The variables are Ising spins $(x_1, \dots, x_N) = \mathbf{x}$, with $x_i \in \{+1, -1\}$. Their joint probability distribution take the form

$$P_\beta(\mathbf{x}) = \frac{1}{Z} e^{-\beta E(\mathbf{x})}, \qquad E(\mathbf{x}) = -\sum_{i=1}^{N-1} x_i x_{i+1} - B \sum_{i=1}^{N} x_i \tag{3.3}$$

The global Markov property is satisfied. In the factor graph the variable nodes represent the Ising spins $\mathbf{x}$. There are two types of factor nodes. Nodes corresponding to the $N-1$ pairwise interactions $F_a(x_i, x_{i+1}) = e^{-\beta x_i x_{i+1}}$ are connected to two neighboring variables nodes. Nodes corresponding to the N magnetic field terms $\phi_a(x_i) = e^{-\beta B x_i}$ are connected to a unique variable node. Thus

$$P(x_1, \dots, x_N) = \prod_{a=1}^{N} F_a(x_i, x_{i+1}) \phi_a(x_i) \tag{3.4}$$

3.2 The Belief Propagation equations

The general problem of computing marginals of a graphical model can take a time exponential in the number of variables. However when the underlying factor graph

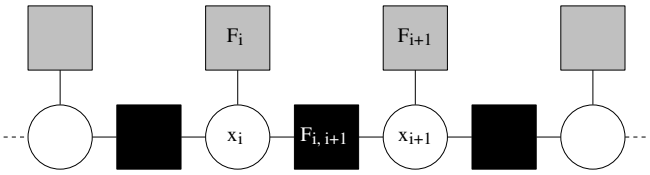


Figure 3.2: Factor graph for an Ising chain. Circles correspond to the Ising spins $x_i \in \{+1,-1\}$. Black squares correspond to the pairwise interaction terms $F_a(x_i,x_{i+1}) = e^{-x_i x_{i+1}}$. Grey squares correspond to the magnetic field terms $\phi_a(x_i) = e^{-Bx_i}$.

is a tree the problem becomes computationally tractable by the use of message-passing algorithms. They operate on messages associated with edges of a factor graph, messages are updated recursively through computations at the vertices of the factor graph. The algorithm design relies on the fact that the probability distribution can be factorized in terms of local functions, i.e factors. Belief Propagation is a message passing algorithm that encodes the update rules yielding exact marginals on trees. The BP equations (as well as the name "Belief Propagation") were introduced in computer science by Pearl [53]. In this section we simply present the algorithm.

Consider the problem of computing marginals of a graphical model with N variables $(x_1,\ldots,x_N)$ taking value in a finite set χ. We are given a set $\mathbf{x} = (x_1,\ldots,x_N)$ of random variables with a joint probability distribution

$$M(\mathbf{x}) = \frac{1}{Z}\prod_a F_a(\mathbf{x}_{\partial a}), \tag{3.5}$$

where $\mathbf{x}_{\partial a} \equiv \{x_i | i \in \partial a\}$ is the set of variables involved in the constraint a. Given the factor graph representing (3.5), functions called messages are associated with every directed edge on the factor graph and they take values in the space of single-variable probability distributions. For each edge (i,a) two messages are defined: $m_{i\to F_a}$ and $p_{F_a\to i}(x_i)$. The fixed-point Belief Propagation (BP) equations

for messages are (fig. 3.3):

$$p_{F_a \to i}(x_i) = \frac{1}{Z_{ai}} \sum_{\{x_j : j \in \partial a \setminus i\}} F_a\left(\{x_i\}_{i \in \partial a}\right) \prod_{j \in \partial a \setminus i} m_{j \to F_a}\left(x_j\right) \tag{3.6}$$

$$m_{i \to F_a}(x_i) = \frac{1}{Z_{ia}} \prod_{b \in \partial i \setminus a} p_{F_b \to i}(x_i). \tag{3.7}$$

It is possible to solve the BP equations by iteration. It is necessary to define messages $m^{(t)}_{i \to F_a}(x_i)$ and $p^{(t)}_{F_a \to i}(x_i)$ at the t-th iteration, as functions of the r.h.s. of the Eqs. (3.6)- (3.7) with messages at time t-1, then iterate the computation until a fixed point is found. Then, the fixed point messages are $m^{(t)}_{i \to F_a}(x_i) = m_{i \to F_a}(x_i)$ and $p^{(t)}_{F_a \to i}(x_i) = p_{F_a \to i}(x_i)$.

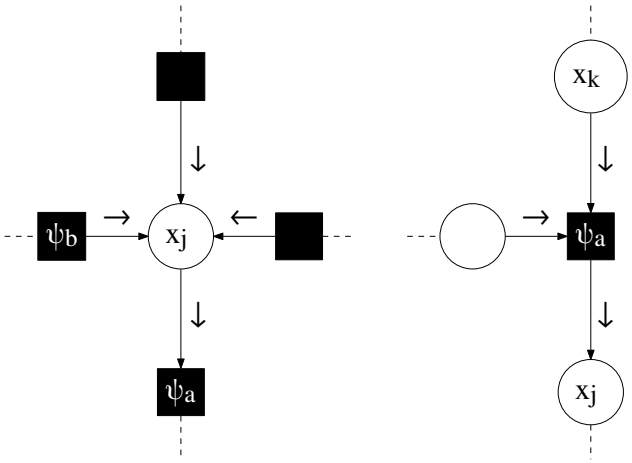


Figure 3.3: Left: the local portion of the factor graph involved in the computation of $m_{j \to F_a}\left(x_j\right)$. This message is a function of the incoming messages $p_{F_a \to j}\left(x_j\right)$ with $b \neq a$. Right: the portion of the factor graph involved in the computation of $p_{F_b \to j}\left(x_j\right)$ as a function of messages $m_{k \to F_a}(x_k)$ with $k \neq j$.

At the fixed point they provide an approximate value for the marginal probability of the variables [52]

$$m_i(x_i) = \frac{1}{Z_i} \prod_{b \in \partial i} p_{F_b \to i}(x_i) \tag{3.8}$$

The marginal probability computed by BP is exact on tree-graphical models.

Theorem 2. *BP is exact on trees. Consider a tree graphical model in which the maximum distance between any two variables nodes is t_*. Then*

1. *Irrespective of the initial condition, BP update equations (3.6-3.7) converge after at most t_* iterations. That means:*

$$t > t_*, \quad m^{(t)}_{i \to F_a}(x_i) = m_{i \to F_a}(x_i), \quad p^{(t)}_{F_a \to i}(x_i) = p_{F_a \to i}(x_i) \qquad \forall (ia)$$

2. *the fixed point messages provide the exact marginals $\mu(x_i)$: for any variable node i and $t > t_*$,*

$$m^{(t)}_i(x_i) = \frac{1}{Z_i} \prod_{b \in \partial i} p^{(t-1)}_{F_b \to i}(x_i) = m_i(x_i). \tag{3.9}$$

A formal proof is sketched in [52]. We will see that the correspondence with the Bethe Approximation gives an explanation to this result. However it was found that BP can be pretty effective on loopy graph, too. The local (in the factor graph definition) nature of BP suggests that it can occur on graphs that are locally tree. In fact when computing equations (3.6)-(3.7) on trees messages incoming from each branch are independent. Therefore we can assume that BP is still a good approximation when correlations between variables tend to zero and this is true in the large size limit on graphs that are locally trees.

The Belief Propagation equations was also found in the spin glass community of physicists using an approach called *cavity method* [54, 55]. The simplest example of cavity method dates back to the work of H. Bethe on Ising ferromagnet [56]. This technique is based on the construction of a new graph (the cavity graph) equivalent to the original, but where a node has been removed. For each neighbors

of the removed one a marginal probability is computed by summing out variables of the branch it belongs. In this way an effective cavity fields is obtained and is then applied along with the interaction between to the removed node and its neighbor.

3.3 The Bethe Approximation

We begin presenting a variational approach to the computation of the free-energy of a system, then we will introduce the Bethe approximation. Let us consider a system of N variables x_i. The state of the system is $\mathbf{x} = (x_1, \ldots, x_N)$ and each state has an energy $E(\mathbf{x})$. In statistical mechanics the Boltzmann's law gives the probability of a state

$$P(\mathbf{x}) = \frac{1}{Z(T)} e^{-\beta E(\mathbf{x})} \tag{3.10}$$

where $Z(T)$ is a normalization constant known as partition function

$$Z(T) = \sum_{\mathbf{x}} e^{-\beta E(\mathbf{x})}, \tag{3.11}$$

where the sum is over all possible states $\mathbf{x} \in \chi^N$ of the system.

Now we want to address an opposite situation. Given the joint probability distribution $M(\mathbf{x}) = \frac{1}{Z} \prod_a F_a(\mathbf{x}_{\partial a})$ we can use the Boltzmann's law as a postulate to define an energy function for the system

$$E(\mathbf{x}) = Z - \log(M(\mathbf{x})) = -\sum_a \log(F_a(\mathbf{x}_{\partial a})), \tag{3.12}$$

where we set $\beta = 1$. In fact, the inverse temperature β has no physical meaning in this context, it simply sets the scale for the units in which the energy is measured and can be set arbitrarily. If one (or more) factor $F_a(\mathbf{x}_{\partial a})$ is equal to zero for some configuration of $\mathbf{x}_{\partial a}$, then the overall probability of states $M(\mathbf{x})$ which contain the forbidden configuration $\mathbf{x}_{\partial a}$ is zero. The corresponding energy is infinite. Deterministic functions are an important class of factors that have forbidden configuration. From the partition function Z the *Helmholtz free energy* is defined as

$$f_H = -\ln Z \tag{3.13}$$

where we used $\beta = 1$ as explained before. Computing the partition function Z is often infeasible, especially for systems with a large number of variables (or particles), because one have to trace over all the possible configurations. Nevertheless, since the free-energy allows computation of measurable system observables, a great effort has been devoted to its approximations. Following [57] in order to approximate f_H one can use a variational approach introducing a trial probability distribution $b(\mathbf{x})$, such that $0 \leq b(\mathbf{x}) \leq 1$, and a *variational free energy*

$$f(b) = U(b) - H(b) \tag{3.14}$$

where the *variational average energy* is

$$U(b) = \sum_{\mathbf{x}} b(\mathbf{x}) E(\mathbf{x}) \tag{3.15}$$

and the *variational entropy*

$$H(b) = -\sum_{\mathbf{x}} b(\mathbf{x}) \ln(b(\mathbf{x})). \tag{3.16}$$

The variational free-energy is actually the *Gibbs free-energy* $f_G = U - TS$ in which we set $T = 1$. It follows that

$$f(b) = f_H + D(b||M) \tag{3.17}$$

where

$$D(b||M) = \sum_{\mathbf{x}} b(\mathbf{x}) \ln\left(\frac{b(\mathbf{x})}{M(\mathbf{x})}\right) \tag{3.18}$$

is the Kullback-Leibler divergence between $b(\mathbf{x})$ and $M(\mathbf{x})$. Since $D(b||M)$ is always nonnegative and is zero if and only if $b(\mathbf{x}) = M(\mathbf{x})$, then $f(b) \geq f_H$, where the equality stands when $b(\mathbf{x}) = M(\mathbf{x})$. Therefore minimizing the variational free-energy $f(b)$ with respect to $b(\mathbf{x})$ leads to computing the exact value of f_H and $M(\mathbf{x})$. In other words, minimizing the Gibbs free-energy with respect to the trial probability distribution gives an upper bound to the Helmholtz free-energy. If we constraint the beliefs to be $b(\mathbf{x}) = e^{-E(\mathbf{x})}/Z$, we obtain the original expression for the Helmholtz free-energy, that is hard to compute. On the contrary, we can

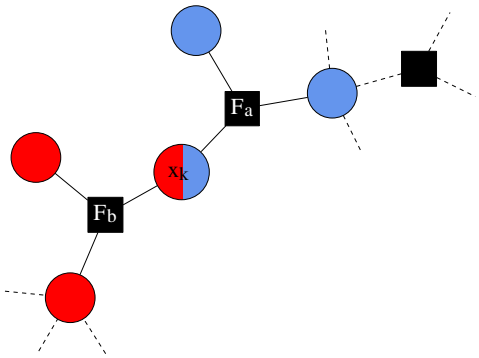


Figure 3.4: Each region in $\mathscr{R}_L$ contains one factor (for example F_a or F_b) and all its neighboring nodes (respectively blue and red ones). Each region in $\mathscr{R}_S$ contains a single variable node. In the example node x_k belongs to both the F_a and F_b regions and to the single variable region in $\mathscr{R}_S$.

minimize $f(b)$ by constraining the beliefs to be a factorized probability distribution

$$b_{MF}(\mathbf{x}) = \prod_{i=1}^{N} b_i(x_i) \tag{3.19}$$

that leads to the *mean-field approximation*.

R. Kikuchi introduced a different class of approximations to the variational free energy [58]. In a *Kikuchi approximation* (*cluster variational method*) the approximate free-energy is a function of trial probability distributions $b_S(\mathbf{x}_S)$ over a set S of variable nodes of the factor graph. Kikuchi recognized that the *Bethe approximation* method for magnets [56] is the simplest example of cluster variational method. We leave the original formulation out and in the following section we refers to the formulation by Jonathan S. Yedidia [57].
In the Bethe approximation, we take a set of region $\mathscr{R}$ of the factor graph of two types (see Fig. 3.4). A set of regions $\mathscr{R}_L$ such that each of the M regions in $\mathscr{R}_L$ contains one factor node and all its neighboring variables nodes. A set $\mathscr{R}_S$ such that each of the N regions in $\mathscr{R}_S$ contains a single variable nodes. The Bethe free

energy is $f_{Bethe} = U_{Bethe} - H_{Bethe}$ where the Bethe average energy is

$$U_{Bethe} = -\sum_{a=1}^{M} \sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) \ln(F_a(\mathbf{x}_{\partial a})) \tag{3.20}$$

and the Bethe entropy is

$$H_{Bethe} = -\sum_{a=1}^{M} \sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) \ln(b_a(\mathbf{x}_{\partial a})) + \sum_{i=1}^{N} (d_i - 1) \sum_{x_i} b_i(x_i) \ln(b_i(x_i)) \tag{3.21}$$

the factor $d_i - 1$ is the counting number which ensures that each node of the factor graph is counted exactly once when regions in $\mathscr{R}_L$ and in $\mathscr{R}_S$ overlap. Moreover, since we require that the beliefs are marginals, we enforce the normalization constraint

$$\sum_{x_i} b_i(x_i) = \sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) = 1, \tag{3.22}$$

the consistency constraint

$$\sum_{\mathbf{x}_{\partial a} \setminus x_i} b_a(\mathbf{x}_{\partial a}) = b_i(x_i) \tag{3.23}$$

and the inequality constraints

$$\begin{aligned} 0 \leq & b_i(x_i) \leq 1 \\ 0 \leq & b_a(\mathbf{x}_{\partial a}) \leq 1. \end{aligned} \tag{3.24}$$

The Bethe approximation for the variational free-energy is also obtained inserting in (3.14) a trial function that considers a two-term factorization

$$b_{Bethe}(\mathbf{x}) = \frac{\prod_{a=1}^{M} b(\mathbf{x}_{\partial a})}{\prod_{i=1}^{N} (b_i(x_i))^{d_i - 1}} \tag{3.25}$$

where in the case of trees $b(\mathbf{x}_{\partial a}) = b_{ij}(x_i, x_j)$. When the underlying factor graph is a tree the true probability distribution factorizes exactly as (3.25), thus the approximation turns into an exact result.

3.4 Correspondence between BP and Bethe Approximation

The connection between Belief Propagation and the Bethe Approximation was firstly pointed out by Kabashima and Saad [59, 60], then investigated in a general formulation by Yedidia, Freeman and Weiss [61, 57]. Following the latter, it is possible to show that the stationary points of the Bethe free energy correspond to the Belief Propagation fixed points with positive beliefs. In order to recover all the constrained stationary points, we can use the Lagrangian formalism. We use Lagrange multipliers γ_a and γ_i for the normalization constraint (3.22), $\lambda_{ai}(x_i)$ for the marginalization constraint (3.23). The conditions on the Lagrangian stationary points are obtained by imposing the *complementary slackness condition*, i.e. enforcing that either the inequality constraints are satisfied with the equality, or the Lagrange multipliers must be equal to zero [57]. As a consequence, we can ignore the Lagrange multipliers for the inequality constraint (3.24), because we are interested in stationary points where the constraint is satisfied with the inequality. The Lagrangian has the form

$$\begin{aligned} L =& f_{Bethe} + \sum_a \gamma_a \left[\sum_{\mathbf{x}_a} b_a(\mathbf{x}_{\partial a}) - 1 \right] + \sum_i \gamma_i \left[\sum_{x_i} b_i(x_i) - 1 \right] \\ &+ \sum_i \sum_{a \in \partial i} \sum_{x_i} \lambda_{ai}(x_i) \left[b_i(x_i) - \sum_{\mathbf{x}_{\partial a} \setminus x_i} b_a(\mathbf{x}_{\partial a}) \right]. \end{aligned} \tag{3.26}$$

The sum over i is over variable nodes with degree $d_i > 2$. In fact the free-energy contributions of single variable nodes connected to only one factor node have already taken into account in the region $\mathscr{R}_{\mathscr{L}}$. Then they have counting number $c_i = d_i - 1$ equal to zero and we do not need to take them into account for the Bethe free-energy. Setting the derivatives of the Lagrangian with respect to the beliefs equal to zero we find an expression for Lagrange multipliers as function of the beliefs, then inverting this expression we get the beliefs at the stationary points

$$\hat{b}_a(\mathbf{x}_{\partial a}) = F_a(\mathbf{z}_a) \exp\left[\gamma_a - 1 + \sum_{i \in \partial a} \lambda_{ai}(x_i) \right] \tag{3.27}$$

and

$$\hat{b}_i(x_i) = \exp\left[\frac{1}{d_i - 1}\left(1 - \gamma_i + \sum_{a \in \partial i} \lambda_{ai}(x_i)\right)\right] \tag{3.28}$$

If we make the identification (see [57])

$$\lambda_{ai} = \ln p_{i \to a}(x_i) = \ln \prod_{b \in \partial i \setminus a} m_{b \to i}(x_i), \tag{3.29}$$

we find the standard fixed-point equations of Belief Propagation (3.6)-(3.7). We can recover the lagrangian stationary point (3.27)-(3.28) inverting (3.29) with respect to $m_{b \to i}(x_i)$, then replacing messages in the BP equation. It is worth to underline that this results are obtained by the minimization of the Bethe free-energy and not the actual Gibbs free-energy (3.14). In fact f_{Bethe} is an approximation of f_G and it is the exact f_G only on trees [57]. Therefore it is not guaranteed that we find an upper bound for the Helmholtz free energy f_H.

