The patient-zero problem with noisy observations

Fabrizio Altarelli\textsuperscript{1,2}, Alfredo Braunstein\textsuperscript{1,2,3}, Luca Dall’Asta\textsuperscript{1,2}, Alessandro Ingrosso\textsuperscript{1} and Riccardo Zecchina\textsuperscript{1,2,3}

\textsuperscript{1} DISAT and Center for Computational Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
\textsuperscript{2} Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy
\textsuperscript{3} Human Genetics Foundation, Via Nizza 52, 10126 Torino, Italy
E-mail: alfredo.braunstein@polito.it

Received 5 August 2014
Accepted for publication 29 August 2014
Published 9 October 2014

Online at stacks.iop.org/JSTAT/2014/P10016
doi:10.1088/1742-5468/2014/10/P10016

Abstract. A belief propagation approach has been recently proposed for the patient-zero problem in SIR epidemics. The patient-zero problem consists of finding the initial source of an epidemic outbreak given observations at a later time. In this work, we study a more difficult but related inference problem, in which observations are noisy and there is confusion between observed states. In addition to studying the patient-zero problem, we also tackle the problem of completing and correcting the observations to possibly find undiscovered infected individuals and false test results.

Moreover, we devise a set of equations, based on the variational expression of the Bethe free energy, to find the patient-zero along with maximum-likelihood epidemic parameters. We show, by means of simulated epidemics, that this method is able to infer details on the past history of an epidemic outbreak based solely on the topology of the contact network and a single snapshot of partial and noisy observations.

Keywords: message-passing algorithms, network dynamics, epidemic modelling, statistical inference
1. Introduction

Epidemic compartment models provide a simple and useful mathematical description of the mechanisms behind disease transmission between individuals, in which only the most prominent aspects are included [1]. One of the most celebrated of these models, susceptible–infected–recovered (SIR) [2], describes diseases in which the person contracting the disease becomes immune to future infections after recovery, including diseases such as measles, rubella, chicken pox and generic influenza. The same model can be applied to lethal diseases such as HIV or ebola, provided that the recovered state is replaced by a removed state. Because they are simple and mathematically appealing, models such as SIR can be employed to study under which conditions and with which consequences large epidemic outbreaks can occur. For most examples of infective diseases, contagion runs over a network of effective contacts between individuals. These contact networks have been inaccessible for decades, but thanks to recent advances in technology miniaturization (e.g. by means of RFID-endowed badges to signal the proximity between individuals) and the popularization of the Internet (e.g. for the construction of databases of self-reported interactions), at least in simple and controlled scenarios, the interaction patterns of individual contacts can be almost entirely reconstructed [3, 4]. Modern computational epidemiology can thus rely on accurate data and powerful computers to run large-scale simulations of stochastic compartment models on real contact networks [5, 6].
In addition to epidemic forecast and control, a problem that has gained attention in recent years is the one of reconstructing the history of an epidemic outbreak [7–17, 18] (e.g. the path of contagion to a specific infected individual). In particular, identifying the origin (or the set of seeds or sources) of an epidemic outbreak in the general case is an open problem, even assuming simple discrete-time stochastic epidemic models such as the susceptible-infected (SI) and SIR models. The reason for this becomes clear when the inference is formulated as a maximum likelihood estimation problem. Estimating the maximum of a properly defined likelihood function corresponds to solve a (generally non-convex) optimization problem in the space of all possible epidemic propagations that are compatible with the data. For propagations with a unique source on regular trees, a maximum likelihood estimator was proposed by Shah and Zaman [7, 9] under the name of rumor centrality (see also [13]) and extended to probabilistic observations in [17]. For a specific continuous-time epidemic process, an optimal estimator on general trees was put forward by Pinto et al [11]. On general graphs, the number of propagation paths grows exponentially with the number of nodes, making exact inference unfeasible in practice. Instead of evaluating the likelihood function, Zhu and Ying put forward a method to select the path that most likely leads to the observed snapshot [16]. For general graphs, other heuristic inference methods are based on centrality measures [8,10], on the distance between observed data and typical outcomes of propagations for given initial conditions [12], or on the assumption that the epidemic propagation follows a breadth-first search tree [11, 15]. Even fewer results exist for epidemic inference with multiple sources [15]. A message-passing approach for the computation of epidemic dynamics was first introduced by Karrer and Newman [19] and applied to source detection in [14], with a further mean-field approximation of the likelihood function.