3.4.1 The Bethe free-energy as a function of BP messages

If we express the trial functions $b_i(x_i)$ and $b_a(\mathbf{x}_{\partial a})$, that are local marginal, in function of the messages (aside from a normalization factor)

$$b_a(\mathbf{x}_{\partial a}) \propto F_a\left(\{x_i\}_{i \in \partial a}\right) \prod_{j \in \partial a \setminus i} m^{(t)}_{j \to F_a}(x_j) \tag{3.30}$$

$$b_i(x_i) \propto \prod_{b \in \partial i \setminus a} p^{(t)}_{F_b \to i}(x_i), \tag{3.31}$$

we can write the Bethe free energy and eqs. (3.20)-(3.21) as (see also [52])

$$-f = \sum_a f_a + \sum_i f_i - \sum_{(ia)} f_{(ia)} \tag{3.32}$$

in which the local contributions can be expressed as function of the Belief Propagation messages

$$f_a = \log\left(\sum_{\{x_i : i\in\partial a\}} F_a(\{x_i\}_{i\in\partial a}) \prod_{i\in\partial a} m_{i\to a}(x_i)\right) \tag{3.33}$$

$$f_i = \log\left(\sum_{x_i} \prod_{b\in\partial i} p_{F_b\to i}(x_i)\right) \tag{3.34}$$

$$f_{(ia)} = \log\left(\sum_{x_i} m_{i\to a}(x_i) p_{F_b\to i}(x_i)\right). \tag{3.35}$$

This form for the Bethe free-energy is useful to computational purposes.

Chapter 4

SIR Prediction via Bayesian approach

In this chapter we introduce a Bayesian approach to epidemic processes on networks that relies on the Belief Propagation approximation. This approach provides marginal probabilities for the state of the nodes at any time [32]. Therefore it can be used to predict epidemic evolution as well as to infer the patient zero. Unlike the methods described in Chapter 2, this approach relies only on the Belief Propagation approximation. On a tree, it takes into account the correlations between pairs of nodes, provides exact marginal probabilities and allows for an exact inference of the source [33]. It can be easily adapted to cope with incomplete or noisy observation, time-varying network structures and epidemic parameters. By the computation of a free-energy the probability of transmission/recovery and the extinction time can also be inferred.

4.1 Bayesian Approach

The idea behind the Bayesian approach is to use evidences to modify some initial assumption and improve predictions. Let $\mathbf{x}$ be a random vector defined by the outcomes of some experiment, the values taken by its elements are called observations. Let $\mathbf{x}$ be dependent on a set of unknown parameters collected in the vector θ. We call *prior* the probability distribution $P(\theta)$ for the realizations of the parameters vector (we assume that parameters take value in an admissible parameter space

$\theta \subset \Theta$). The probability distribution, called *likelihood*, of vector $\mathbf{x}$ is the conditional probability distribution $P(\mathbf{x}|\theta)$ given the realization θ of the parameters vector. The probability $P(\theta|\mathbf{x})$ of a realization of θ given the observation $\mathbf{x}$ is called the *posterior*. The two quantities are related by the *Bayes' Theorem*,

$$P(\theta|\mathbf{x}) = \frac{P(\mathbf{x}|\theta)P(\theta)}{P(\mathbf{x})}. \tag{4.1}$$

In the following we will show how the problem of predicting a SIR epidemic process given an observation can be tackled using a Bayesian approach. The *posterior* probability of a configuration $\mathbf{x}^t$ at time t given an observation $\mathbf{x}^{T_{obs}}$ at time T_{obs} can be written as

$$P\left(\mathbf{x}^t|\mathbf{x}^{T_{obs}}\right) = \frac{P\left(\mathbf{x}^t,\mathbf{x}^{T_{obs}}\right)}{P\left(\mathbf{x}^{T_{obs}}\right)} \propto \sum_{\mathbf{x}_0} P\left(\mathbf{x}^t,\mathbf{x}^{T_{obs}}|\mathbf{x}^0\right) P\left(\mathbf{x}^0\right), \tag{4.2}$$

where in the last expression we neglected the *a priori* probability $P\left(\mathbf{x}^{T_{obs}}\right)$ of the observed state because in our analysis it is only a normalization constant, and $P\left(\mathbf{x}^0\right)$ is the *prior* on the initial conditions. In order to obtain the posterior probability $P\left(\mathbf{x}^t|\mathbf{x}^{T_{obs}}\right)$ it is required to compute the joint probability distribution $P\left(\mathbf{x}^t,\mathbf{x}^{T_{obs}}|\mathbf{x}^0\right)$ of the states at the observation time T_{obs} and at some later time t given the initial condition $\mathbf{x}_0$, for all possible initial conditions. In principle, this quantity could be evaluated experimentally, by taking into account all possible realizations compatible with the constraints imposed by the SIR dynamics and the observation and discarding the others. However, the number of possible epidemic trajectories of length t grows as 3^{tN}, making this brute-force approach computationally unfeasible even for small systems and very early observations.

4.2 The Belief Propagation approximation

Although the sum on the right-hand side of Eq. (4.2) still runs over a possibly huge number (exponentially large in N) of configurations, the posterior $P\left(\mathbf{x}^t|\mathbf{x}^{T_{obs}}\right)$ can be efficiently computed by the use of the Belief Propagation technique [32, 33]. It is necessary to address the joint probability distribution $P\left(\mathbf{x}^t,\mathbf{x}^{T_{obs}}|\mathbf{x}^0\right)$ by using a probabilistic graphical model defined by the static representation presented in

Section 4.2.1 and then write down a factorized form for the posterior (4.2).

When the underlying contact network is a tree, the factor graph on which the graphical model is defined can be also reduced to a tree, and the joint probability distribution can be computed exactly by solving a set of local fixed-point equations known as Belief Propagation (BP) equations. On more general graphs, the BP equations can be considered as an algorithm that provides a good approximation of the real probability distribution [52], see Sections 1.5 and 3.

4.2.1 Factor graph for the SIR model

We aim to introduce a factor graph representation of (4.2) in order to use the BP approximation, therefore we address the posterior probability to get an expression factorized over the single node variables. Since the SIR model is an irreversible process every realization of the trajectory $(\mathbf{x}^0,\dots,\mathbf{x}^t)$ is in one-to-one correspondence with a static configuration of individual infection times $\mathbf{t}=\{t_i\}_{i\in V}$ and recovery times $\mathbf{g}=\{g_i\}_{i\in V}$. Given the initial configuration $\mathbf{x}^0$, for each node $i\in V$, a *recovery delay* g_i is randomly drawn according to the geometric distribution

$$\mathscr{G}_i(g_i)=\mu_i(1-\mu_i)^{g_i} \tag{4.3}$$

and the infection *transmission delays* s_{ij} from node i to node j are generated from the conditional distribution

$$\omega\left(s_{ij}|g_i\right)=\begin{cases}\lambda_{ij}(1-\lambda_{ij})^{s_{ij}} & s_{ij}\leq g_i\\ \sum_{s>g_i}\lambda_{ij}(1-\lambda_{ij})^{s} & s_{ij}=\infty.\end{cases} \tag{4.4}$$

Infection times are then given by the deterministic equation

$$t_i=\min_{j\in\partial i}(t_j+s_{ji})+1. \tag{4.5}$$

In the static representation of the epidemic dynamics, we can express the posterior probability of a configuration $\mathbf{x}^t$ at time t given an observation $\mathbf{x}^{T_{obs}}$ at time T_{obs},

by the Bayes' Theorem, as

$$P\left(\mathbf{x}^t|\mathbf{x}^{T_obs}\right) \propto \sum_{\mathbf{x}^0} P\left(\mathbf{x}^t,\mathbf{x}^{T_{obs}}|\mathbf{x}^0\right) P(\mathbf{x}^0)$$
$$= \sum_{\mathbf{t},\mathbf{g},\mathbf{x}^0} P\left(\mathbf{x}^t|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right) P\left(\mathbf{x}^0\right). \tag{4.6}$$

$P\left(\mathbf{x}^t|\mathbf{t},\mathbf{g}\right)$ and $P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right)$ are deterministic functions of the set $(\mathbf{t},\mathbf{g})$, indeed they connect configurations in term of $(\mathbf{t},\mathbf{g})$ in the static representation to configurations $\mathbf{x}^t$ and $\mathbf{x}^{T_{obs}}$ given in terms of epidemic states of the nodes. $P\left(\mathbf{x}^0\right)$ is the prior probability for the initial configuration. $P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right)$ is the joint probability distribution of infection and recovery times conditioned on the initial configuration $\mathbf{x}^0$. Since it needs special attention we will explicitly analyze this term last.

For each node with can write the deterministic function $\zeta_i^{T_{obs}}(g_i,t_i,x_i^{T_{obs}}) = P(x_i^{T_{obs}}|t_i,g_i)$. The ζ_i gives one if the infection and recovery time (g_i,t_i) are compatible to the observed state given by variable $x_i^{T_{obs}}$ otherwise it gives zero

$$\zeta_i^{T_{obs}}\left(t_i,g_i,x_i^t\right) = \mathbb{I}\left[x_i^{T_{obs}} = S, T_{obs} < t_i\right] +$$
$$+\mathbb{I}\left[x_i^{T_{obs}} = I, t_i \leq T_{obs} \leq (t_i+g_i)\right] + \mathbb{I}\left[x_i^{T_{obs}} = R, T_{obs} > t_i+g_i\right]. \tag{4.7}$$

This function depends on a single node variables, by consider all nodes we get the factorized form

$$P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right) = \prod_i \zeta_i^{T_{obs}}\left(t_i,g_i,x_i^t\right). \tag{4.8}$$

The same is valid for $P\left(\mathbf{x}^t|\mathbf{t},\mathbf{g}\right) = \prod_i \zeta_i^t\left(t_i,g_i,x_i^t\right)$. The function $\zeta_i^t\left(t_i,g_i,x_i^t\right)$ ensures that we take into account only values of (t_i,g_i) compatible with the final desired state x_i^t when we compute the posterior probability. Omitting this term we get the probability $P(\mathbf{t},\mathbf{g}|\mathbf{x}^{T_{obs}})$, i.e we get a posterior for any pairs (t_i,g_i) for every node i.

The prior probability on the initial configuration is also factorized over each node $P\left(\mathbf{x}^0\right) = \prod_i \gamma_i(x_i^0)$ with

$$\gamma_i(x_i^0) = \gamma_i \mathbb{I}[x_i^0 = I] + (1-\gamma_i)\mathbb{I}[x_i^0 = S] \tag{4.9}$$

where γ is the probability that each node is infected in the initial configuration. We finally consider to the joint probability distribution of infection and recovery times $\mathbf{t}, \mathbf{g}$ conditioned on the initial configuration. Making explicit the dependence on the transmission delays s_{ij}, we write

$$\begin{aligned} P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right) =& P\left(\mathbf{t}|\mathbf{x}^0,\mathbf{g}\right) P\left(\mathbf{g}\right) = \\ &= \sum_{\left\{s_{ij}\right\}} P(\mathbf{s}|\mathbf{g}) P(\mathbf{t}|\mathbf{x}^0,\mathbf{g},\mathbf{s}) P(\mathbf{g}). \end{aligned} \tag{4.10}$$

The description of the static representation in Section 1.3 provides the required probability distributions. In fact, the conditional probability of a configuration of transmission delays $\mathbf{s}$ is given by the product of the truncated geometric distribution ω_{ij}, defined by equation (4.4), over all pairs of nodes (i, j)

$$P(\mathbf{s}|\mathbf{g}) = \prod_{i,j} \omega_{ij}\left(s_{ij}|g_i\right). \tag{4.11}$$

Since the recovery delays depend only on single node recovery probabilities μ_i, the probability distribution for the recovery delays $\mathbf{g}$ is factorized over the nodes, each one characterized by the geometric distribution, i.e. (4.3)

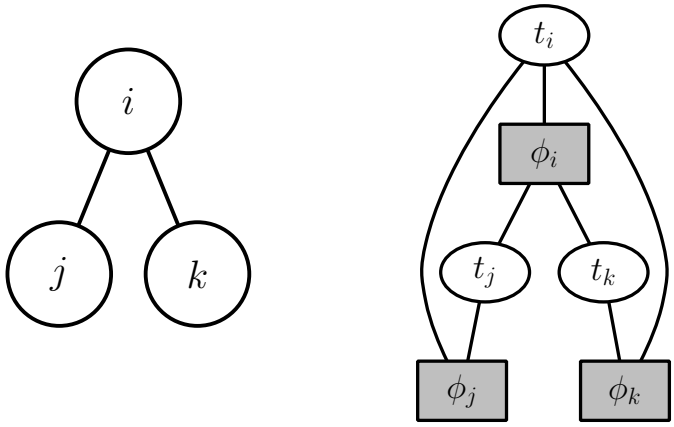
$$P\left(\mathbf{g}\right) = \prod_i \mathscr{G}_i\left(g_i\right). \tag{4.12}$$

Equation (4.5) gives the time t_i at which node i gets infected and it can be cast in a function ϕ_i such that

$$\phi_i\left(t_i, \{t_k, s_{ki}\}_{k\in\partial i}\right) = \delta(t_i, \mathbb{I}[x_i^0 \neq I](\min_{k\in\partial i}(t_k + s_{ki}) + 1)), \tag{4.13}$$

where $\mathbb{I}[x_i^0 \neq I]$ imposes the value $t_i = 0$ if node i is infected in the initial configuration. The probability of finding a configuration for $\mathbf{t}$ given the initial configuration $\mathbf{x}_0$, the recovery delays $\mathbf{g}$ and the transmission delays $\mathbf{s}$ is

$$P(\mathbf{t}|\mathbf{x}^0,\mathbf{g},\mathbf{s}) = \prod_i \phi_i\left(t_i, \{t_k, s_{ki}\}_{k\in\partial i}\right). \tag{4.14}$$



(a) (b)

Figure 4.1: (a): original graph with nodes (i, j, k). (b): the factor graph representation. Even in simple cases it shows a loopy structure.

In conclusion the posterior probability (4.6) is given by

$$P\left(\mathbf{x}^t|\mathbf{x}^{T_obs}\right) \propto \sum_{\mathbf{t},\mathbf{g},\mathbf{x}^0} P\left(\mathbf{x}^t|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right) P\left(\mathbf{x}^0\right)$$
$$= \sum_{\mathbf{t},\mathbf{g},\mathbf{s},\mathbf{x}^0} \prod_{i,j} \omega_{ij} \prod_i \phi_i \zeta_i^{T_{obs}} \zeta_i^t \mathscr{G}_i \gamma_i. \tag{4.15}$$

The posterior probability is the product of local terms that can be represented by using a factor graph. Nonetheless with this definition the corresponding factor graph has a loopy structure that can decrease the accuracy of the Belief Propagation approximation even in the case of trees. We can verify the existence of such loopy structure focusing on the dynamical constraint (4.5). In fact, the infection time of each node i, given by $\phi_i\left(t_i, \{t_k, s_{ki}\}_{k\in\partial i}\right)$, depends on the infection times of its neighboring node $k \in \partial i$ and vice versa. This is true for every pairs of nodes (i,k) connected by an edge.

4.2.2 Disentangled factor graph representation

In order to guarantee, at least, that the stationary points of BP give the exact values for the marginal probability (Theorem n.2 in Section 3.2) when the underlying network is a tree, we require a factor graph representation that maintains the topological properties of the original graph of contacts between nodes. We meet this requirement by using the approach proposed in [32] (see also [31, 30]). The representation can be disentagled by grouping pairs of infection times $\left(t_i, t_j\right)$ in the same variable node associated to the edge $(i, j) \in E$. Then a factor node enforcing the dynamical constraint (4.13) is associated with the nodes of the original graph. For each edge of the original graph from node i to node j we introduce a copy $t_i^{(j)}$ of the infection time t_i that is forced to take the common value t_i by the constraint $\prod_{k\in\partial i}\delta(t_i^{(k)}, t_i)$. The function (4.13) for a node i depends on infection times t_j and

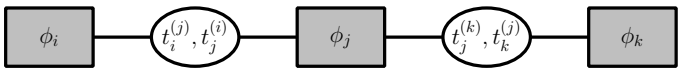


Figure 4.2: The factor graph representation for the graph in fig. 4.1(a) when pairs of variables (t_i, t_j) are grouped together.

transmission delays s_{ji} for $j \in \partial i$ only through their sum, thus it is convenient to introduce the variable $t_{ji} = t_j + s_{ji}$, i.e the time at which the neighboring node j transmits the infection i. It is also convenient to group the recovery time g_i with the infection time t_i. For each edge emerging from the node i we introduce the copies $g_i^{(j)}$ of the recovery time g_i and the identity constraint $\prod_{k\in\partial i}\delta(g_i^{(k)}, g_i)$, then we can define the triplet $(t_i^{(j)}, t_{ji}, g_i^{(j)})$ as a new variable node.

The new function enforcing the dynamical constraint for each node i becomes

$$\begin{aligned}\psi_i &= \delta(t_i, \mathbb{I}[x_i^0 \neq I](\min_{j\in\partial i}(t_{ji}+1)))\prod_{j\in\partial i}\delta(t_i^{(j)}, t_i)\delta(g_i^{(j)}, g_i) = \\ &= \phi_i(t_i, \left\{t_{ji}\right\}_{j\in\partial i})\prod_{j\in\partial i}\delta(t_i^{(j)}, t_i)\delta(g_i^{(j)}, g_i). \end{aligned} \tag{4.16}$$

For each edge between a pair of nodes (i, j) a function ϕ_{ij} is defined

$$\phi_{ij} = \omega_{ij}(t_{ij} - t_i^{(j)}|g_i^{(j)})\omega_{ji}(t_{ji} - t_j^{(i)}|g_j^{(i)}), \tag{4.17}$$

accounting for the interaction between the pair of nodes. Then we can write the posterior distribution (4.15) in the new "disentangled" representation as

$$\begin{aligned} P\left(\mathbf{x}^t|\mathbf{x}^{T_{obs}}\right) &\propto \sum_{\mathbf{t},\mathbf{g},\mathbf{s},\mathbf{x}^0} P\left(\mathbf{x}^t|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right) P\left(\mathbf{x}^0\right) \\ &= \sum_{\mathbf{t},\mathbf{g},\mathbf{s},\mathbf{x}_0} \mathscr{Q}(\mathbf{g},\mathbf{t},\left\{t_{ij}\right\},\mathbf{x}_0) \end{aligned} \tag{4.18}$$

where

$$\mathscr{Q}(\mathbf{g},\mathbf{t},\left\{t_{ij}\right\},\mathbf{x}_0) = \frac{1}{Z}\prod_{i<j}\phi_{ij}\prod_i \psi_i \zeta_i^{T_{obs}}\zeta_i^t \mathscr{G}_i \gamma_i. \tag{4.19}$$

Fig. 4.3 shows the factor graph for the distribution (4.19).

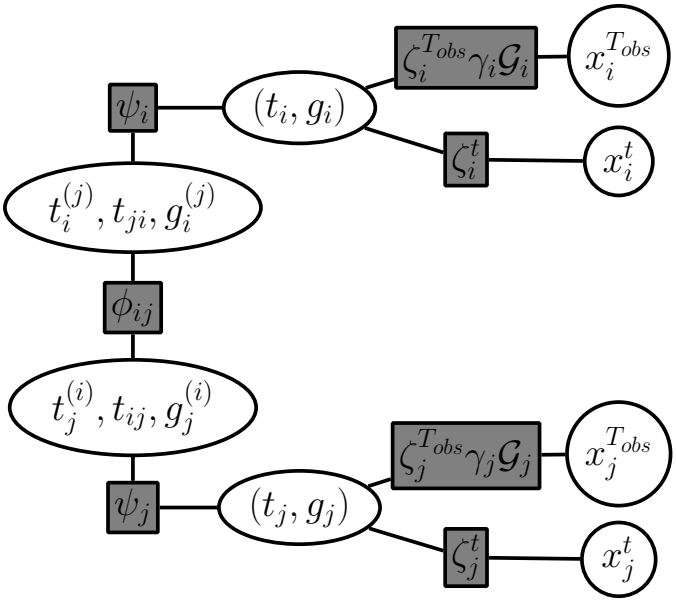


Figure 4.3: The factor graph representation for the graphical model associated to the distribution (4.18).

.

4.2.3 Unknown initial time

In most realistic cases, the observation of one (or more) infected individuals cannot provide information about the initial time nor the age of the epidemic process. Given an upper bound ΔT for the uncertainty on the epidemic age, we can consider the dynamical process to start from the all-susceptibles configuration and let the nodes have a probability of spontaneous infection at arbitrary time before the observation. In our factor graph representation it is equivalent to adding a fictitious neighbor to every node whose activation depends only on a prior probability

$$\varepsilon(g_i'', t_i'', t_i) = \delta(t_i'', \infty)(1-\rho) + (1-\delta(t_i'', \infty))\rho. \tag{4.20}$$

t_i'' and g_i'' are the copies of the variables t_i and g_i for the edge to the fictitious node. This edge is connected to the factor ψ_i, that has to be modified accordingly adding the identity constraints $\delta(t_i'', t_i)$ and $\delta(g_i'', g_i)$. The fictitious node provides the prior probability that node i gets spontaneously infected at any time step, then it is equivalent to take the transmission delay between the fictitious node and i equal to zero. Thus, the contact time is exactly t_i. Considering the process running on a time interval $T + \Delta T$, we can find the probability that a node i is the origin of the epidemic. Fig.4.4 shows the corresponding factor graph.

4.2.4 BP updates for the SIR model

In order to get Belief Propagation approximation of the probability distribution (4.6) we look for the stationary points of the BP equations (3.6) - (3.7) and (3.8). In this case factors $F_a(\underline{z}_a)$ are the ones presented in the SIR factor graph representation in Sec. 4.2.1 and 4.2.2 – i.e $\psi_i, \phi_{ij}, \zeta_i^t, \mathscr{G}_i, \zeta_i^{T_{obs}}, \gamma_i$ and variables nodes z_i are the couples (t_i, g_i) and triplets $(t_i^{(j)}, t_{ji}, g_i^{(j)})$. In general equations (3.7) and (3.8) can be computed efficiently. Instead in (3.6)

$$p^{(t)}_{F_a \to i}(z_i) = \frac{1}{Z_{ai}} \sum_{\{z_j : j \in \partial a \setminus i\}} F_a\left(\{z_i\}_{i \in \partial a}\right) \prod_{j \in \partial a \setminus i} m^{(t)}_{j \to F_a}\left(z_j\right)$$

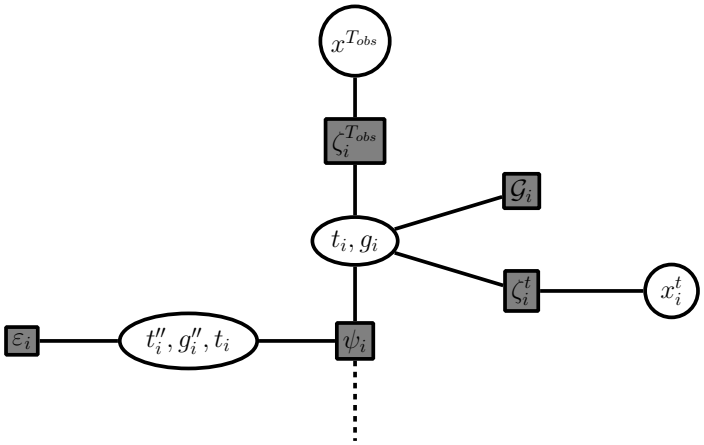


Figure 4.4: The factor graph representation for the graphical model associated to the distribution (4.18) with unknown initial time.

.

the sum over $\{z_{i\in\partial a}\}$ need a time which is exponential in the number of participating variables. The messages for the factor ζ^t are

$$m_{\zeta_i^t \to x_i^t}\left(x_i^t\right) \propto \sum_{t_i,g_i} \zeta_i^t(t_i,g_i,x_i^t) m_{i\to\zeta_i^t}(t_i,g_i) \tag{4.21}$$

and

$$m_{\zeta_i^t \to i}\left(x_i^t\right) \propto \sum_{x_i^t} \zeta_i^t(t_i,g_i,x_i^t) m_{x_i^t\to\zeta_i^t}(x_i^t). \tag{4.22}$$

It is the same for factor nodes $\zeta_i^{T_{obs}}$ where the variable $x_i^{T_{obs}}$ has a value that is the evidence fixed by the observation. $\mathcal{G}_i$ and γ_i do not need any update because they are connected by a single edge to the variable node (t_i, g_i), so their outgoing messages have a fixed value given by the definition of the factor. The outgoing messages from variable node (t_i, g_i) is

$$m_{i\to\psi_i}(t_i,g_i) \propto \sum_{t_i,g_i} \mathcal{G}_i(g_i)\gamma_i(t_i,g_i) m_{\zeta_i^t\to i}(t_i,g_i) m_{\zeta_i^{T_{obs}}\to i}(t_i,g_i). \tag{4.23}$$

In the following we provide efficient forms for the update equations concerning

factor nodes ψ_i and ϕ_{ij} [32, 33]. The update rule for messages coming from factors ψ_i and going to a variable node $(t_i^{(j)}, t_{ji}, g_i^{(j)})$ is

$$\begin{aligned}
p_{\psi_i \to j}(t_i^{(j)}, t_{ji}, g_i^{(j)}) \propto & \sum_{g_i, t_i} \sum_{\left\{t_i^{(k)}, t_{ki}, g_i^{(k)}\right\}} m_{i \to \psi_i}(t_i, g_i) \times \\
& \times \prod_{k \in \partial i \setminus j} m_{k \to \psi_i}(t_i^{(k)}, t_{ki}, g_i^{(k)}) \psi_i \left(t_i, g_i, \left\{(t_i^{(k)}, t_{ki}, g_i^{(k)})\right\}_{k \in \partial i}\right) \\
\propto & \sum_{g_i, t_i} \sum_{\left\{t_i^{(k)}, t_{ki}, g_i^{(k)}\right\}} m_{i \to \psi_i}(t_i, g_i) \times \\
& \times \prod_{k \in \partial i \setminus j} m_{k \to \psi_i}(t_i^{(k)}, t_{ki}, g_i^{(k)}) \times \\
& \times \left[\delta(t_i, 0) + \delta(t_i, (1 + \min_{k \in \partial i}\{t_{ki}\}))\right] \prod_{k \in \partial i} \delta(t_i^{(k)}, ti) \delta(g_i^{(k)}, g_i) \\
\propto & m_{i \to \psi_i}(t_i^{(j)}, g_i^j) \sum_{t_{ki}} \prod_{k \in \partial i \setminus j} m_{k \to \psi_i}(t_i^{(j)}, t_{ki}, g_i^{(j)}) \times \\
& \times \left[\delta(t_i^{(j)}, 0) + \delta(t_i^{(j)}, (1 + \min_{k \in \partial i}\{t_{ki}\}))\right].
\end{aligned} \tag{4.24}$$

In order to reduce the computational complexity we exploit that

$$\delta(t_i, (1 + \min_{j \in \partial i}\{t_{ji}\})) = \prod_{j \in \partial i} \mathbb{I}(t_i \leq t_{ji} + 1) - \prod_{j \in \partial i} \mathbb{I}(t_i < t_{ji} + 1) \tag{4.25}$$

and substituting it into the factor ψ_i we get

$$\begin{aligned}
p_{\psi_i \to j}(t_i^{(j)}, t_{ji}, g_i^{(j)}) \propto & \, \delta(t_i^{(j)}, 0) m_{i \to \psi_i}(0, g_i^{(j)}) \prod_{k \in \partial i \setminus j} \sum_{t_{ki}} m_{k \to \psi_i}(0, t_{ki}, g_i^{(j)}) \\
& + m_{i \to \psi_i}(t_i^{(j)}, g_i^{(j)}) \mathbb{I}(t_i^{(j)} \leq t_{ji} + 1) \prod_{k \in \partial i \setminus j} \sum_{t_{ki} \geq t_i^{(j)} - 1} m_{k \to \psi_i}(t_i^{(j)}, t_{ki}, g_i^{(j)}) \\
& - m_{i \to \psi_i}(t_i^{(j)}, g_i^{(j)}) \mathbb{I}(t_i^{(j)} < t_{ji} + 1) \prod_{k \in \partial i \setminus j} \sum_{t_{ki} > t_i^{(j)} - 1} m_{k \to \psi_i}(t_i^{(j)}, t_{ki}, g_i^{(j)}).
\end{aligned} \tag{4.26}$$

Now it is possible to use a simpler representation for messages that reduces the operations per update to $O(TG)$ where T and G are the maximum infection time and maximum recovery delay. In fact all the information needed resides in the

relative timing between the infection time $t_i^{(j)}$ for node i and the transmission time t_{ji} at which neighbor j transmits the infection to i. In order to simplify the representation we introduce the variable $\sigma_{ij} = 1 + \text{sign}(t_{ji} - (t_i^{(j)} - 1))$. Then in the σ-representation Eq.(4.26) becomes

$$\begin{aligned} p_{\psi_i \to i^{(j)}}(t_i^{(j)}, \sigma_{ji}, g_i^{(j)}) \propto\ & \delta(t_i^{(j)}, 0) m_{i\to\psi_i}(0, g_i^{(j)}) \prod_{k\in\partial i\setminus j} \sum_{\sigma_{ki}} m_{k\to\psi_i}(0, \sigma_{ki}, g_i^{(j)}) \\ & + m_{i\to\psi_i}(t_i^{(j)}, g_i^{(j)}) \mathbb{I}(\sigma_{ji} = 1,2) \prod_{k\in\partial i\setminus j} \sum_{\sigma_{ki}=1,2} m_{k\to\psi_i}(t_i^{(j)}, \sigma_{ki}, g_i^{(j)}) \\ & - m_{i\to\psi_i}(t_i^{(j)}, g_i^{(j)}) \mathbb{I}(\sigma_{ji} = 2) \prod_{k\in\partial i\setminus j} m_{k\to\psi_i}(t_i^{(j)}, 2, g_i^{(j)}). \end{aligned} \tag{4.27}$$

The update rule for messages going to the variable node (t_i, g_i) is

$$\begin{aligned} p_{\psi_i\to i}(t_i, g_i) \propto & \sum_{\left\{t_i^{(k)}, t_{ki}, g_i^{(k)}\right\}} m_{i\to\psi_i}(t_i, g_i) \prod_{k\in\partial i\setminus j} m_{k\to\psi_i}(t_i^{(k)}, t_{ki}, g_i^{(k)}) \times \\ & \times \psi_i\left(t_i, g_i, \left\{(t_i^{(k)}, t_{ki}, g_i^{(k)})\right\}_{k\in\partial i}\right) \\ \propto & \delta(t_i, 0) \prod_{k\in\partial i\setminus j} \sum_{\sigma_{ki}} m_{k\to\psi_i}(0, \sigma_{ki}, g_i) \\ & + \prod_{k\in\partial i\setminus j} \sum_{\sigma_{ki}=1,2} m_{k\to\psi_i}(t_i, \sigma_{ki}, g_i) \\ & - \prod_{k\in\partial i\setminus j} m_{k\to\psi_i}(t_i, 2, g_i). \end{aligned} \tag{4.28}$$

We use the σ-representation to show the update equation for ϕ_{ij} nodes

$$p_{\phi_{ij}\to j}(t_j, \sigma_{ij}, g_j) \propto \sum_{t_i, \sigma_{ji}, g_i} \Omega(t_i, t_j, \sigma_{ij}, \sigma_{ji}, g_i, g_j) p_{i\to\phi_{ij}}(t_i, \sigma_{ji}, g_i). \tag{4.29}$$

The function Ω accounts for the probability that a configuration of variables (t, σ, g) for nodes i and j is realized given the infection rate λ:

$$\Omega(t_i,t_j,\sigma_{ij},\sigma_{ji},g_i,g_j) = \begin{cases} \chi(t_i,t_j,\sigma_{ij},g_i) & : t_i < t_j, \sigma_{ji} = 2, \sigma_{ij} \neq 2 \\ \chi(t_i,t_j,\sigma_{ij},g_i) + (1-\lambda)^{g_i+1} & : t_i < t_j, \sigma_{ji} = 2, \sigma_{ij} = 2 \\ \chi(t_j,t_i,\sigma_{ji},g_j) & : t_j < t_i, \sigma_{ij} = 2, \sigma_{ji} \neq 2 \\ \chi(t_j,t_i,\sigma_{ji},g_j) + (1-\lambda)^{g_j+1} & : t_j < t_i, \sigma_{ij} = 2, \sigma_{ji} = 2 \\ 1 & : t_i = t_j, \sigma_{ij} = \sigma_{ji} = 2 \\ 0 & : \text{otherwise} \end{cases} \tag{4.30}$$

where

$$\chi(t_1,t_2,\sigma,g) = \sum_{t=t_1}^{t_1+g} \delta(\sigma(t_2,t),\sigma)\lambda(1-\lambda)^{t-t_1}. \tag{4.31}$$

The first and third row represent the probability that node i (j respectively) transmits an infection signal to j (i) before its infection time $t_j(t_i)$. In this case infection is possible, but it is worth noting that the target node may have been already infected, this event is taken into account by factor ψ. The second and fourth row represent the probability that node i (j respectively) transmits an infection signal to j (i) after the time of infection $t_j(t_i)$. The symmetry between rows 1-2 and 3-4 can be exploited for a more efficient computation of the update. When one loops over variable (t_j, σ_{ij}, g_j) to compute the output message $p_{\phi_{ij}\to j}(t_j, \sigma_{ij}, g_j)$, one can loop on variables (t_i, σ_{ji}, g_i) of the input message $p_{i\to\phi_{ij}}(t_i, \sigma_{ji}, g_i)$ just by switching indices.

4.3 Temporal networks

In Section 1.6 we introduced temporal networks, that are realistic representations of evolving contact patterns. In the framework already presented we can deal with temporal networks assuming that the transmission probability λ_{ij} depends on time. We call T_{ij} the set of time steps in $\{0,\dots,T\}$ at which nodes i and j make contact. Then the transmission probability takes value λ_{ij} when an edge (i,j) exists and

zero otherwise:

$$\lambda_{ij}^{t} = \begin{cases} \lambda_{ij} & \text{if} \quad t \in T_{ij} \\ 0 & \text{otherwise.} \end{cases} \tag{4.32}$$

Then the delay probabilities in equation (4.4) and the factor ϕ_{ij} in (4.17) are consequently modified with

$$\omega_{ij}(s_{ij}|g_i) = \begin{cases} \lambda_{ij}^{t} \prod_{s=0}^{t-1}(1-\lambda_{ij}^{s}) & \text{if } s_{ij} \leq g_i \\ \prod_{s>g_i}(1-\lambda_{ij}^{s}) & \text{if } s_{ij} = \infty. \end{cases} \tag{4.33}$$

The rest of the formalism remains the same.

4.4 The patient-zero problem via Belief Propagation

The approach proposed naturally allows for the inference of the source of an epidemics [33]. Unlike the naive Bayes methods described in Chapter 2, the Bayesian framework proposed does not introduce further approximations in addition to those due to the Belief propagation technique. On trees, the method coherently exploits information on correlations between the state of neighboring nodes to infer the epidemic origin and avoids misleads due to the mean-field approximation. An example is provided by the case described in Figure 2.1 (Section 2.3), in which disregarding correlations between nodes leads to an incorrect inference of the source. Recalling equation (4.6) we now want to infer $\mathbf{x}_0$ given the observed configuration $\mathbf{x}^{T_{obs}}$, then it becomes

$$P\left(\mathbf{x}^0|\mathbf{x}^{T_{obs}}\right) \propto P(\mathbf{x}^{T_{obs}}|\mathbf{x}^0)P(\mathbf{x}^0) \propto \sum_{\mathbf{t},\mathbf{g}} P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right) P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right) P(\mathbf{x}^0),$$

where the terms involved in the expression are the same already presented in the previous sections. In fact, the Belief Propagation approach provides distribution probabilities for all pairs of (t_i, g_i) in the selected range. The single site posterior marginal is

$$P(x_i^0 = I|\mathbf{x}^{T_{obs}}) = \sum_{g_i,t_i} \zeta_i^t(t_i, g_i, x_i^0 = I) m_{i\to\zeta^t}(t_i, g_i). \tag{4.34}$$

The patient-zero problem has been studied in the case of different observation models, such as the one in which the observed state is partially unknown (noisy observation) or the case in which the observation is erroneous (confused observation) [33].