Recently, a belief propagation (BP) approach was proposed for the Bayesian inference of the origin of epidemics [18]. The main idea consists of exploiting a graphical model representation of the stochastic dynamics of the SI and SIR models to devise an efficient message-passing algorithm for the evaluation of the posterior distribution of the epidemic sources. The BP approach is exact on trees and also works very well on general graphs, outperforming other methods on many graph topologies, in the presence of one or more sources and also when observations are by large extent incomplete. In this work, we build on the BP approach by studying a more difficult variant of the inference problem in which either (a) it is not possible to distinguish between recovered or susceptible individuals or (b) the observation is noisy, i.e. there is a non-zero probability of making an error in the observation of each individual. The first case has already been described in [16] using the most likely infection path method on trees and similar heuristics on general graphs. The second case was not directly addressed in the literature, although the problem of determining the causative network of epidemiological data in the presence of false negatives and positives has recently attracted some attention [20,21]. We show that the BP approach allows for more complete inference, for example, (a) to infer the epidemic parameters, i.e. the probability of transmission in each contact and the distribution of recovery times; and (b) to infer missing data in a partial observation, e.g. correcting errors or finding which of the two states, S or R, are in the confused state setup. The work is organized as follows. Section 2 provides a detailed description of the graphical model representation of the stochastic epidemic dynamics. The BP equations of the model and the details of their efficient implementation are discussed in section 3 (and in the

doi:10.1088/1742-5468/2014/10/P10016
appendices). The results of the Bayesian inference under different observation models are reported in section 4. In section 5, we present an efficient on-line method for the inference of the epidemic parameters by maximization of the log-likelihood by gradient ascent in the Bethe approximation.

2. Graphical model representation of the epidemic process

We consider a discrete-time version of the (SIR) model [22] on a graph $G = (V, E)$ that represents the contact network of a set $V$ of individuals. A node $i$ can be in one of three possible states: susceptible ($S$), infected ($I$), or recovered/removed ($R$). The state of node $i$ at time $t$ is represented by variable $x^t_i \in \{S, I, R\}$. At each time step (e.g. a day) of the stochastic dynamics, an infected node $i$ can first spread the disease to each susceptible neighbor $j$ with given probability $\lambda_{ij}$, then recover with probability $\mu_i$. Once recovered, individuals can no longer become sick. This process is Markovian and satisfies

$$P(x^{t+1}_i | x^t) = \prod_i P(x^{t+1}_i)$$

where

$$P(x^{t+1}_i = S | x^t) = \mathbb{I}[x^t_i = S] \prod_j (1 - \lambda_{ji} \mathbb{I}[x^t_j = I])$$

$$P(x^{t+1}_i = I | x^t) = (1 - \mu_i) \mathbb{I}[x^t_i = I] + \mathbb{I}[x^t_i = S] \prod_j (1 - \lambda_{ji} \mathbb{I}[x^t_j = I])$$

$$P(x^{t+1}_i = R | x^t) = \mu_i \mathbb{I}[x^t_i = I] + \mathbb{I}[x^t_i = R].$$

A realization of the stochastic dynamics is fully specified by knowing, for each individual $i$, her infection time $t_i = \min\{t : x^t_i = I\}$ and her recovery time $g_i = \min\{g : x^{t_i+g+1}_i = R\}$. It is easy to show that, for a given initial configuration $\{x^0_i\}$, a realization of the stochastic process can be generated by drawing the recovery time $g_i$ of each node $i$ and an infection transmission delay $s_{ij}$ from node $i$ random node $j$, for all pairs $(ij)$. The recovery times $\{g_i\}$ are independent random variables extracted from geometric distributions

$$G_i(g_i) = \mu_i (1 - \mu_i)^{g_i},$$

while the delays $\{s_{ij}\}$ are conditionally independent random variables distributed according to a truncated geometric distribution,

$$\omega_{ij}(s_{ij} | g_i) = \begin{cases} \lambda_{ij} (1 - \lambda_{ij})^{s_{ij}}, & s_{ij} \leq g_i \\ \sum_{s > g_i} \lambda_{ij} (1 - \lambda_{ij})^s, & s_{ij} = \infty, \end{cases}$$

in which, for convenience, we concentrate on the value $s_{ij} = \infty$ of the mass of the distribution beyond the hard cut-off $g_i$ imposed by the recovery time. Infection times are related by the deterministic equation

$$t_i = 1 + \min_{j \in \partial i} \{t_j + s_{ji}\},$$

which is a constraint encoding the infection dynamics of the SIR model. Then, individual $i$ recovers at time $t_i + g_i$.