4.5 Extinction Time

The characterization of the extinction time distribution of an observed outbreak on given network structures is a major issue in epidemic studies [16]. This quantity cannot be directly computed from the knowledge of the local marginals, because it depends on the global evolution of the epidemic process. On the contrary a crucial point of the BP algorithm is that it is very convenient for computing local quantities, such as marginal probability distributions for the single variables or pair-correlations.

Conveniently the posterior probability distribution $\mathscr{P}\left(T_{ext}|\mathbf{x}^{T_{obs}}\right)$ that the discrete-time epidemic process dies out at time T_{ext} when it is conditioned on the observation of the configuration $\mathbf{x}^{T_{obs}}$ can be expressed as the difference between two terms involving the free energies of the related graphical models when the epidemics are constrained to vanish before time T_{ext} and $T_{ext}-1$, respectively:

$$\mathscr{P}\left(T_{ext}|\mathbf{x}^{T_{obs}}\right)=\mathscr{P}\left(t_{ext}<T_{ext}|\mathbf{x}^{T_{obs}}\right)-\mathscr{P}\left(t_{ext}<T_{ext}-1|\mathbf{x}^{T_{obs}}\right). \tag{4.35}$$

In the static representation, the cumulative distribution in the right hand side of (4.35) can be written using the Bayes' theorem as follows

$$\begin{aligned}\mathscr{P}\left(t_{ext}<T_{ext}|\mathbf{x}^{T}_{obs}\right)&=\sum_{\mathbf{t},\mathbf{g}}P\left(t_{ext}<T_{ext}|\mathbf{t},\mathbf{g}\right)P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^{T_{obs}}\right)\\&\propto\sum_{\mathbf{t},\mathbf{g}}P\left(t_{ext}<T_{ext}|\mathbf{t},\mathbf{g}\right)P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right)P\left(\mathbf{t},\mathbf{g}\right)\\&=\sum_{\mathbf{t},\mathbf{g},\mathbf{x}^0}P\left(t_{ext}<T_{ext}|\mathbf{t},\mathbf{g}\right)P\left(\mathbf{x}^{T_{obs}}|\mathbf{t},\mathbf{g}\right)P\left(\mathbf{t},\mathbf{g}|\mathbf{x}^0\right)P\left(\mathbf{x}^0\right).\end{aligned} \tag{4.36}$$

The trace in Eq. (4.36) can be computed from the free-energy of the related graphical model in the Bethe approximation. The posterior probability $P\left(t_{ext} < T_{ext} | \mathbf{x}_{obs}^{T}\right)$ of the extinction time can be written as

$$P\left(t_{ext} < T_{ext} | \mathbf{x}_{obs}^{T}\right) \propto \sum_{\mathbf{t},\mathbf{g},\mathbf{x^0}} \mathscr{Q}\left(\mathbf{x}^{T_{obs}}, T_{ext}, \mathbf{t}, \mathbf{g}, x^0\right) = Z\left(T_{ext}, \mathbf{x}^{T_{obs}}\right). \tag{4.37}$$

The terms in equation (4.36) and in the latter expression are the same as in (4.6), with the exception of $P\left(t_{ext} < T_{ext} | \mathbf{t}, \mathbf{g}\right)$ term factorized over the nodes

$$P\left(t_{ext} < T_{ext} | \mathbf{t}, \mathbf{g}\right) = \prod_i \mathbb{I}\left[(t_i + g_i) < T_{ext}\right]. \tag{4.38}$$

This term constrains the calculation to epidemics that vanish before T_{ext}: we give null probability to every single site configuration with $t_i + g_i$ larger than T_{ext} (except for $t_i = T_{inf}$ that describes susceptible nodes). We call the corresponding factor $\chi^{T_{ext}}(t_i, g_i)$. The logarithm of the partition function is the free energy of the model, hence

$$-f\left(T_{ext}, \mathbf{x}_{obs}^{T}\right) = \log Z\left(T_{ext}, \mathbf{x}_{obs}^{T}\right) = \log P\left(t_{ext} < T_{ext} | \mathbf{x}_{obs}^{T}\right). \tag{4.39}$$

Given a value of the external parameters T_{ext} and the observation $\mathbf{x}_{obs}^{T}$ it is possible to compute the free energy of the graphical model. In the factor graph representation, the Bethe free-energy of the graphical model can be expressed as function of the Belief Propagation messages as described in (3.32). The contributions relating

to the factor nodes are

$$\begin{aligned}\sum_a f_a = \sum_{i\in V}\Bigg\{ &\log\left[\sum_{t_i,g_i,x_i^{T_{obs}}} \zeta_i^{T_{obs}}(t_i,g_i,x_i^{T_{obs}}) m_{i\to\zeta^{T_{obs}}}(t_i,g_i) m_{x_i\to\zeta^{T_{obs}}}(x_i)\right] \\ &+\log\left[\sum_{t_i,g_i}\chi^{T_{ext}}(t_i,g_i) m_{i\to\chi^{T_{ext}}}(t_i,g_i)\right] + \log\left[\sum_{t_i,g_i}\mathscr{G}(g_i) m_{i\to\mathscr{G}}(t_i,g_i)\right] \\ &+\log\left[\sum_{t_i,g_i}\gamma(t_i,g_i) m_{i\to\gamma}(t_i,g_i)\right] \\ &+\log\left[\sum_{g_i,t_i}\sum_{\{t_i^{(k)},t_{ki},g_i^{(k)}\}} \psi_i(t_i,g_i,\{t_i^{(k)},t_{ki},g_i^{(k)}\}_{k\in\partial i}) m_{i\to\psi_i}(t_i,g_i)\prod_{k\in\partial i} m_{k\to\psi_i}(t_i^{(k)},t_{ki},g_i^{(k)})\right]\Bigg\} \\ &+\sum_{(i,j)\in E:i<j}\log\Bigg[\sum_{\{t_i^{(j)},t_{ji},g_i^{(j)}\}}\sum_{\{t_j^{(i)},t_{ij},g_j^{(i)}\}} \phi_{ij}(\{t_i^{(j)},t_{ji},g_i^{(j)}\},\{t_j^{(i)},t_{ij},g_j^{(i)}\}) \\ &\times m_{i\to\phi_{ij}}(t_i^{(j)},t_{ji},g_i^{(j)}) m_{j\to\phi_{ij}}(t_j^{(i)},t_{ij},g_j^{(i)})\Bigg]. \end{aligned} \tag{4.40}$$

Contributions from the variable nodes are

$$\begin{aligned}\sum_i f_i = \sum_{i\in V}\Bigg\{ &\log\left[\sum_{t_i,g_i} m_{\psi_i\to i}(t_i,g_i) m_{\chi^{T_{ext}}\to i}(t_i,g_i) m_{\mathscr{G}\to i}(t_i,g_i) m_{\zeta_i^{T_{obs}}\to i}(t_i,g_i) m_{\gamma\to i}(t_i,g_i)\right] \\ &+\sum_{j\in\partial i}\log\left[\sum_{\{t_i^{(j)},t_{ji},g_i^{(j)}\}} m_{\phi_{ij}\to i}(t_i^{(j)},t_{ji},g_i^{(j)}) m_{\psi_i\to i}(t_i^{(j)},t_{ji},g_i^{(j)})\right]\Bigg\}. \end{aligned} \tag{4.41}$$

Finally the contributions (to be subtracted) from edges connecting variable and factor nodes

$$\begin{aligned}
\sum_{(ia)} f_{(ia)} = \sum_{i\in V} \Bigg\{ & \log\left[\sum_{t_i,g_i} m_{\zeta_i^{T_{obs}}\to i}(t_i,g_i)\, m_{i\to\zeta^{T_{obs}}}(t_i,g_i)\right] \\
& + \log\left[\sum_{t_i,g_i} m_{\zeta_i^{T_{obs}}\to x_i}(t_i,g_i,x_i^{T_{obs}})\, m_{x_i\to\zeta^{T_{obs}}}(t_i,g_i,x_i^{T_{obs}})\right] \\
& + \log\left[\sum_{t_i,g_i} m_{i\to\chi^{T_{ext}}}(t_i,g_i)\, m_{\chi^{T_{ext}}\to i}(t_i,g_i)\right] \\
& + \log\left[\sum_{t_i,g_i} m_{i\to\gamma}(t_i,g_i)\, m_{\gamma\to i}(t_i,g_i)\right] \\
& + \log\left[\sum_{t_i,g_i} m_{i\to\mathscr{G}}(t_i,g_i)\, m_{\mathscr{G}\to i}(t_i,g_i)\right] + \log\left[\sum_{g_i,t_i} m_{i\to\psi_i}(t_i,g_i)\, m_{\psi_i\to i}(t_i,g_i)\right] \\
& + \sum_{k\in\partial i} \log\left[\sum_{\{t_i^{(k)},t_{ki},g_i^{(k)}\}} m_{\psi_i\to k}(t_i^{(k)},t_{ki},g_i^{(k)})\, m_{k\to\psi_i}(t_i^{(k)},t_{ki},g_i^{(k)})\right] \\
& + \sum_{j\in\partial i} \log\left[\sum_{\{t_i^{(j)},t_{ji},g_i^{(j)}\}} m_{i\to\psi_i}(t_i^{(j)},t_{ji},g_i^{(j)})\, m_{\psi_i\to i}(t_i^{(j)},t_{ji},g_i^{(j)})\right] \\
& + \sum_{j\in\partial i} \log\left[\sum_{\{t_i^{(j)},t_{ji},g_i^{(j)}\}} m_{\phi_{ij}\to i}(t_i^{(j)},t_{ji},g_i^{(j)})\, m_{i\to\phi_{ij}}(t_i^{(j)},t_{ji},g_i^{(j)})\right] \Bigg\}
\end{aligned} \tag{4.42}$$

4.6 Non-Markovian models

In Section 1.7 we pointed out that realistic epidemic models often have parameters changing in time and in this case they cannot be represented as Markov processes. The Belief Propagation approach proposed does not make any preliminary assumption on the recovery and transmission delays, therefore it can be adapted to time varying epidemic parameters. In most cases it is assumed that the infection probability between two individuals changes depending on the time *after* the infection: $\lambda_{ij}(t-t_i)$. In this case only the term for the transmission delays (4.4) is affected.

Thus

$$\omega\left(s_{ij}|g_i\right) = \begin{cases} \lambda_{ij}(s_{ij})\prod_{r=0}^{s_{ij}}(1-\lambda_{ij}(r))^r & s_{ij} \leq g_i \\ \sum_{s>g_i}\lambda_{ij}(s)\prod_{r=0}^{s}(1-\lambda_{ij}(r))^r & s_{ij} = \infty. \end{cases} \tag{4.43}$$

As a consequence we must change consequently the update rules in Section 4.2.4, by adapting equation (4.31) as follows:

$$\chi(t_1,t_2,\sigma,g) = \sum_{t=t_1}^{t_1+g}\delta(\sigma(t_2,t),\sigma)\lambda(t-t1)\prod_{r=0}^{t-t1}(1-\lambda(r))^r. \tag{4.44}$$

Another case is represented by a SIR model in which the recovery delay is fixed with value G. The distribution is changed as

$$\mathscr{G}(g_i) = \begin{cases} 0 & \text{if } g_i < G \\ 1 & \text{if } g_i = G. \end{cases} \tag{4.45}$$

In this case the model can be reduced to an susceptible-infected model, where the marginal probability of being recovered at time t is obtained by the marginal probability of being infected at time $t-G$: $P(x_i^t = R) = P(x_i^{t-G} = I)$. In this case the messages $m(t_i,g_i)$ can be casted in $m(t_i)$ and $m(t_i^{(j)},t_{ji},g_i^{(j)})$ becomes $m(t_i^{(j)},t_{ji})$. Then it is not necessary to consider the trace over variable g_i in the update rules in 4.2.4. The compatibility factor for the observation becomes:

$$\begin{aligned}\zeta_i^{T_{obs}}\left(t_i,x_i^t\right) =& \mathbb{I}\left[x_i^{T_{obs}} = S, T_{obs} < t_i\right] + \mathbb{I}\left[x_i^{T_{obs}} = I, t_i \leq T_{obs} \leq (t_i+G)\right] \\ &+ \mathbb{I}\left[x_i^{T_{obs}} = R, T_{obs} > t_i + G\right].\end{aligned} \tag{4.46}$$

Part II

Results

Chapter 5

Predicting epidemic evolution from partial observations

In this chapter we show the results obtained by using the Belief Propagation approach introduced in Chapter 4. We address the realistic situation in which the epidemic forecast is performed at some time after the initial infection event, when a number of infected cases is discovered in the population, and the information available involves only a fraction of the total population [62]. In this case the intrinsic stochasticity of the process makes difficult to predict the evolution of the epidemics. In order to focus only on the effects of partial observation, we assume that the structure of the contact network is completely known.

In other words, we assume that at time T_{obs} the state $x_i^{T_{obs}}$ is made available for a set of nodes $i \in V_{obs} \subset V$ and no information about the state is supplied for the nodes not in V_{obs}. For a given value of the infection parameters and a given initial condition $\mathbf{x}^0 = \{x_i^0\}_{i \in V}$ a huge number of different realizations of the stochastic process exists, although some of these outcomes are more likely to occur than others. We aim to provide probability distributions for these outcomes. We will focus on the probability marginals (Section 5.3) for the states of individual nodes $P(x_i^t)$ at any time and on the probability distribution for time at which the epidemic process dies out (i.e *extinction time*) $P(T_{ext} = t)$ (Section 5.4).

5.1 Sampling methods

Since the SIR stochastic process is Markovian, when all nodes in the network are observed ($V_{obs} = V$, complete observation), it is possible to estimate the probability of the future states $\mathbf{x}^t$ for $t > T_{obs}$ performing a *direct sampling* of a large number of epidemic realizations. It is necessary to generate a large number M_s of virtual realizations of the Markov chain from the same initial conditions (a complete observation at T_{obs}), then one can compute the probability of an event from its relative frequency of occurrence. In particular the individual probability marginal $P(x_i^t = X|\mathbf{x}^{T_{obs}})$ can be estimated from the experimental average

$$\hat{P}(x_i^t = X|\mathbf{x}^{T_{obs}}) = \frac{1}{M_s}\sum_{\ell=1}^{M_s}\mathbb{I}\left[x_{i,\ell}^t = X\right] \tag{5.1}$$

where $x_{i,\ell}^t$ is the value of the variable i at time t in the ℓ-th realization of the stochastic process from the same initial conditions. The probability distributions (5.1) converge to the correct value with a standard deviation that decreases as $\propto 1/\sqrt{M_s}$. In the case of an incomplete observation $V_{obs} \subset V$ we do not know the configuration of states $\mathbf{x}^{T_{obs}}$, thus the uncertainty about the future evolution of an epidemic state is much larger. For example, Fig. 5.1 shows five different evolutions of the epidemic process after the same partial observation. In order to apply the direct sampling method to the realistic case of a partial observation we first need a way to complete the missing information about the configuration $\mathbf{x}^{T_{obs}}$. We consider two simple ways to choose the states of unobserved nodes at T_{obs}:

- *random sampling*: given the incomplete observation of the system, the states of unobserved nodes at time T_{obs} are drawn randomly, independently and uniformly with the same probability $1/q$ where q is the number of possible states (for the SIR $q = 3$).

- *density sampling*: given the incomplete observation of the system, the fraction of observed nodes in each state $X \in \{S, I, R\}$ at time T_{obs} is used as an empirical probability to assign, independently and uniformly at random, the state of the unobserved nodes. The method can be generalized to include dependence on node attributes, such as the degree, by assigning to the unob-

served nodes a state with a probability computed from the knowledge of the states of observed nodes with the same attributes.

In both cases once we get a set of initial condition we use the direct sampling to find the probability distributions $P(x_i^t)$. It is worth remarking that unlike the case of complete information, the estimators obtained from these methods through direct sampling have non-zero bias.

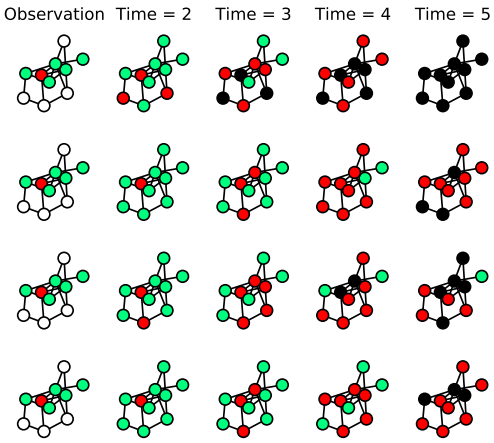


Figure 5.1: Each line represents a different realization for the SIR epidemic process given the (same) incomplete observation of the initial condition. Configurations in the leftmost column represent the observed state of the system, the other columns represent the time evolution of the epidemic process in that specific realization. Nodes colors: Green = Susceptible, Red = Infected, Black = Recovered, White = Unobserved.