The exact mapping from realizations of the epidemic process to realizations of transmission delays and recovery times can be exploited to provide a graphical model
representation of the stochastic dynamics of the SIR model on a graph. For a given initial
condition, the joint probability distribution of infection and recovery times conditioned
on the initial state is

\[ P(t, g \mid x^0) = \sum_s P(s \mid g) P(t, g, s) P(g) = \sum_s \prod_{i,j} \omega_{ij} (s_{ij} \mid g_i) \prod_i \phi_i(t_i, \{t_k, s_{ki}\}_{k \in \partial i}) \mathcal{G}_i(g_i), \]

where

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

is a characteristic function, which imposes on each node \( i \) the dynamical constraint (2).

In the following, we derive a method to reconstruct information about the origin of an
epidemic given some observation at a later time. We first need to compute the posterior
probability of the initial configuration given an observation at time \( T \). This is done by
assuming a probabilistic prior on the initially infected nodes and applying Bayes formula

\[ P(x_0 \mid x_T) \propto \sum_{t, g} P(t, g \mid x_T) P(t, g, x^0) P(x^0) \]

where \( P(x^0) = \prod_i \gamma_i(x_i^0) \) is a factorized prior on the initial infection with

\[ \gamma_i(x_i^0) = \gamma \delta(x_i^0, I) + (1 - \gamma) \delta(x_i^0, S) \]

for a generally small constant \( \gamma \) and where we exploited the fact that the state \( x^t \) is a
deterministic function of the set of infection and recovery times \( (t, g) \), which gives us

\[ P(x_T \mid t, g) = \prod_i \zeta_i^T (t_i, g_i, x_i^T) \]

\[ P(t, g \mid x^0) \propto \prod_i \zeta_i^0 (t_i, g_i, x_i^0) \]

with

\[ \zeta_i^t = \mathbb{I}[x_i^t = S, t < t_i] + \mathbb{I}[x_i^t = I, t_i \leq t < t_i + g_i] + \mathbb{I}[x_i^t = R, t_i + g_i \leq t] . \]

The above formula can be generalized to the case in which the parameters \( \mu \) and \( \lambda \)
have an explicit dependence on \( t_i \). The problem of computing the marginals from (6) is,
in general, intractable (NP-hard) and we need to resort to an efficient approximation.
Here, we choose to implement the BP approximation, which preserves some non-trivial
correlations between variables and is exact on acyclic graphs.

A general theory for the correctness of the BP equations is unfortunately lacking. Correctness
has been established for arbitrary topologies in a few particular cases in the zero temperature limit, e.g. \[23–25\] and in the case of Gaussian potentials \[26\]. Regarding the problem at non-zero temperature, approximation bounds and correctness
of the marginals and the free energy, in the infinite volume limit, can be ensured for locally

\[ \text{doi:10.1088/1742-5468/2014/10/P10016} \]

5
tree-like factor graphs with potentials leading to map contractiveness of the BP equations or more general correlation decay conditions such as De Dobrushin’s (see e.g. [27]). In addition to the exact result given here for acyclic graphs, the BP approach for the inference of the patient-zero in the noiseless case has been validated experimentally by means of simulations on various types of random graphs and on real contact networks [18].

We proceed by introducing a factor graph representation of (6), namely a bipartite graph composed of factor nodes and variable nodes. In the standard definition, each variable appearing in the problem is identified by a variable node, while each factorized term of the probability weight in (6) is represented by a factor node. A factor node is connected to the set of variable nodes appearing in the corresponding factorized term. However, with this definition, the factor graph of (6) has a loopy structure both at local and global scales and this could compromise the accuracy of the BP approximation. The existence of short loops can be easily verified even focusing only on the expression of the dynamical constraint (2): a pair of variable nodes corresponding to the infection times $t_i$ and $t_j$ of neighboring individuals are indeed involved in the two factors $\phi_i$ and $\phi_j$, inducing a short loop in the factor graph (see figure 1(a)). We would use a factor graph representation that maintains the same topological properties of the original graph of contacts to guarantee that BP is exact when the original graph of contacts is a tree. Following an approach proposed in [28,29], the factor graph can be disentangled by grouping pairs of infection times $(t_i, t_j)$ in the same variable node as in figure 1(b). For convenience, we keep all variable nodes $\{t_i\}$, but also introduce, for each edge $(i, j)$ emerging from a node $i$, a set of copies $t_i^{(j)}$ of the infection time $t_i$ that is forced to take the common value $t_i$ by including the constraint $\prod_{k \in \partial i} \delta(t_i^{(k)}, t_i)$ in the factor $\phi_i$. 