We call *similarity sampling* an approximate sampling method to estimate $P\left(\mathbf{x}^t|\mathbf{x}^{T_{obs}}\right)$ inspired by the Soft-Margin algorithm [24] that we introduced in Section 2.2 and that was used to infer the epidemic origin. The similarity sampling method consists in evaluating $P\left(\mathbf{x}^t, \mathbf{x}^{T_{obs}}|\mathbf{x}^0\right)$ by computing the frequency of occurrence of each configuration over a large number of epidemic realizations. Each realization contributes to the posterior probability on the basis of the similarity to the observed states at T_{obs}. Every node in the set of infected and removed nodes at the time of observation T_{obs} is used as single seed for a given number of realizations. We include unobserved nodes if they have at least one not susceptible

neighbor (otherwise they cannot be the seed for sure). In case of unknown initial time we assume to know that the epidemics started within a ΔT_0 of time steps, i.e computing marginal probabilities we consider realizations with origin in a range $[-\Delta T_0, \Delta T_0]$. The similarity between a generic realization $\hat{\mathbf{x}}$ and the real one $\mathbf{x}$ is measured by computing the Jaccard similarity function described in Section 2.2. Then the individual marginal probability computed by similarity sampling reads:

$$\hat{P}(\hat{x}_i^t = X|\mathbf{x}^{T_{obs}}) \propto \sum_{\ell=1}^{M_s} \mathbb{I}\left[\hat{x}_{i,\ell}^t = X\right] e^{-(\phi(\hat{\mathbf{x}},\mathbf{x})-1)^2/a^2}. \tag{5.2}$$

According to the Soft-Margin estimator method, we use a convergence criterion to set the number M_s of realizations: the maximum of the differences between individual marginals after M_s and $M_s/2$ realizations is smaller than 0.1, i.e $\max(|P_{M_s}(x_i) - P_{\frac{M_s}{2}}(x_i)|) < 0.1$. Then, we choose the smallest value of a at which the convergence criterion is met. The accuracy of such a method usually relies on fine-tuning of the parameters and requires a very high computational power. Therefore, in all cases we initially set $a = 0.125$. If the convergence criterion is not met for $M_s \leq 8 \times 10^5$, then we use $a = 0.5$, that guarantees convergence for any instance. We must stress that this criterion is evaluated for the marginal probability to be in each of the three states at each time and for each node, then a large number of conditions must be satisfied and convergence is harder to achieve. Instead, in the case of Soft-Margin estimator the criterion applies only on the marginal probability of a single node to be infected at time zero.

5.2 Computational Cost

The similarity sampling computational cost is $\mathcal{O}(n \times M_s \times \bar{N}_I^T \times \bar{k} \times T)$. It is proportional to the time T over which we predict the epidemics process and the average number of infected node $\bar{N}_I^T$ up to time T, the number of potential sources in the observation n, the number of virtual realizations M_s and the average degree $\bar{k}$.

In the case of the Belief Propagation approach, the computational cost is $\mathcal{O}(N \times \bar{k} \times T \times G)$, where N is the number of nodes and G is the maximum recovery delay. This computational cost is fixed when BP converges, although the

convergence is only guaranteed on trees. On the contrary, Similarity Sampling computational cost depends on the number of virtual realizations of the process to meet the convergence criterion, which is not known a priori.

5.3 Epidemic Size and individual classification

In order to validate the proposed technique, we compute, for every node i, the marginal probabilities $P(x_i^t = X|\mathbf{x}^{T_{obs}})$ with $X \in \{S,I,R\}$ and compare with the methods proposed in the previous section. In order to quantify the performances of a prediction method a relevant measure is the correct classification of the epidemic stage for each individual, in particular the problem of discriminating whether an node has been reached by the disease at a time $t' \leq t$. This can be achieved by the use of a Receiver Operating Characteristic (ROC) curve [32, 33]. We build the ROC curve in the following way. Firstly we rank the nodes in decreasing order of magnitude of the probabilities $P(x_i^t = I|\mathbf{x}^{T_{obs}}) + P(x_i^t = R|\mathbf{x}^{T_{obs}})$. Given the ordered list of nodes and starting from the origin of the axes, we move upward by one unit whenever a node is correctly classified as already infected at time t (true positive) or rightward in case it is not (false positive). The *area under the ROC curve* (AUC) expresses the probability that a randomly chosen node infected before time t is ranked higher in terms of the corresponding probability marginal than a randomly chosen susceptible one. When the ranking is equal to the real one, the area under the ROC is 1, whereas a completely random ordering gives an area equal to 0.5. The area under the ROC curve gives indication of the fraction of the correctly classified nodes, but it does not have a clean dependence on the actual values of the marginal probabilities. The latter ones have instead a direct effect on the size of the epidemic outbreak, i.e. the number of nodes reached by the infection. Unlike the extinction time, the *average epidemic size* at time t can be expressed as function of the local marginals as [63]

$$size\,(t) = \frac{1}{N}\sum_i \left[P\left(x_i^t = I|\mathbf{x}^{T_{obs}}\right) + P\left(x_i^t = R|\mathbf{x}^{T_{obs}}\right)\right]. \tag{5.3}$$

In the following paragraphs we show the results obtained relatively to the quantity of interest. We studied the classification problem and the average size prediction

when the underlying network is either Random Regular graph or Barabási-Albert random graph.

Random Regular Graphs. In Fig.5.2 a set of results is displayed for random regular graphs of size $N = 1000$ nodes and degree $k = 4$, the information provided is an observation of a randomly chosen fraction of 10% of the nodes at $T_{obs} = 3$. We set the error on the initial time $\Delta T_0 = 1$. Fig.5.2 displays (a) the average values of AUC and (b) the average epidemic size as function of the time steps $t > T_{obs}$ for the different prediction methods: random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red). Results for the direct sampling with complete observation (black) are plotted as a reference, since they represent an upper bound for the quality of the prediction. The average values are computed on $M_o = 10^3$ independent realizations of the epidemic process, each one providing an observation at the same time T_{obs}. For each observation, the direct sampling algorithms are performed on $M_s = 2.5 \cdot 10^5$ realizations of virtual epidemic processes. As already noticed the similarity sampling method seldom converges in a number of realizations $M_s = 8 \cdot 10^5$ when $a = 0.125$, therefore most of the results are obtained using $a = 0.5$.

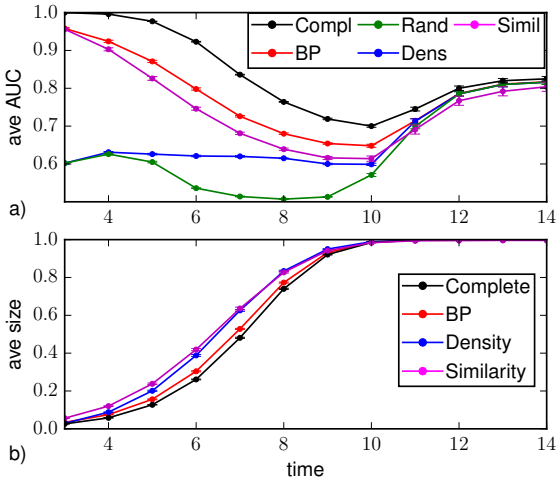


Figure 5.2: a) Area under the ROC curve as function of the time $t > T_{obs} = 3$ on a random regular graph of $N = 1000$ nodes and average degree $k = 4$. The average is computed over $M_o = 10^3$ epidemic realizations (with homogeneous parameters $\lambda = 0.7$, $\mu = 0.5$); the vertical bars represent the standard error of the mean. The prediction is obtained after the observation at T_{obs} of a 10%-fraction of the nodes chosen randomly (*random observation*). b) Predicted average epidemic size on random regular graphs ($N = 1000$, $k = 4$, $\lambda_{ij} = 0.7$, $\mu_i = 0.5$) as function of time for a random observation of 10% of the nodes at $T_{obs} = 3$. The inference methods used are direct sampling with complete observation (black), random sampling (green), density sampling (blue), similarity sampling (magenta) and belief propagation (red).

In a stochastic process as the SIR model it is not possible to predict exactly the state of the nodes, rather a distribution probability can only be provided. This leads to possible errors in the classification of the nodes (AUC values smaller than 1), even in the case of direct sampling with complete observation. In fact, the corresponding AUC values decrease after the observation time and recover only at late times since the epidemics dies out and almost all nodes are either in R or S states. We can interpret the average values of the area under the ROC as a proxy for the epidemic predictability, in the case of a complete observation (best-case scenario) the behavior observed is compatible with the effects due to epidemic heterogeneity reported in [64]. The process becomes more difficult to predict at intermediate time, when a large fraction of the population have non-zero probability to be in each of the three epidemic states. The Belief Propagation

technique with partial observation gives values of averaged AUC that are closer to those from complete observation than the other methods (Fig.5.2a). BP and similarity sampling perform largely better in the first stage after the observation, corresponding to the exponential outbreak phase [15]. In particular, similarity sampling gives an AUC value similar to BP at the time of observation, but a lower AUC value in the subsequent time steps.

In Fig. 5.2b results for the average epidemic size are displayed. The density sampling strongly overestimates the average epidemic size with respect to results from complete observation; this is probably an effect of the homogeneous deployment over the graph of infected nodes used to complete the information on the configuration at T_{obs}, that leads to the prediction of a larger epidemic spreading. In other words density sampling disregards existing correlations between the 90% of the nodes, taking in consideration realizations possibly not compatible with the observation. This scheme could lead to the overestimation of the probability of being infected – in a similar way to mean field approximations. In similarity sampling the overestimation of the epidemic size is likely due to two concurrent causes: the procedure itself because takes into consideration the M_s virtual epidemic realizations even if they do not match the actual observation with a Gaussian weight; and to the approximation on the initial time. Belief Propagation also slightly overestimates the epidemic size, but we think this is essentially due to the fact that in most of the instances the algorithm does not properly converge to the correct marginals. Moreover the underlying graph is loopy and it leads to an overestimation of the infection probability (see Section 1.5).

The heat-plots in Fig.5.3 display the same set of data arranged as function of the number of observed nodes that were infected before the observation time (on the horizontal axes), respectively for density sampling, similarity sampling and belief propagation. Results for direct sampling with complete observation are presented as a reference. In the case of the average AUC (Fig.5.3), BP performs better than both density sampling and similarity sampling in all regimes, in particular the performance is very good in the first steps after the observation, almost independently of the actual number of infected and recovered nodes in the observation. For all methods the results slightly improve when a larger number of nodes reached by the epidemics is observed at T_{obs}. For the average epidemic size,

Fig. 5.3b shows that the early-stage prediction by density sampling is negatively affected by the observation of a larger number of infected and recovered nodes. The opposite occurs, though to a lesser extent, for BP: when few infected nodes are observed BP overestimates the epidemic size, the worst prediction by BP giving an average size about 20% larger than that obtained from complete observation. The deviation observed by similarity sampling is also more evident when a lower number of infected and recovered nodes is observed, but the overestimation is more homogeneously distributed. Interestingly, the poor performance at large times is localized on realizations in which only a few of the observed nodes already got infected at T_{obs}. This may be due to the fact that the observation does not provide enough information to make more precise predictions.

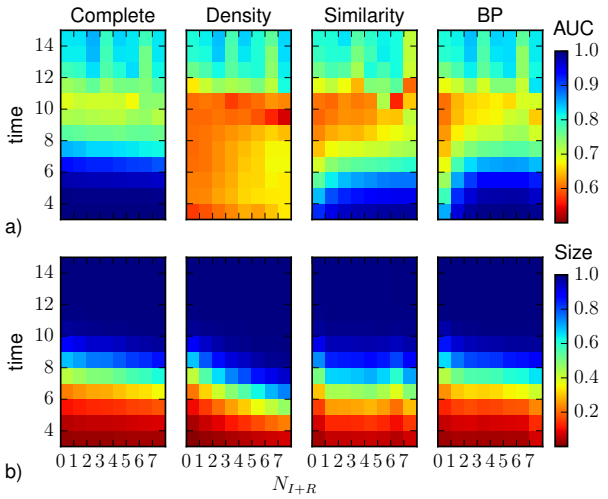


Figure 5.3: a) The heatplots represent the average AUC as function of time and of the number of observed nodes that were infected before T_{obs}, computed by density sampling, similarity sampling, belief propagation. b) The average epidemic size predicted by density sampling, similarity sampling and belief propagation is also shown as function of the number of infected and recovered nodes in the observation. As a reference, in both panels, we plot results obtained, for the same realizations of the SIR process, by direct sampling with complete observation. The horizontal axis refers to the number of infected or recovered nodes present in the 10% observation (also in the case of complete observation).

Barabási-Albert random graph. In the case of heterogeneous graphs, such as those obtained with the Barabási-Albert (BA) growing network model, in addiction

to the *random observation*, it is possible to define other interesting observation schemes for the same density of observed nodes:

- *degree-based observation*: nodes are observed in descending order of their degree;

- *local observation*: a connected cluster of observed nodes is generated from a randomly chosen infected node.

We investigated the effect of different observation schemes on random sampling, density sampling, similarity sampling and BP.

The results for the average AUC, obtained with observation of 30% of the nodes at $T_{obs} = 3$, are reported in Figs. 5.4-5.5. In the case of complete observation, direct sampling produces monotonically decreasing AUC values for increasing times. The reason is that in finite size networks the parameters chosen give a non-zero probability of finding susceptible nodes in the last stage of the epidemic evolution. Therefore, it is possible to make a wrong prediction and the AUC remains considerably below one. For random observation, Fig.5.4a shows that Belief Propagation always gives larger AUC values than the other sampling methods, especially in the first stage of the epidemics, i.e. during the exponential outbreak. The same behavior is found plotting the results as function of the actual number of observed nodes (see heat-plots in Fig.5.5a) that were already infected/recovered at the time of observation; in particular, the performances of BP are better when this number is small.

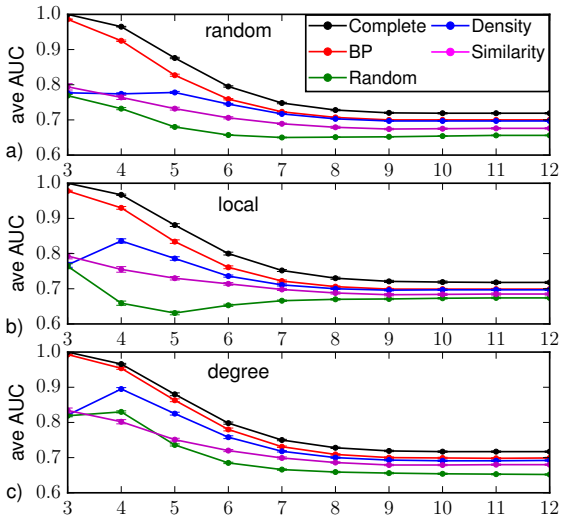


Figure 5.4: Area under the ROC curve as function of the time $t > T_{obs} = 3$ on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ (with homogeneous epidemic parameters $\lambda = 0.5$, $\mu = 0.6$), in the case of observation of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The average is computed over $M = 201$ epidemic realizations. The inference methods used are direct sampling with complete observation (black), random sampling (green), density sampling (blue), similarity sampling (magenta) and belief propagation (red).

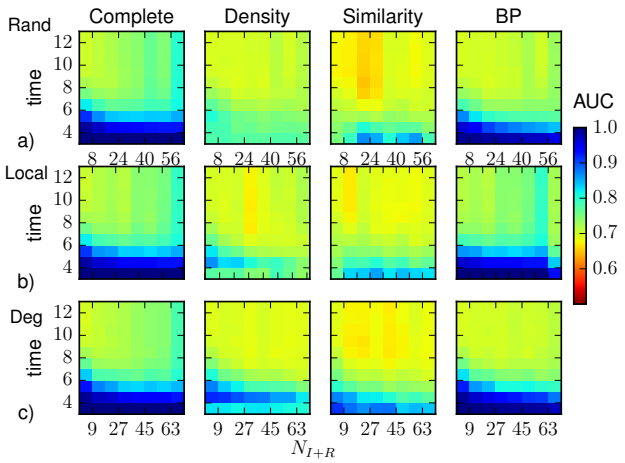


Figure 5.5: The heatplots represent the average AUC as function of time and of the number of observed nodes that were infected before $T_{obs} = 3$, computed by density sampling, similarity sampling, belief propagation, on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ with homogeneous parameters $\lambda = 0.5$, $\mu = 0.6$. As a reference, we also plot results obtained, for the same realizations of the SIR process, by direct sampling with complete observation. The prediction is obtained after the observation at T_{obs} of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The horizontal axis refers to the number of infected or recovered nodes present in the 30% observation (also in the case of complete observation).

When the observation is provided for a 30%-fraction of nodes forming a connected subgraph the overall results (Fig. 5.4b-5.5b) are similar to those with random observation, even though density sampling and BP perform slightly better in the time steps immediately after the observation, while similarity sampling is slightly worse in the same regime. It turns out that a degree-based observation is particularly convenient for heterogeneous networks. In fact, we found that the average values of the ROC area increase in the first stage of the epidemics for all prediction methods (Fig.5.4c). In particular, the difference between values obtained by BP and those from direct sampling with complete observation is less than 2%. The results for the AUC in function of the number of infected/recovered nodes observed (see the heat-plots in Fig.5.5c) are qualitatively similar to those from the other observation schemes. However we notice slightly better prediction performances at early times, when the number of infected/recovered nodes in the observation is small.

Since the information provided involves a connected subgraph generated from an infected node, the case in which few nodes are detected as infected/recovered likely corresponds to smaller epidemics whose initial evolution is more predictable.

Results on the prediction of the average epidemic size are reported in Fig. 5.6 and Fig.5.7. Random sampling strongly overestimates the size in all regimes and observation schemes, then we omit the relative results. With a random observation scheme (see Fig. 5.6a), density sampling and BP provide very accurate prediction along the whole dynamics, while similarity sampling provides strong overestimate of the size value at early time and underestimate at late times. Density sampling and BP provide a lower accuracy when a small number of infected and recovered nodes is observed (Fig.5.7a). When the number of nodes reached by the infection at T_{obs} is larger, BP performs better than density sampling (4.5% of the nodes larger than the direct sampling with complete observation). Similarity sampling provides a bad estimation at late times. The heat-plot in Fig.5.7a suggests that it is mostly due to a very strong underestimation of the average size, when the observation contains only few infected/recovered nodes. On the contrary, at early times, overestimate of the average size occurs when a large number of infected/recovered nodes is observed. The cause may be the following: similarity sampling actually does not fix the information provided by the observation, but uses it to assign a Gaussian weight whose expected value is the observed configuration. Then, if the observation contains few infected node, similarity sampling assigns weights with respect to an expected configuration that contains less infected/recovered nodes than the actual epidemic process.