\[ \delta(t_i^{(k)}, t_i) \]

Figure 1. (a) Example of a loopy factor graph representation induced by constraints such as those in (2). (b) Disentangled factor graph. (c) A more convenient representation of the disentangled factor graph employed in the present work. For simplicity, the dependency on $\{s_{ij}\}$ is not considered.
We also observe that the factors $\phi_i$ depend on infection times and transmission delays only through the sums $t_i^{(j)} + s_{ij}$. It is thus more convenient to introduce the variables $t_{ij} = t_i^{(j)} + s_{ij}$ and express the dependencies through the pairs $(t_i^{(j)}, t_{ij})$.

Finally, it is convenient to group the variable $g_i$ with the corresponding infection times $t_i$ in the same variable node, replace $g_i$ and $g_j$ by their copies $g_i^{(i)}$ and $g_j^{(j)}$ in the edge constraints $\omega_{ij}(t_{ij} - t_i^{(j)}|g_i^{(i)})$ and $\omega_{ji}(t_{ji} - t_j^{(j)}|g_j^{(j)})$ and impose the identity $\prod_{k \in \partial i} \delta(g_k, g_i)$ for each node $i$.

We can now define the new factors

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

and

$$\psi_i = \delta(t_i, \mathbb{I}[x_i^0 \neq I](1 + \min_{j \in \partial i}(t_{ji}))) \prod_{j \in \partial i} \delta(t_i^{(j)}, t_i)\delta(g_i^{(j)}, g_i)$$

$$= \phi_i(t_i, \{t_{ji}\}_{j \in \partial i}) \prod_{j \in \partial i} \delta(t_i^{(j)}, t_i)\delta(g_i^{(j)}, g_i).$$

The posterior distribution can be written as

$$\mathcal{P}(x^0|x^T) \propto \sum_{t, \{t_{ij}\}, \mathbf{g}} \mathcal{Q}^{T}(\mathbf{g}, \mathbf{t}, \{t_{ij}\}, x_0)$$

where

$$\mathcal{Q}^{T}(\mathbf{g}, \mathbf{t}, \{t_{ij}\}, x_0) = \frac{1}{Z} \prod_{i<j} \phi_{ij} \prod_i \psi_i \mathcal{G}_i \gamma_i \zeta_i^T \zeta_i^0.$$
From (13), it is possible to compute single posterior marginals \( P(x_0^i|x^T) = \sum_{\{x_0^j:j\neq i\}} P(x_0^j|x^T) \). For the patient-zero problem, we simply rank nodes in decreasing order of \( P(x_0^i = I|x^T) \).

3. BP equations

Belief propagation consists of a set of equations for single-site probability distributions labeled by directed graph edges. These equations are solved by iteration and, on a fixed point, give an approximation for single-site marginals and other quantities of interest such as the partition function \( Z \) (see e.g. [30,31] for a general introduction).

We recall the general form of the BP equations in the following. For a factorized probability measure on \( z = \{z_i\} \),

\[
M(z) = \frac{1}{Z} \prod_a F_a(z_a)
\]

(15)

where \( z_a \) is the subvector of variables that \( F_a \) depends on, the general form of the equations is

\[
p_{F_a\rightarrow i}(z_i) = \frac{1}{Z_{z_{\text{partial}}}} \sum_{\{z_j\}_{j\in \partial a \setminus i}} F_a(\{z_i\}_{i\in \partial a}) \prod_{j\in \partial a \setminus i} m_{j\rightarrow F_a}(z_j)
\]

(16)

\[
m_{i\rightarrow F_a}(z_i) = \frac{1}{Z_{z_{\text{partial}}}} \prod_{b\in \partial a \setminus i} p_{F_b\rightarrow i}(z_i)
\]