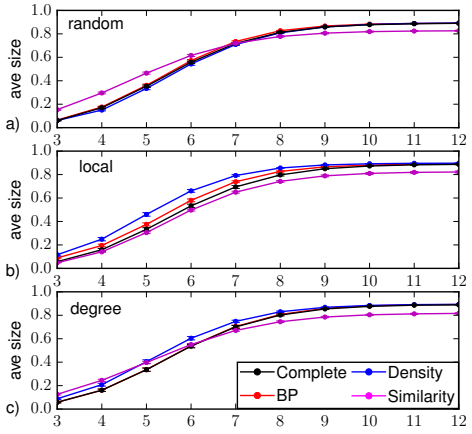


Figure 5.6: Predicted average epidemic size as function of the time $t > T_{obs} = 3$ on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ (with homogeneous epidemic parameters $\lambda = 0.5$, $\mu = 0.6$), in the case of observation of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The average is computed over $M = 201$ epidemic realizations. The inference methods used are direct sampling with complete observation (black), density sampling (blue), similarity sampling (magenta) and belief propagation (red).

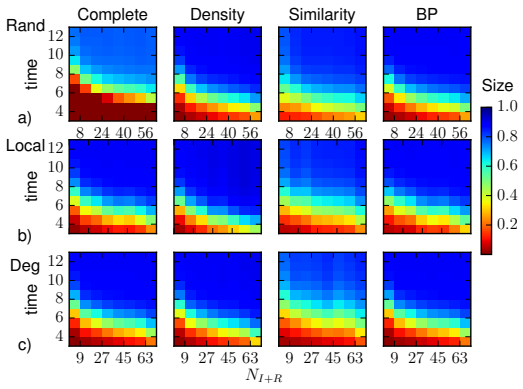


Figure 5.7: The heatplots represent the average epidemic size as function of time and of the number of observed nodes that were infected before $T_{obs} = 3$, computed by density sampling, similarity sampling, belief propagation, on a Barabási-Albert random graph of $N = 1000$ nodes and average degree $\langle k \rangle \approx 4$ with homogeneous parameters $\lambda = 0.5$, $\mu = 0.6$. As a reference, we also plot results obtained, for the same realizations of the SIR process, by direct sampling with complete observation. The prediction is obtained after the observation at T_{obs} of a 30%-fraction of (a) nodes chosen at random uniformly and independently, (b) nodes forming a connected subgraph, (c) the most connected nodes. The horizontal axis refers to the number of infected or recovered nodes present in the 30% observation (also in the case of complete observation).

In Fig.5.6b we show the prediction of the average epidemic size when the partial observation is performed considering a connected subgraph of 30% of the nodes. In this case all methods overestimate the epidemic size, with BP performing considerably better than the others. The poor performances of density sampling are expected because it completely neglects the topological information in the observation. For example, if infected nodes are surrounded by susceptible ones, the probability of infection for unobserved nodes is lower, but this is not taken into account in the density sampling approach. BP performs instead poorly when there are very few infected nodes in the observed area. This is expected, because in such a situation this method is not able to correctly reconstruct missing information. Finally, similarity sampling gives good results for small and intermediate time steps but again it strongly deviates at large times, mostly because of observations with few infected/recovered nodes (see Fig.5.7b).

In the case of the degree-based observation, BP predicts very accurately the epidemic size (we already noticed BP good performances for AUC), see Fig.5.6c. The difference between the average epidemic size predicted by Belief Propagation and the one obtained by direct sampling with complete observation is less than 2% of the nodes in the network. Instead, density sampling overestimates the average epidemic size, especially in the first epidemic outbreak and for a large number of infected and recovered nodes in the observation (see Fig.5.7c). Density sampling does not make use of the connectivity knowledge, which is a valuable information: an observed highly connected node is more likely to be infected, ignoring this fact leads to assign the same infection probability of the hubs to every node in the network, leading to larger predicted epidemic sizes. In this respect, one could expect that better results could be obtained simply by introducing a degree-dependence in the infection probability inferred from the observation; nevertheless, preliminary results show no significant improvement in the quality of the prediction.

5.3.1 A case study of real contact network

We consider a real network dataset of the sexual encounters of internet-mediated prostitution previously studied by Luis E. C. Rocha, Fredrik Liljeros and Petter Holme [45, 65], that was obtained analyzing a Brazilian web community exchang-

ing information between male sex buyers. The original dataset is in the form of a bipartite temporal network, in which an edge between a "sex buyer" A and "sex seller" B is drawn if A posted a comment in a thread about B. The dataset covers the period September 2002 to October 2008 (2,232 days) and 50,185 contacts are recorded between 6,642 sex sellers and 10,106 sex buyers. In our analysis, we do not consider separate classes of vertices and we focus on a sample network comprising a time window between day 1000 and day 1100. The resulting network (SC) has $N = 1293$ nodes, $E = 1571$ edges, average degree $\langle k \rangle \approx 2.4$ and maximum degree $k_{max} = 55$.

We study the AUC and the prediction for the average epidemic size on a static projection of the sexual contact network when the observation takes place at times $T_{obs} = 4, 8$ as representatives of early and later time observation. In both cases, density sampling and random sampling make unreliable predictions of the classification of individual states of the nodes (see Fig.5.8a-5.8c). For $T_{obs} = 4$, BP gives good results only in the time steps immediately after the observation, then the performances rapidly deteriorate. BP results slightly improve increasing the observation time. Nevertheless BP is better than other methods. For the average epidemic size, Fig. 5.8b shows that similarity sampling gives the best prediction at $T_{obs} = 4$ (though underestimating the epidemic size), whereas BP performs as bad as density sampling (and random sampling even worse). BP results improve considerably for $T_{obs} = 8$ while similarity sampling turns out to overestimate the epidemic size at time steps close to the observation (Fig. 5.8d).

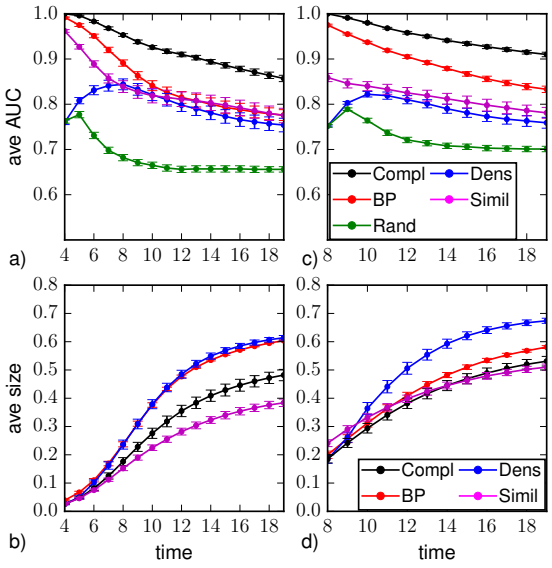


Figure 5.8: Average area under the ROC curve (a,c) and average epidemic size (b,d) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.5$, $\mu = 0.4$) on the SC network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 4$ (a,b) and $T_{obs} = 8$ (c,d). In all plots direct sampling from a complete observation is shown for comparison (black).

As we noticed in Section 5.3 this results can be strongly influenced by the number of infected and recovered nodes in the observation. In Fig. 5.9 we repeat all measurements considering observations at $T_{obs} = 4$ containing a number of infected and recovered nodes equal to $N_{I+R} \geq 6$ (corresponding to the 46% of all instances), and at $T_{obs} = 8$ with $N_{I+R} \geq 18$ (75% of all instances). BP performances improve considerably at $T_{obs} = 4$, outperforming all other methods in the case of $T_{obs} = 8$. These results can be better understood if we consider that the network is characterized by a well connected core surrounded by many low degree nodes. When few infected nodes are observed, they typically are low-degree ones and the epidemic process spreads slowly at early time. In this situation similarity sampling is facilitated because the trajectories leading to the observed states are a small set. Therefore, the observation provides a good expected configuration for the Gaussian weights and smaller biases are introduced (see Section 5.3). On the contrary, it is less accurate when many infected nodes are observed or the

observation occurs at later times. However, results can likely be improved by an higher computational power and algorithmic optimization. Belief propagation is more likely to overestimate the nodes probability of being infected when the size of epidemics is small at the time of observation. In fact, in this case the effect of the existence of short loops in the network is more important and BP is more likely to overestimate [30]. It is worth noting that we provide to similarity sampling the information about the initial time $t = 0 \pm \Delta T_0$ of the epidemic spreading, on the contrary, we do not provide such an information to BP.

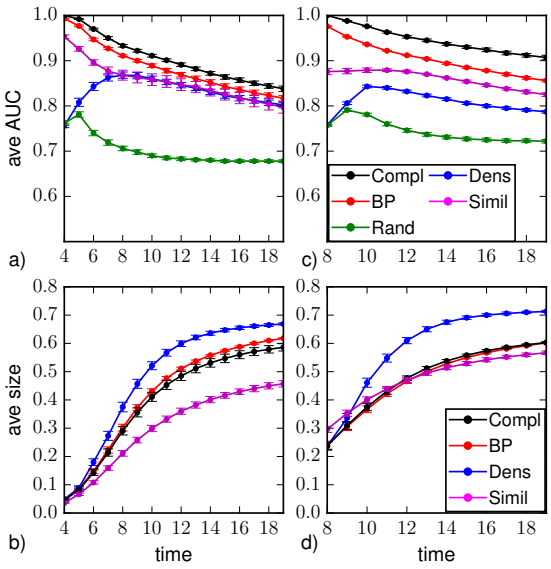


Figure 5.9: Average area under the ROC curve (a,c) and average epidemic size (b,d) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.5$, $\mu = 0.4$) on the SC network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 4$ (a,b) and $T_{obs} = 8$ (c,d). For $T_{obs} = 8$, only instances with a number of observed infected and recovered nodes $N_{I+R} > 18$ is considered (75% of instances). For $T_{obs} = 4$, only instances with observed infected and recovered nodes $N_{I+R} > 6$ is considered (46% of instances). In all plots direct sampling from a complete observation is shown for comparison (black).

In order to keep memory of information about the time evolution of the contact pattern between individuals we also consider a weighted static projections (see

Section 1.6) of the sexual contact network (WSC). We assign a weight w_{ij} to every existing edge ij that corresponds to the number of contacts between node i and node j during the period under consideration. Then, we define the probability that node i infects node j as $\lambda_{ij} = 1-(1-\lambda)^{w_{ij}}$. By doing so, individuals sharing more contacts during the time interval under consideration have a larger probability of infecting each other. Fig.5.10 shows results for the average AUC and the average epidemic size. Belief Propagation provides higher values for the AUC than all other methods at all times, even though AUC decreases with time much faster compared to direct sampling with complete observation. At the time steps immediately after the observation, BP also provided the best prediction of the average epidemic size, while at late times similarity sampling works better.

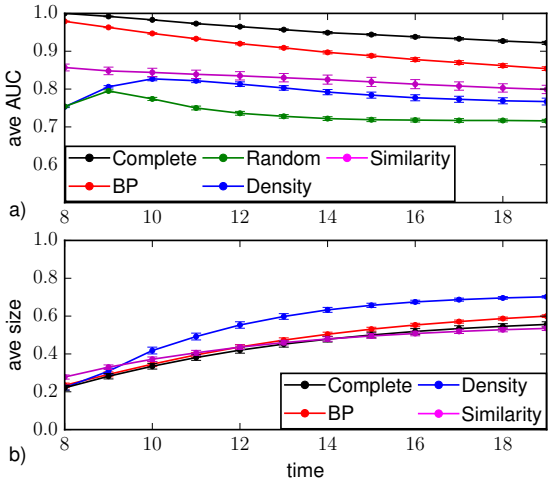


Figure 5.10: Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.5$, $\mu = 0.4$) on the WSC network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 8$. Direct sampling from a complete observation is shown for comparison (black).

5.4 Extinction Time probability distribution

The results that we show here concern the probability distribution for the extinction time when a (possibly partial) observation is provided. Even in the case of complete

observation, the results are highly non-trivial, in particular on networks with peculiar topological structure. In Fig. 5.11 we show extinction time distributions $P_{ext}(t) = P\left(t = T_{ext}|\mathbf{x}^{T_{obs}}\right)$ given different observed configuration. The underlying networks are regular trees (a) and regular random graphs (b). In the case of trees the probability distribution is highly variable: depending on the observation, the width and the maximum value of the distribution can change significantly. This effect is mostly due to the peculiar topology of trees. This can be explained focusing on Figs. 5.11c-5.11e. They show the configuration of three different realizations at the time of observation T_{obs}. In terms of the number of infected node and their average degree, the snapshots in panels (c) and (e) are similar, but their extinction time probability distributions are rather different (for example, they have the maximum respectively at $T_{peak} = 16$ and $T_{peak} = 23$). On the contrary, panels (c) and (d) show very different realizations at T_{obs}. Nevertheless, they have similar extinction time probability distributions ($T_{peak} = 23$ and $T_{peak} = 21$). This effect is related to the positions where infected and recovered nodes are observed at T_{obs}: the configuration in Fig.5.11e does not allow the disease to access the root of the tree, because a recovered individual interposes. Thus, for $t > T_{obs}$, the epidemics can spread only on the already infected branches and the diffusion to other regions of the graph is blocked. Conversely, in Figs.5.11c-5.11d the epidemics can spread throughout the graph and the distribution reaches the maximum at larger times. The heterogeneity of the extinction time distribution is peculiar of trees and graphs with topological bottlenecks, while random graphs, or graphs with small-world properties in general, are characterized by very similar distributions for different realizations of the epidemic process (with same epidemic parameters and observation).

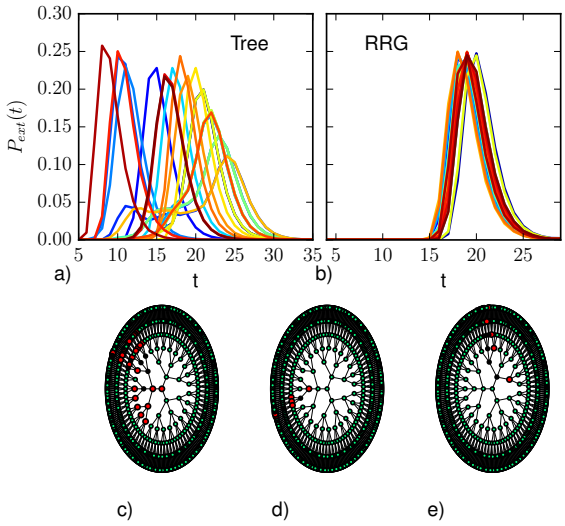


Figure 5.11: Extinction time distributions for different complete observations: a) on trees with branching ratio $k = 3$ and $N = 1092$ (epidemic parameters $\lambda = 0.7$, $\mu = 0.5$ and observation time $T_{obs} = 5$); b) random regular graphs of degree 4 and $N = 1000$ nodes ($\lambda = 0.7$, $\mu = 0.5$ and $T_{obs} = 4$). (c)-(e): similar realizations of the epidemic process at T_{obs} on a tree graph corresponding to rather different predicted extinction time distributions with maximum value respectively at $T = 21$ (c), $T = 23$ (d), and $T = 16$ (e). Nodes color: Green= Susceptible, Red= Infected, Black= Recovered.

The results for the extinction time probability distribution given a partial observations are displayed in Fig. 5.12. We first considered the case of regular trees of branching ratio equal to 4 (average degree $\langle k \rangle \approx 2$). We already noticed that trees provide strong variability of the extinction time distribution. The observation is made by sampling randomly the state of 10% of the nodes at $T_{obs} = 5$. Fig 5.12a displays the average difference between the extinction time distribution probability predicted using direct sampling with complete observation and that obtained using Belief Propagation (red), density sampling (blue), and similarity sampling (magenta). All methods present two regions of higher discrepancy with respect to the prediction with complete observation. The heat-plots in Fig.5.13b show that this is usually due to an underestimation of the probability of extinction in the early stage of propagation and to an overestimation of the probability of extinction at

large times. For example, let us consider a configuration similar to Fig.5.11c. If we do not observe the roots (that is in a recovered state), the computed extinction time probability distribution must take into account the probability that the disease diffuses throughout the network, then it underestimates the distribution with respect to the complete observation. Nevertheless, BP is usually able to qualitatively identify the most probable extinction time even when the other methods instead assign a higher probability to the disease vanishing at larger times. In fact, we know that BP overcomes the lack of information by taking into account correlations in the dynamics. Heat-plots show that the two-peak discrepancy is especially due to observations with few infected and recovered nodes, while the discrepancies between the distributions move mostly at intermediate times when this number is increased. BP performs better than the other methods at every time step, although it presents the same qualitative weaknesses. Interestingly, the similarity sampling method overestimates the probability for the epidemics to die out at early time step. The cause is intrinsic of the method: given the partial observation, the expected configuration in the Gaussian weight is set with a number of infected node smaller than the actual one. Then, it is more probable the epidemics with a high weight dies out shortly after T_{obs}. Thus, it leads to an overestimation of the extinction probability at early time steps.

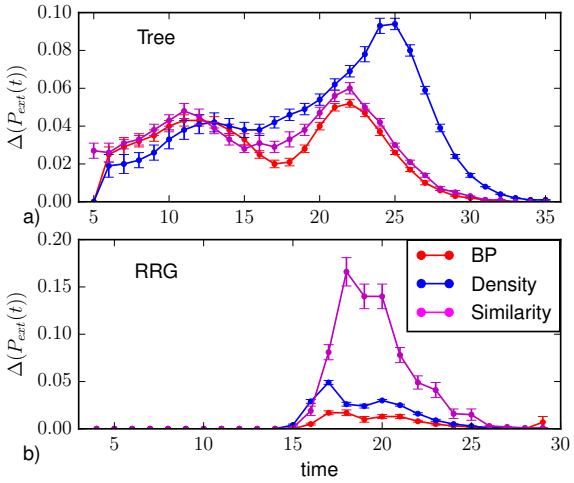


Figure 5.12: Absolute value of the difference between the extinction time distribution $P_{ext}(t)$ computed from direct sampling with complete information and those calculated with density sampling (blue), BP (red) and similarity sampling (magenta). a) On trees of $N = 1092$ nodes, with branching ratio 3 ($\langle k \rangle \approx 2$) and with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 10% of the nodes at $T_{obs} = 5$ and averaging over $M_o = 210$ such realizations. b) On random regular graphs of $N = 1000$ nodes and degree $k = 4$ with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 30% of the nodes at $T_{obs} = 4$ and averaging over $M_o = 150$ such realizations.