(17)

doi:10.1088/1742-5468/2014/10/P10016 8
The patient-zero problem with noisy observations

$$m_i(z_i) = \frac{1}{Z_i} \prod_{b \in \partial i} p_{F_a \to i}(z_i)$$  \hspace{1cm} (18)

where $F_a$ is a factor (i.e. $\psi_i$, $\phi_{ij}$, $\gamma_i$, $\zeta^0_i$, $\zeta^T_i$, or $G_i$ in our case), $z_i$ is a variable (i.e. $(t_i, g_i)$, $(t^{(j)}_i, g^{(j)}_i, t_{ji})$, $x^0_i$, or $x^T_i$ in our case), $\partial i$ is the subset of indices of variables in factor $F_a$ and $\partial i$ is the subset of factors that depend on $z_i$. The terms $Z_{ia}, Z_{ai}$ and $Z_{i}$ are normalization factors that can be calculated once the rest of the right-hand side is computed. Equation 18 for variable $x^0_i$ directly gives the posterior estimation of the probability of node $i$ being in state $I$ at time 0, i.e. being the patient-zero and equation 18 for variable $x^T_i$ gives the posterior estimation on the real state of individual $i$ at time $T$.

While equations (17)–(18) can be always computed efficiently in general, the computation of the trace in (16) may need a time which is exponential in the number of participating variables.

In appendix A, we show the derivation of an efficient version of the computation of the message $p_{\psi_{i} \to j}(t^{(j)}_i, t_{ji}, g^{(j)}_i)$ in equation (16) for node factor $\psi_i$ in (12) that can be computed in linear time in the degree of vertex $i$.

Factor $\phi_{ij}$ in equation (11) involves two (aggregated) variables: $(t^{(j)}_i, g^{(j)}_i, t_{ji})$ and $(t^{(i)}_j, g^{(i)}_j, t_{ij})$ (see figure 2). Variables $t^{(j)}_i, t^{(i)}_j, t_{ij}, t_{ji}$ have $T + 2$ states, while we limit variables $g^{(i)}_j, g^{(j)}_i$ to take values in $0, \ldots, G$. Given a distribution of recovery delays, it is sufficient to take $G$ such that the weight of the tail of the distribution $\sum_{g=G}^\infty P(g)$ is small enough when compared to $1/N$. In the case of the geometric distribution of recovery delays, one can take, for example, $\sum_{g=G}^\infty P(g) = \sum_{g=G}^\infty p(1-p)^g = (1-p)^G \sim 1/N$, i.e. $G$ needs to grow only logarithmically with $N$ and can, of course, be truncated at $T$. A naive implementation of the BP equations for factor $\phi_{ij}$ thus takes $O(T^4G^2)$ operations, which can be still too expensive in practical applications.

It is possible to use a simpler representation for the messages in which we only retain information on the relative timing between infection time $t^{(j)}_i$ for a node $i$ and the infection propagation time $t_{ji}$ on its link with node $j$, introducing the variables

$$\sigma_{ji} = 1 + \text{sign} \left( t_{ji} - (t^{(j)}_i - 1) \right), \hspace{1cm} (19)$$

effectively reducing the complexity of messages from $O(T^2G)$ to $O(TG)$ real numbers. In appendix A, we show the computation of $p_{\phi_{ij} \to j}(t_j, \sigma_{ij}, g_j)$ from equation (16) for factor $\phi_{ij}$ with this simplification for the messages, bringing down the computation time from $O(T^4G^2)$ to $O(TG^2)$ operations.

We also show in appendix A how to efficiently compute the BP equation for factor $\psi_i$, corresponding to the simplified version of the messages $p_{\psi_{i} \to j}(t^{(j)}_i, \sigma_{ji}, g^{(j)}_i)$. The computation of all updates of each factor $\psi_i$ takes $O(TG|\partial i|)$ operations, amounting to $O(TG|\partial i|)$ operations for all $\psi$ nodes per iteration.

The update equations for the remaining factors, $\gamma_i$ in equation (7), $\zeta^0_i, \zeta^T_i$ in equation (10) and $G_i$ in equation (1), can be computed in a straightforward way from (16), as they each involve a very small number of variables states.