Fig. 5.12b displays the results for the average difference from the distribution provided by direct sampling with complete observation, in the case of random regular graphs of degree $k = 4$ with partial observation of the 30% of the nodes at $T_{obs} = 4$. Although all prediction methods under study are able to reproduce the existence of a unique peak, there are remarkable quantitative differences with the results from direct sampling with complete observation. The BP algorithm provides the best performances, in particular for observations with a large number of infected and recovered nodes. When this number is low, instead, BP gives a larger average difference with respect to density sampling. This effect is mostly due to the non-convergence of the BP algorithm in some instances of the epidemic process, leading to an overestimation of the probability of long extinction times. At the time steps close to the peak, the similarity sampling gives the largest average

difference. Fig. 5.13b shows that the main contribution to the average difference comes when low number of infected and recovered nodes are observed.

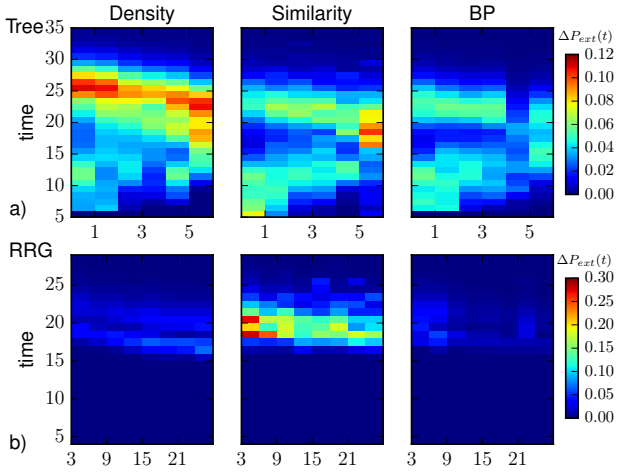


Figure 5.13: Absolute value of the difference between the extinction time probability distribution $P_{ext}(t)$ computed from direct sampling with complete information and those calculated with density sampling, BP and similarity sampling as a function of the number of infected and recovered nodes in the observed subset of nodes. a) On trees of $N = 1092$ nodes, with branching ratio 3 ($\langle k \rangle \approx 2$) and with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 10% of the nodes at $T_{obs} = 5$ and averaging over $M_o = 210$ such realizations. b) On random regular graphs of $N = 1000$ nodes and degree $k = 4$ and with uniform epidemic parameters $\lambda = 0.7$, $\mu = 0.5$. The partial observation is performed sampling uniformly the state of 30% of the nodes at $T_{obs} = 4$ and averaging over $M_o = 150$ such realizations.

Chapter 6

Generalization to Temporal Networks and Non-Markovian Models

When studying epidemic processes it is common to assume approximations to get models easier to deal with. In Sections 1.6 and 1.7 we introduced tzemporal networks and non-markovian epidemic models, that are two approximated attempts to catch two different features of realistic contagion processes. In Sections 4.3 and 4.6 we showed that the Belief Propagation approach can easily be adapted to study epidemic processes on time-varying network structures and non-Markovian epidemic models. In this chapter we will study the problem of predicting the epidemic evolution given incomplete information in these two cases. We will show some preliminary results.

6.1 Temporal Networks

We consider a real world network based on the contact patterns occurred in a high school in Marseille, France, collected by the means of the RFID (Radio-frequency identification) technology [66] and firstly studied by R. Mastrandrea, J. Fournet and A. Barrat [14]. The dataset records the face-to-face contacts (in the range of 1-1.5 meters) between individuals with a temporal resolution of 20 seconds occurred during the week of December 2-6, 2013. We take into account data collected

during the day in which the largest number of contacts has been recorded. We obtain a discrete time network (from here out called HN) by partitioning the time line into subintervals of length Δ (time-steps) and aggregating all contacts falling in the interval $[t\Delta,(t+1)\Delta]$. To keep memory of the number of contacts between any pair of individuals (i,j) recorded during a time step, we compute a time dependent probability of infection $p_{ij} = 1-(1-\lambda)^{c_{ij}}$, where c_{ij} is the number of contacts in the t-th time step [11, 32, 33].

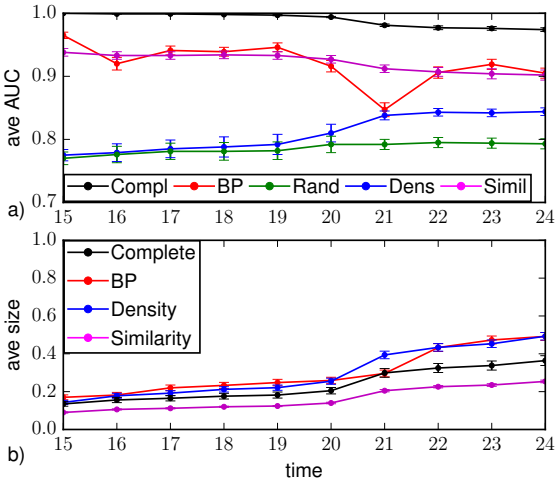


Figure 6.1: Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.6$, $\mu = 0.02$) on the HN. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 15$. Direct sampling from a complete observation is shown for comparison (black). The average is computed on $M = 50$ epidemic realizations.

We modify the Belief Propagation approach as explained in Section 4.3 and we use as comparison the direct Monte Carlo sampling methods described in Section 5.1. We partitioned the time interval in $T = 25$ time steps, we set the infection parameter $\lambda = 0.6$, the probability of recovery at each time step $\mu = 0.02$ and the observation at $T_{obs} = 15$. In Fig. 6.1b displays the average size predicted by the four methods. BP and the density sampling overestimate the average size, especially at late time. On the contrary, similarity sampling underestimates the average epidemic

size. It is worth noting that the structure of the temporal network under study slows down the spread of the disease. It is expected because causal relations must be taken into account. In fact, individuals spread the disease only through contacts occurred after their infection. Fig. 6.1a displays the AUC obtained by the four methods. The Monte Carlo sampling with complete observation perfectly classifies the state of the nodes until late time, when the AUC decreases. On the contrary, density and random sampling give the worst prediction and the AUC slowly increases only at late time. BP and similarity sampling provide good predictions at any time step and the AUC slowly decreases with time. Interestingly, the AUC obtained for the BP method drops at time step $t = 21$ in our partition of the total time interval. In Fig. 6.2, we show the number of contacts for each time step in our partition and it comes out that $t = 21$ is the time step with the largest number of contacts. Further investigation are necessary, however the drop in the BP accuracy is possibly due to the high number of contacts occurred at $t = 21$.

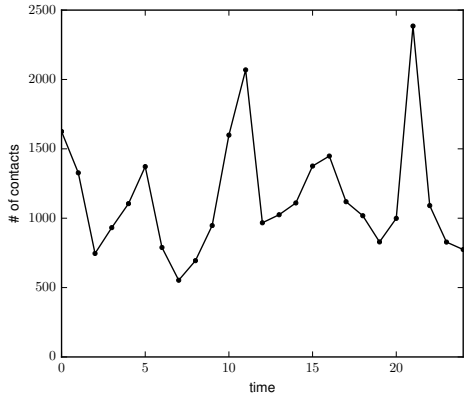


Figure 6.2: Number of contacts per time step in the High School network with a partition composed of $T = 25$ time steps.

6.2 Non-Markovian SIR model: fixed recovery delay

The assumption of a constant probability of recovering is not a good approximation for all diseases. For example, it is known that individuals infected by measles take a fixed time interval to recover. We modify the Belief Propagation approach

to study a discrete time SIR model in which the recovery occurs after a fixed number of time steps for each individual. As a comparison, we study the results of the direct Monte Carlo sampling methods described in Section 5.1. Since the model is not Markovian, providing only the states of the observed nodes is not a sufficient information for the direct sampling methods. In order to have a best-case benchmark, the Monte Carlo sampling is provided with a complete observation of the configuration at T_{obs} and is also provided with the time of infection of every node. On the contrary, in the density sampling and random sampling method only the states of the observed nodes are known; the time of their infection is unknown and it is randomly drawn from the interval $[T_{obs} - G, T_{obs}]$.

Fig. 6.3 displays the results for a SIR model in which we set the recovery at $G = 10$ time steps after the infection (the infection probability is $\lambda = 0.2$).

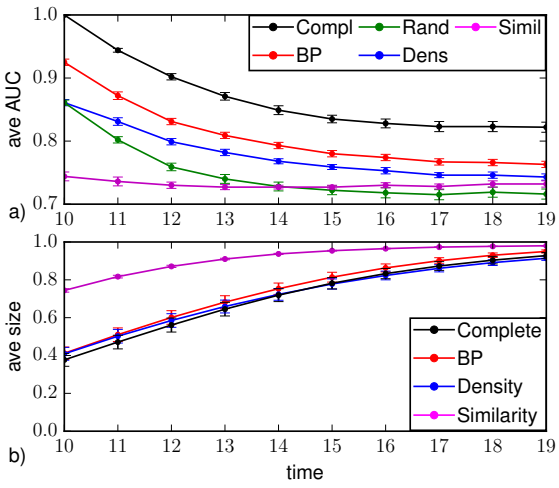


Figure 6.3: Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.2$, the recovery occurs after $G = 10$ time steps) on a BA network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 10$. Direct sampling from a complete observation is shown for comparison (black). The average is computed on $M = 50$ epidemic realizations.

In Fig. 6.3a results for the AUC are displayed. The AUC has a large value shortly after the observation and becomes smaller at later time. The pattern is

similar to the Markovian case studied in Section 5.3 on BA networks. The AUC provided by BP is the largest at any time, instead similarity sampling gives the worst prediction. The average size predicted by BP, random, and density sampling (Fig. 6.3b) is consistent at any time with the one predicted with the complete observation. Instead similarity sampling largely overestimates the average size at any time. We already pointed out in Section 5.3 that similarity sampling assumes as a typical configuration the observed one and samples realizations assigning a Gaussian weight consequently. This assumption leads to inaccurate predictions because the actual realization is just one of a huge number of possible configurations, that may be rather different because of intrinsic heterogeneity of epidemic processes. Selecting one of them as a reference point is like promoting a rare event to be considered as a typical one. This effect is expected to be larger when the observation is provided at intermediate time, because at this stage of the epidemic evolution there is an exponential number of possible configurations. Although the method can be improved by fine-tuning of the parameters, the cause of the inaccuracy is intrinsic.

6.3 Non-Markovian SIR: non-infective period

In some realistic cases, a symptomatic individual becomes able to infect some time after his infection. We consider a SIR model in which an infected individual shows the symptoms of the disease, but is not able to spread the infection for a fixed amount of time L. We provided the direct Monte Carlo sampling with the same information as in the previous section. For both the density sampling method and the random sampling method, the infection times of the nodes that are infected in the observation are drawn independently and uniformly at random in the interval $[T_{obs} - L, T_{obs}]$. Fig. 6.4 displays results for the described SIR in which we set $L = 1$ and epidemic parameters $\lambda = 0.6$, $\mu = 0.3$ on BA networks and $T_{obs} = 5$. Shortly after the observation, BP gives values of the AUC (Fig. 6.4a) similar to the ones provided by the direct Monte Carlo with complete observation. Predictions become less accurate at later times, remaining in line with the other methods. Similarity sampling also provides good predictions. The AUC computed by the density sampling improves with time, instead random sampling provides unreliable prediction at any time. Fig. 6.4b displays the results for prediction of

the average size. However, it is worth noting that the non-infective period leads to a slower spread, thus only few infected individuals are observed at T_{obs}. BP overestimates the epidemic size at any time step, especially at late time. Although we observed similar results on BA networks in Section 5.3 and it is known that a loopy network structure leads to overestimation on the marginal probability that a node is infected, further investigations are required.

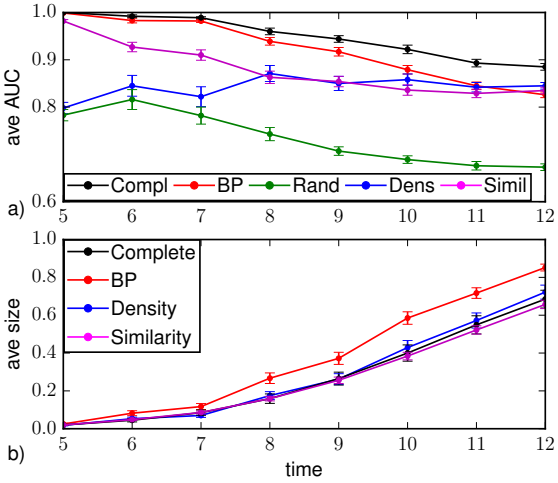


Figure 6.4: Average area under the ROC curve (a) and average epidemic size (b) as function to the time $t \geq T_{obs}$ for SIR dynamics ($\lambda = 0.6$, $\mu = .3$ and a node start infecting $L = 1$ time steps after the infection) on a BA network. Results are obtained with random sampling (green), density sampling (blue), similarity sampling (magenta) and Belief Propagation (red) from a random observation of 30% of the nodes at $T_{obs} = 5$. Direct sampling from a complete observation is shown for comparison (black). The average is computed on $M = 50$ epidemic realizations.

Conclusions

The prediction of the epidemic evolution in contact networks, from unknown initial conditions and partial observations, is an important issue of great interest for health-care organizations and policy makers. Even in the case of an irreversible stochastic processas the Susceptible-Infected-Recovered model, that gives a simple mathematical description for disease that cause permanent immunization or death (such as the Ebola virus, Measles or Smallpox), the computational problem associated with epidemic forecast from partial information is highly non-trivial.

In this thesis we tackled the problem by defining an accurate and efficient method to compute the marginal probability of the state of a node at any time. We presented a technique based on the Belief Propagation approximation, that is exact on trees and usually a good approximation on general graphs. This technique is based on the computation of a partition function in which the variables are temporal trajectories, on which constraints are imposed by the epidemic dynamical rule and by the observation data. The idea behind the method is that it is possible to compensate for the missing information by taking advantage of the reconstruction of the causal relations built during the dynamic evolution that preceded the observation time.

The results presented in this thesis show that Belief Propagation is in general accurate in the classification of the states of individuals and in the prediction of the average epidemic size even from a very partial information. The comparisons with simulations provided with complete information show that BP is actually able to compensate incomplete data by the inference of causality relations before the time in which the observation is performed. In fact, this is especially verified in the early time following the observation. Instead, at late times, the epidemic process

itself is progressively less predictable, even in the case of complete information. Moreover, we showed that the BP approach can be also exploited to evaluate some global quantities, such as the distribution of extinction times, for which it provided a pretty good characterization in most cases under study. We found that, in loopy graphs, BP reliability improves as the number of infected individuals discovered increases. This is expected, because, in this cases, the BP approximation becomes more reliable. Although Belief Propagation turned out to be a good approximation on loopy graphs, we faced convergence related issues, especially on dense networks. Still, the BP algorithm seemed to provide reasonably good results even in several cases in which a strict convergence of the equations was not reached.

The BP approach to the SIR model that was presented in this thesis is based on a very general principle, therefore it can be easily generalized to the study of several other inference problems with irreversible epidemic processes. We showed how the method can be extended to spreading on temporal networks and non-Markovian epidemic models. In addition, it can provide predictions on the disease evolution and the inference of the epidemic parameters at the same time. However, in realistic conditions, even if a disease is spreading within a population, surveillance programs may perform several observations without detecting any infected individual. As a consequence, whenever some cases are discovered, it is important to provide reliable estimations of the actual size of the outbreak. This discovery-incidence problem is another typical inference problem that can be tackled with the methods presented here [67, 68]. Another important issue is the generalization of the BP approach to the meta-population framework. Exploiting the knowledge of the statistical properties of the first outbreak time, i.e. the time at which an infected individual arriving into a susceptible population from an infected one first generates an outbreak [7, 69, 70], it is possible to work out an effective model connecting the original continuous-time microscopic outbreak, inside each population, to a discrete time macroscopic spreading process among different populations. The resulting effective model should be sufficiently simplified to allow for a direct analysis by means of BP technique.

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. *Histoire de l'Acad. Roy. Sci.(Paris) avec Mém. des Math. et Phys. and Mém*, pages 1–45, 1760.

[2] Daryl J Daley, Joe Gani, and Joseph Mark Gani. *Epidemic modelling: an introduction*, volume 15. Cambridge University Press, 2001.

[3] Maria D Van Kerkhove, Tommi Asikainen, Niels G Becker, Steven Bjorge, Jean-Claude Desenclos, Thais dos Santos, Christophe Fraser, Gabriel M Leung, Marc Lipsitch, Ira M Longini Jr, et al. Studies needed to address public health challenges of the 2009 h1n1 influenza pandemic: insights from modeling. *PLoS Med*, 7(6):e1000275, 2010.

[4] Jan Medlock and Alison P Galvani. Optimizing influenza vaccine distribution. *Science*, 325(5948):1705–1708, 2009.

[5] T Déirdre Hollingsworth, Neil M Ferguson, and Roy M Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature medicine*, 12(5):497–499, 2006.

[6] Ira M Longini, Azhar Nizam, Shufu Xu, Kumnuan Ungchusak, Wanna Hanshaoworakul, Derek AT Cummings, and M Elizabeth Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.

[7] Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med*, 4(1):e13, 2007.

[8] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC medicine*, 5(1):1, 2007.

[9] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC medicine*, 10(1):1, 2012.

[10] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 115, pages 700–721. The Royal Society, 1927.

[11] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.

[12] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE*, 6(8):e23176, 08 2011.

[13] Alain Barrat, C Cattuto, AE Tozzi, Philippe Vanhems, and Nicolas Voirin. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection*, 20(1):10–16, 2014.

[14] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9):e0136497, 2015.

[15] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[16] Petter Holme. Extinction times of epidemic outbreaks in networks. *PloS one*, 8(12), 2013.

[17] Petter Holme. Information content of contact-pattern representations and predictability of epidemic outbreaks. *arXiv preprint arXiv:1503.06583*, 2015.

[18] Petter Holme and Taro Takaguchi. Time evolution of predictability of epidemics on networks. *Physical Review E*, 91(4):042811, 2015.

[19] Daniel ben Avraham and Joachim Köhler. Mean-field (n, m)-cluster approximation for lattice models. *Physical Review A*, 45(12):8358, 1992.

[20] Kieran J Sharkey. Deterministic epidemiological models at the individual level. *Journal of Mathematical Biology*, 57(3):311–331, 2008.

[21] Brian Karrer and Mark EJ Newman. Message passing approach for general epidemic models. *Physical Review E*, 82(1):016101, 2010.

[22] Pedro C Pinto, Patrick Thiran, and Martin Vetterli. Locating the source of diffusion in large-scale networks. *Physical review letters*, 109(6):068702, 2012.

[23] Nino Antulov-Fantulin, Alen Lancic, Hrvoje Stefancic, Mile Sikic, and Tomislav Smuc. Statistical inference framework for source detection of contagion processes on arbitrary network structures. In *Self-Adaptive and Self-Organizing Systems Workshops (SASOW), 2014 IEEE Eighth International Conference on*, pages 78–83. IEEE, 2014.

[24] Nino Antulov-Fantulin, Alen Lančić, Tomislav Šmuc, Hrvoje Štefančić, and Mile Šikić. Identification of patient zero in static and temporal networks: Robustness and limitations. *Physical review letters*, 114(24):248701, 2015.

[25] Devavrat Shah and Tauhid Zaman. Detecting sources of computer viruses in networks: theory and experiment. *ACM SIGMETRICS Performance Evaluation Review*, 38(1):203–214, 2010.

[26] Wenxiang Dong, Wenyi Zhang, and Chee Wei Tan. Rooting out the rumor culprit from suspects. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 2671–2675. IEEE, 2013.

[27] Devavrat Shah and Tauhid Zaman. Rumors in a network: who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.

[28] Devavrat Shah and Tauhid Zaman. Rumor centrality: a universal source detector. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 199–210. ACM, 2012.

[29] Andrey Y Lokhov, Marc Mézard, Hiroki Ohta, and Lenka Zdeborová. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E*, 90(1):012801, 2014.

[30] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, and Riccardo Zecchina. Large deviations of cascade processes on graphs. *Physical Review E*, 87(6):062115, 2013.

[31] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, and Riccardo Zecchina. Optimizing spread dynamics on graphs by message passing. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(09):P09011, 2013.

[32] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, Alejandro Lage-Castellanos, and Riccardo Zecchina. Bayesian inference of epidemics on networks via belief propagation. *Phys. Rev. Lett.*, 112:118701, Mar 2014.

[33] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, Alessandro Ingrosso, and Riccardo Zecchina. The patient-zero problem with noisy observations. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(10):P10016, 2014.

[34] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.

[35] Luís M.A. Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, and Carlos Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364:513 – 536, 2006.

[36] William Goffman. Mathematical approach to the spread of scientific ideas–the history of mast cell research. *Nature*, 212(5061):449–452, 1966.

[37] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.

[38] Kieran J Sharkey. Deterministic epidemic models on contact networks: correlations and unbiological terms. *Theoretical population biology*, 79(4):115–129, 2011.

[39] Kieran J Sharkey, Istvan Z Kiss, Robert R Wilkinson, and Peter L Simon. Exact equations for sir epidemics on tree graphs. *Bulletin of mathematical biology*, 77(4):614–645, 2015.

[40] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34. IEEE, 2003.

[41] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, Joseph Rushton Wakeling, and Riccardo Zecchina. Containing epidemic outbreaks by message-passing techniques. *Physical Review X*, 4(2):021024, 2014.

[42] Andrey Y Lokhov, Marc Mézard, and Lenka Zdeborová. Dynamic message-passing equations for models with unidirectional dynamics. *Physical Review E*, 91(1):012811, 2015.

[43] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[44] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):1–30, 2015.

[45] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Information dynamics shape the sexual networks of internet-mediated prostitution. *Proceedings of the National Academy of Sciences*, 107(13):5706–5711, 2010.

[46] NG Van Kampen. Stochastic processes in chemistry and physics. *Amsterdam: North Holland*, 1:120–127, 1981.

[47] SP Blythe and RM Anderson. Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (hiv). *Mathematical Medicine and Biology*, 5(1):1–19, 1988.

[48] Martin Eichner and Klaus Dietz. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, 158(2):110–117, 2003.

[49] Hiroshi Nishiura and Martin Eichner. Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiology and infection*, 135(07):1145–1150, 2007.

[50] Norman TJ Bailey. On estimating the latent and infectious periods of measles: I. families with two susceptibles only. *Biometrika*, pages 15–22, 1956.

[51] Gerardo Chowell and Hiroshi Nishiura. Transmission dynamics and control of ebola virus disease (evd): a review. *BMC medicine*, 12(1):1, 2014.

[52] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[53] Judea Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible reasoning, 1988.

[54] Marc Mézard and Giorgio Parisi. The bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233, 2001.

[55] Marc Mézard and Giorgio Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111(1-2):1–34, 2003.

[56] Hans A Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 150(871):552–575, 1935.

[57] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.

[58] Ryoichi Kikuchi. A theory of cooperative phenomena. *Physical review*, 81(6):988, 1951.

[59] Yoshiyuki Kabashima and David Saad. Belief propagation vs. tap for decoding corrupted messages. *EPL (Europhysics Letters)*, 44(5):668, 1998.

[60] Yoshiyuki Kabashima and David Saad. The tap approach to intensive and extensive connectivity systems. *Advanced Mean Field Methods–Theory and Practice*, 6:65–84, 2001.

[61] Jonathan Yedidia. An idiosyncratic journey beyond mean field theory. *Advanced mean field methods: Theory and practice*, pages 21–36, 2001.

[62] Tom Britton and Federica Giardina. Introduction to statistical inference for infectious diseases. *arXiv preprint arXiv:1411.3138*, 2014.

[63] F. Altarelli, A. Braunstein, L. Dall'Asta, J. R. Wakeling, and R. Zecchina. Containing epidemic outbreaks by message-passing techniques. *Phys. Rev. X*, 4:021024, May 2014.

[64] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. *Bulletin of mathematical biology*, 68(8):1893–1921, 2006.

[65] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput Biol*, 7(3):e1001109, 2011.

[66] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

[67] S Parnell, TR Gottwald, WR Gilks, and F Van Den Bosch. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. *Journal of theoretical biology*, 305:30–36, 2012.

[68] S Parnell, TR Gottwald, NJ Cunniffe, V Alonso Chavez, and F van den Bosch. Early detection surveillance for an emerging plant pathogen: a rule of thumb to predict prevalence at first discovery. In *Proc. R. Soc. B*, volume 282, page 20151478. The Royal Society, 2015.

[69] Aurélien Gautreau, Alain Barrat, and Marc Barthelemy. Global disease spread: statistics and estimation of arrival times. *Journal of theoretical biology*, 251(3):509–522, 2008.

[70] Marc Barthélemy, Claude Godreche, and Jean-Marc Luck. Fluctuation effects in metapopulation models: percolation and pandemic threshold. *Journal of theoretical biology*, 267(4):554–564, 2010.

[71] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[72] Nicholas C Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.

[73] Edward A Bender and E Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978.

[74] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[75] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.

Appendix A

Networks

A network is any system that can be represented as a graph $G(V,E)$, whose elements are represented as N nodes in V and their relations or interactions are represented by the edges in E, i.e pairs of nodes. In the following we focus on undirected graphs whose edges are not ordered pairs, i.e $(i,j) = (j,i)$. The topological structure of a network is encoded in the adjacency matrix $\mathbf{A} = \{a_{ij}\}$, a $N \times N$ matrix defined as

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{if } (i,j) \notin E \end{cases} \tag{A.1}$$

Each node i that has k neighbors is said to have degree $k = \sum_{j \in \partial i} a_{ij}$. A graph is characterized by the degree distribution $P(k)$ defined as the probability that a randomly chosen vertex has degree k. Then the average degree is

$$\langle k \rangle = \sum_k kP(k). \tag{A.2}$$

A graph is called *connected* if any node i can be reached by any other node j. A *loop* is a closed path (i.e starting and ending in the same node) in which all nodes and edges are distinct. A *tree* is a graph in which any pair of nodes is connected by exactly one path, i.e it has no closed paths. *Barabási-Albert graphs* [71] are characterized by the following degree distribution

$$P(k) = 2m^2 k^{-3}, \tag{A.3}$$

where $m = \langle k \rangle$. A BA network can be described as the long time limit result of the evolution of an initial network described by to rules: at each time step a new vertex with m new edges is added and the new edges are connected to the old nodes with a probability proportional to their degree. They are an example of the wider class of power-law degree distributions $P(k) \simeq k^{-\gamma}$. In this class of network the nodes degree has large fluctuations: a large fraction of nodes have a small degree value and a small fraction of nodes (called *hubs*) has a very large degree, sometimes several orders of magnitude larger then the average value.

In a *random regular graph* every node has a fixed degree d [72]. They can be obtained, for example, by the *configuration model* [73, 74]: given the degree of each node (all of them have the same degree d in random regular graph), a corresponding number of stubs is attached to each node, then two randomly chosen stubs from two nodes are connected in order to form an edge. Once all the stubs are connected a graph with the desired degree sequence is obtained.

A *weighted network* [75] is characterized by the topological structure as well as by the intensity of the interactions that is encoded in a matrix ω_{ij} specifying the weight of the edge connecting i and j. The weight may assume any value representing the intensity of the interaction between the nodes.

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. *Histoire de l'Acad. Roy. Sci.(Paris) avec Mém. des Math. et Phys. and Mém*, pages 1–45, 1760.

[2] Daryl J Daley, Joe Gani, and Joseph Mark Gani. *Epidemic modelling: an introduction*, volume 15. Cambridge University Press, 2001.

[3] Maria D Van Kerkhove, Tommi Asikainen, Niels G Becker, Steven Bjorge, Jean-Claude Desenclos, Thais dos Santos, Christophe Fraser, Gabriel M Leung, Marc Lipsitch, Ira M Longini Jr, et al. Studies needed to address public health challenges of the 2009 h1n1 influenza pandemic: insights from modeling. *PLoS Med*, 7(6):e1000275, 2010.

[4] Jan Medlock and Alison P Galvani. Optimizing influenza vaccine distribution. *Science*, 325(5948):1705–1708, 2009.

[5] T Déirdre Hollingsworth, Neil M Ferguson, and Roy M Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature medicine*, 12(5):497–499, 2006.

[6] Ira M Longini, Azhar Nizam, Shufu Xu, Kumnuan Ungchusak, Wanna Hanshaoworakul, Derek AT Cummings, and M Elizabeth Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.

[7] Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med*, 4(1):e13, 2007.

[8] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC medicine*, 5(1):1, 2007.

[9] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC medicine*, 10(1):1, 2012.

[10] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 115, pages 700–721. The Royal Society, 1927.

[11] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.

[12] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE*, 6(8):e23176, 08 2011.

[13] Alain Barrat, C Cattuto, AE Tozzi, Philippe Vanhems, and Nicolas Voirin. Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection*, 20(1):10–16, 2014.

[14] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9):e0136497, 2015.

[15] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[16] Petter Holme. Extinction times of epidemic outbreaks in networks. *PloS one*, 8(12), 2013.

[17] Petter Holme. Information content of contact-pattern representations and predictability of epidemic outbreaks. *arXiv preprint arXiv:1503.06583*, 2015.

[18] Petter Holme and Taro Takaguchi. Time evolution of predictability of epidemics on networks. *Physical Review E*, 91(4):042811, 2015.

[19] Daniel ben Avraham and Joachim Köhler. Mean-field (n, m)-cluster approximation for lattice models. *Physical Review A*, 45(12):8358, 1992.

[20] Kieran J Sharkey. Deterministic epidemiological models at the individual level. *Journal of Mathematical Biology*, 57(3):311–331, 2008.

[21] Brian Karrer and Mark EJ Newman. Message passing approach for general epidemic models. *Physical Review E*, 82(1):016101, 2010.

[22] Pedro C Pinto, Patrick Thiran, and Martin Vetterli. Locating the source of diffusion in large-scale networks. *Physical review letters*, 109(6):068702, 2012.

[23] Nino Antulov-Fantulin, Alen Lancic, Hrvoje Stefancic, Mile Sikic, and Tomislav Smuc. Statistical inference framework for source detection of contagion processes on arbitrary network structures. In *Self-Adaptive and Self-Organizing Systems Workshops (SASOW), 2014 IEEE Eighth International Conference on*, pages 78–83. IEEE, 2014.

[24] Nino Antulov-Fantulin, Alen Lančić, Tomislav Šmuc, Hrvoje Štefančić, and Mile Šikić. Identification of patient zero in static and temporal networks: Robustness and limitations. *Physical review letters*, 114(24):248701, 2015.

[25] Devavrat Shah and Tauhid Zaman. Detecting sources of computer viruses in networks: theory and experiment. *ACM SIGMETRICS Performance Evaluation Review*, 38(1):203–214, 2010.

[26] Wenxiang Dong, Wenyi Zhang, and Chee Wei Tan. Rooting out the rumor culprit from suspects. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 2671–2675. IEEE, 2013.

[27] Devavrat Shah and Tauhid Zaman. Rumors in a network: who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.

[28] Devavrat Shah and Tauhid Zaman. Rumor centrality: a universal source detector. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 199–210. ACM, 2012.

[29] Andrey Y Lokhov, Marc Mézard, Hiroki Ohta, and Lenka Zdeborová. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E*, 90(1):012801, 2014.

[30] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, and Riccardo Zecchina. Large deviations of cascade processes on graphs. *Physical Review E*, 87(6):062115, 2013.

[31] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, and Riccardo Zecchina. Optimizing spread dynamics on graphs by message passing. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(09):P09011, 2013.

[32] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, Alejandro Lage-Castellanos, and Riccardo Zecchina. Bayesian inference of epidemics on networks via belief propagation. *Phys. Rev. Lett.*, 112:118701, Mar 2014.

[33] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, Alessandro Ingrosso, and Riccardo Zecchina. The patient-zero problem with noisy observations. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(10):P10016, 2014.

[34] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.

[35] Luís M.A. Bettencourt, Ariel Cintrón-Arias, David I. Kaiser, and Carlos Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364:513 – 536, 2006.

[36] William Goffman. Mathematical approach to the spread of scientific ideas–the history of mast cell research. *Nature*, 212(5061):449–452, 1966.

[37] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.

[38] Kieran J Sharkey. Deterministic epidemic models on contact networks: correlations and unbiological terms. *Theoretical population biology*, 79(4):115–129, 2011.

[39] Kieran J Sharkey, Istvan Z Kiss, Robert R Wilkinson, and Peter L Simon. Exact equations for sir epidemics on tree graphs. *Bulletin of mathematical biology*, 77(4):614–645, 2015.

[40] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34. IEEE, 2003.

[41] Fabrizio Altarelli, Alfredo Braunstein, Luca Dall'Asta, Joseph Rushton Wakeling, and Riccardo Zecchina. Containing epidemic outbreaks by message-passing techniques. *Physical Review X*, 4(2):021024, 2014.

[42] Andrey Y Lokhov, Marc Mézard, and Lenka Zdeborová. Dynamic message-passing equations for models with unidirectional dynamics. *Physical Review E*, 91(1):012811, 2015.

[43] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[44] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):1–30, 2015.

[45] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Information dynamics shape the sexual networks of internet-mediated prostitution. *Proceedings of the National Academy of Sciences*, 107(13):5706–5711, 2010.

[46] NG Van Kampen. Stochastic processes in chemistry and physics. *Amsterdam: North Holland*, 1:120–127, 1981.

[47] SP Blythe and RM Anderson. Distributed incubation and infectious periods in models of the transmission dynamics of the human immunodeficiency virus (hiv). *Mathematical Medicine and Biology*, 5(1):1–19, 1988.

[48] Martin Eichner and Klaus Dietz. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, 158(2):110–117, 2003.

[49] Hiroshi Nishiura and Martin Eichner. Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiology and infection*, 135(07):1145–1150, 2007.

[50] Norman TJ Bailey. On estimating the latent and infectious periods of measles: I. families with two susceptibles only. *Biometrika*, pages 15–22, 1956.

[51] Gerardo Chowell and Hiroshi Nishiura. Transmission dynamics and control of ebola virus disease (evd): a review. *BMC medicine*, 12(1):1, 2014.

[52] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[53] Judea Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible reasoning, 1988.

[54] Marc Mézard and Giorgio Parisi. The bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233, 2001.

[55] Marc Mézard and Giorgio Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111(1-2):1–34, 2003.

[56] Hans A Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 150(871):552–575, 1935.

[57] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.

[58] Ryoichi Kikuchi. A theory of cooperative phenomena. *Physical review*, 81(6):988, 1951.

[59] Yoshiyuki Kabashima and David Saad. Belief propagation vs. tap for decoding corrupted messages. *EPL (Europhysics Letters)*, 44(5):668, 1998.

[60] Yoshiyuki Kabashima and David Saad. The tap approach to intensive and extensive connectivity systems. *Advanced Mean Field Methods–Theory and Practice*, 6:65–84, 2001.

[61] Jonathan Yedidia. An idiosyncratic journey beyond mean field theory. *Advanced mean field methods: Theory and practice*, pages 21–36, 2001.

[62] Tom Britton and Federica Giardina. Introduction to statistical inference for infectious diseases. *arXiv preprint arXiv:1411.3138*, 2014.

[63] F. Altarelli, A. Braunstein, L. Dall'Asta, J. R. Wakeling, and R. Zecchina. Containing epidemic outbreaks by message-passing techniques. *Phys. Rev. X*, 4:021024, May 2014.

[64] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. *Bulletin of mathematical biology*, 68(8):1893–1921, 2006.

[65] Luis EC Rocha, Fredrik Liljeros, and Petter Holme. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput Biol*, 7(3):e1001109, 2011.

[66] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

[67] S Parnell, TR Gottwald, WR Gilks, and F Van Den Bosch. Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. *Journal of theoretical biology*, 305:30–36, 2012.

[68] S Parnell, TR Gottwald, NJ Cunniffe, V Alonso Chavez, and F van den Bosch. Early detection surveillance for an emerging plant pathogen: a rule of thumb

to predict prevalence at first discovery. In *Proc. R. Soc. B*, volume 282, page 20151478. The Royal Society, 2015.

[69] Aurélien Gautreau, Alain Barrat, and Marc Barthelemy. Global disease spread: statistics and estimation of arrival times. *Journal of theoretical biology*, 251(3):509–522, 2008.

[70] Marc Barthélemy, Claude Godreche, and Jean-Marc Luck. Fluctuation effects in metapopulation models: percolation and pandemic threshold. *Journal of theoretical biology*, 267(4):554–564, 2010.

[71] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[72] Nicholas C Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.

[73] Edward A Bender and E Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978.

[74] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, 6(2-3):161–180, 1995.

[75] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.