The overall required computation time of the update is dominated by the updates of $\phi_{ij}$ nodes and amounts to $O(TG^2|E|)$ operations per iteration, where $|E|$ is the number of edges of the original contact graph. To give a rough estimation, in our C++ implementation, in a graph with 1000 nodes and 4000 contacts (edges), with
The patient-zero problem with noisy observations

\[ T = G = 20, \text{ the computation takes about one minute per iteration on a single cpu.} \]
\[ \text{The full computation requires a few hundred iterations, i.e. about three hours on a single cpu. Note that computation can be easily parallelized even on a single instance (the computation for different nodes can be carried out simultaneously); in our parallel implementation, the computation time gets divided almost entirely by the number of employed CPUs. Values of } G \text{ smaller than } T \text{ (within the estimation given above) can be used to reduce the computation time for larger } T. \]
\[ \text{Regarding memory usage, the BP equations with simplified messages need to store } O(TG|E|) \text{ real numbers.} \]

4. Observation models

In the inference of the origin of epidemic propagations, it is often assumed that the state of every node is known at the observation time \( T \) with no uncertainty. This is also the case studied in [18], where the BP approach for this problem was first introduced. In practice, every clinical test for determining the state of an individual is affected by some amount of error and this possibility must be taken into account in the inference problem. Therefore, it is realistic to assume that each observation carries some level of noise. We introduce a general concept of observation model for the inference problem, which allows dealing with several different cases of incomplete and noisy data using common notation. We assume that the noise level is known (as is the case for the majority of clinical tests) and introduce a new variable \( y^T_i \in \{S, I, R\} \) for the observed state of node \( i \) and an additional evidence term that reflects the probability \( n_i(y^T_i|x^T_i) \) of the observed state \( y^T_i \) given the true state \( x^T_i \). In the factor graph, the observed-state variables \( y^T_i \) are fixed to their values given by the experimental observation (by means of a delta function representing an infinite external field), while the true-state variables \( x^T_i \) are traced over in the compatibility function \( \zeta^T_i \). More explicitly, the modified factor graph shown in figure 3(b), contains a \( \zeta^T_i \) factor node attached to the true-state variable \( x^T_i \), which is linked to the observed state \( y^T_i \) (which is a constant) through the node \( n_i(y^T_i|x^T_i) \). The posterior distribution now takes the form

\[
P(x^0|y^T) \propto \sum_{x^T, t, t_{ij}, g} Q'(x^T, g, t, t_{ij}, x_0) \tag{20}
\]

where

\[
Q'(x^T, g, t, t_{ij}, x_0) = \frac{1}{Z} \prod_{i<j} \phi_{ij} \prod_i \psi_i G_i \zeta_i n_i. \tag{21}
\]

In what follows, we introduce, for convenience, a map \( \rho(s) \) from indices \( i \in \{1, 2, 3\} \) into configurations of the \( x \) variables, such that \( \rho(1) = S, \rho(2) = I, \rho(3) = R \) and then define the observational transition matrix (OTM) \( O^{(i)}_{s,t} \), whose elements are the transition probabilities

\[
O^{(i)}_{s,t} = n_i(\rho(s), \rho(t)). \tag{22}
\]

The case in which observations are complete and noiseless corresponds to an identity matrix \( O^{(i)}_{s,t} = \delta_{st} \). In the following sections, we provide some interesting examples.
The patient-zero problem with noisy observations

of applications of this scheme to confused and noisy observations. Note that, in this
generalized scheme, we can also take into account the case of partial observations by
assuming a totally uniform OTM $O^{(i)}_{s,i} \equiv \frac{1}{3}$ for unobserved nodes.

4.1. Inference of epidemic source from confused observations

In some situations, it can be hard to distinguish between nodes that already recovered
from a disease and nodes that did not contract it. To take this fact into account, we follow
the approach of [16] and explore the efficiency of our inference machinery in a setting in
which observations on susceptible and recovered nodes are confused. More specifically, we
allow only two types of observed states $x^T_i \in \{I, N\}$, where $N$ stands for not-infected.
This situation corresponds to choose the following OTM:

$$O^{(i)} = \begin{pmatrix}
\frac{1}{2} & 0 & \frac{1}{2} \\
0 & 1 & 0 \\
\frac{1}{2} & 0 & \frac{1}{2}
\end{pmatrix}.$$

We verified the performance of the BP algorithm on a completely uniform setting
provided by random regular graphs with identical infection parameters $(\lambda, \mu)$ for all nodes
and links. All epidemic propagations were initiated from a unique seed (the patient-zero).
For each node, the BP algorithm provides an estimate of the posterior probability that
the node became infected at a certain time and thus, also the probability that the node
was the origin of the epidemic. We can thus rank the nodes in decreasing order with
respect to the estimated probability of their being the origin of the observed epidemics:
the position of the true origin in the ranking provided by the algorithm is a good measure
of the efficacy of the method. In what follows, we indicate with $i_0$ the ranking of the true
origin of the epidemic and with $|G|$ the number of nodes in the graph $G$.

An important by-product of the algorithm is the ability to infer the true state of a
node from the marginal of the infection time, providing a method for ‘correct’ observations.
More precisely, in the present example, we consider the problem of discriminating between
susceptible and recovered nodes. An effective method for quantifying the accuracy of such
binary classification problem is the receiver operating characteristic (ROC) curve, namely
a plot of the ‘true positive rate’ against the ‘false positive rate’. Constructing the ROC
curve in the present case is very easy: we select the $N$ nodes and rank them on the base
of their marginal $P(t_i = \infty | x^0)$. We then take one step upward in the ROC whenever a
true positive case is encountered ($y^T_i = x^T_i = S$) or one step rightward in the case of a
false positive ($y^T_i \neq x^T_i$). We performed this discrimination analysis for each sample and
then computed the average value of the area under the ROC curve, which gave indication
of the fraction of correctly classified nodes. It turns out that the proposed algorithm can
be effectively used as an ex-post-facto tool for discriminating susceptible from recovered
individuals.

Figure 4 displays the absolute rank of the true infected site $i_0$, for a set of $M = 1000$
simulated epidemic propagations with $\lambda = 0.6$ and $\mu = 1$ on random regular graphs of
size $N = 1000$ and degree $d = 4$ ($T = 10$). The probability of perfect inference of the
patient-zero is also reported. The quantities of interest are plotted as functions of the
normalized epidemic size $N_{IR} = \frac{|I| + |R|}{|G|}$ (i.e. the fraction of infected or recovered sites),

\[ \text{doi:10.1088/1742-5468/2014/10/P10016} \]
The patient-zero problem with noisy observations

Figure 4. Probability of perfect inference of the patient-zero (blue solid line), average absolute rank of the true patient-zero (violet solid line), average ROC area (red solid line) and average fraction of infected nodes (green dotted line) as a function of the rescaled epidemic size \( \frac{|I| + |R|}{|G|} \). The fraction of the \( M \) samples belonging to each bin of the rescaled epidemic size is also indicated. The realization of the epidemic process is propagated for \( T = 10 \) steps with \( \lambda = 0.6 \) and \( \mu = 1 \). Observations are confused, i.e. \( x^t_i \in \{I, N\} \). Simulations were run over \( M = 1000 \) samples of random regular graphs with \( N = 1000 \) nodes and degree \( d = 4 \).

whose values are discretized with intervals of width equal to 0.05. Note that in all the figures we show in the paper, we discarded the rare cases with very low epidemic size (\( N_{IR} < 0.3 \) in figure 5, \( N_{IR} < 0.2 \) elsewhere) where the number of infected is extremely low and the inference is practically unfeasible.

For each set of data, the symbols report the mean value obtained by averaging over the samples belonging to that interval and the error bars indicate the corresponding standard deviation. The average fraction of infected nodes and the fraction of samples in each bin are reported as a reference. The algorithm is very effective in identifying the patient-zero average for all values of the normalized epidemic size. We also show the average ROC area, which reveals that the inference algorithm allows very good discrimination between \( S \) and \( R \) nodes.

The same analysis for a random graph with power-law degree distribution, obtained using the Barabasi–Albert model [32], is reported in figure 5. When the observation time \( T \) is sufficiently small (\( T = 7 \) in figure 5), the performance of the algorithm is high. When longer observation times are considered, epidemics tend to cover the whole network and convergence issues emerge. In this regime, most of the infected nodes have already recovered at observation time \( T \) (and thus they can no longer be distinguished from the susceptible ones). This causes a rapid decay of the available information content that explains the performance degradation. A similar effect arises also on random regular graphs, but at longer times, as we show in section 4.2.

In summary, even when supplied with confused observations, BP shows striking ability to discriminate between recovered and susceptible nodes, provided that there is enough information at the chosen observation time \( T \).

\[ \text{doi:10.1088/1742-5468/2014/10/P10016} \]
The patient-zero problem with noisy observations

4.2. Inference of the epidemic source with noisy observations

Let us now consider a simple type of observational noise. Suppose that a node on state $x$ has probability $1 - \nu$ of being correctly observed in state $x$ and probability $\nu$ of being observed incorrectly in one of the two remaining states, distributed uniformly among the two. For example, node $i$ could be $I$ (infected) at the observation time $T$ and, for a given noise level $\nu$, there will be an equal probability $\frac{\nu}{2}$ for node $i$ to be observed in the $R$ (recovered) or $S$ (susceptible) state. This setting corresponds to the following $OTM$:

$$O^{(i)} = \begin{pmatrix}
1 - \nu & \frac{\nu}{2} & \frac{\nu}{2} \\
\frac{\nu}{2} & 1 - \nu & \frac{\nu}{2} \\
\frac{\nu}{2} & \frac{\nu}{2} & 1 - \nu
\end{pmatrix}.$$

We simulated a set of $M = 1000$ single-source epidemic propagations with $\lambda = 0.6$ and $\mu = 1$ on random regular graphs with $N = 1000$ nodes and degree $d = 4$. In figure 6, we show the average rank of the true origin of the epidemics (left) and the probability of inferring the true patient-zero (right) for various levels of observational noise up to $\nu = 0.4$. The low values of the average rank obtained demonstrate that the BP algorithm is able to perform extremely well up to very high levels of noise. The corresponding ROC curves are plotted in figure 7.

We investigated the role of observation time $T$ in relation to the amount of information needed for inferring the patient-zero: simulations were run for given realizations of the epidemic process and observation time was systematically varied. In figure 8, we show
Figure 6. Average normalized rank of the true patient-zero (left) and probability of ranking it in the first place (right) as a function of epidemic size $|I| + |R|/|G|$ for various levels of noise $\nu$ in the observation (the error-bars indicate the standard deviation computed on the sub-sample corresponding to a given epidemic size). Each curve refers to $M = 1000$ samples of random regular graphs with $N = 1000$ nodes and degree $d = 4$. The epidemic is propagated until $T = 10$ with $\lambda = 0.6$ and $\mu = 1$.

Figure 7. Average area of the ROC curve of the inference of the state of noisy variables as a function of epidemic size $|I| + |R|/|G|$ for various levels of noise $\nu$ in the observation (the error-bars indicate the standard deviation computed on the subsample corresponding to a given epidemic size). Each curve refers to $M = 1000$ samples of random regular graphs with $N = 1000$ nodes and degree $d = 4$. The epidemic is propagated until $T = 10$ with $\lambda = 0.6$ and $\mu = 1$.

a representative situation in random graphs (the picture is similar in scale-free graphs, as we argued in section 4.1). It turns out that the ratio of infected nodes to epidemic size is critical for inference: when observation time is too long so that the majority of infected individuals have recovered, it is much more difficult to find the patient-zero in
The patient-zero problem with noisy observations

Figure 8. Normalized rank of the true patient-zero (solid lines), fraction of infected nodes $|I|/|G|$ (green dotted line), and rescaled epidemic size $(|I| + |R|)/|G|$ (dotted purple line) as a function of observation time $T$ for a single realization of the epidemic process, propagated with $\lambda = 0.6$ and $\mu = 1$, on a random regular graph with $N = 1000$ nodes and degree $d = 4$. Observations are complete (black solid line, superimposed to x axis), confused (blue solid line) and with noise 20% noise level (red solid line). Values of the normalized rank greater than 0.5 are meaningless: they are the realization of a random variable with an average close to 0.5 and are evidence that for large $T$, the inference algorithm is unable to identify the patient-zero with better precision than pure chance.

5. Inference of epidemic parameters

We have shown that if the parameters of the epidemic are known in advance, our inference method can effectively detect the patient-zero. It is reasonable to assume that, for certain types of diseases, clinical information could dictate some plausible range for the average rates of infection and recovery. In the simplified case in which the infection parameters are uniform among the population, the BP method can be generalized in a way to infer both the patient-zero and the epidemic parameters at the same time.

From our Bayesian approach, $-f(\lambda, \mu) = \log Z(\lambda, \mu)$ is the log-likelihood of the epidemic parameters for the observation $x_T$. Indeed, the log–likelihood of the parameters $-f(\lambda, \mu)$ equals $\log \mathcal{P}(x_T|\lambda, \mu)$ and the latter can be computed as

$$\mathcal{P}(x_T|\lambda, \mu) = \sum_{t,g,x^0} \mathcal{P}(x_T|t,g) \mathcal{P}(t,g|x^0) \mathcal{P}(x^0) = Z(\lambda, \mu).$$

In [18], this observation was used to infer the epidemic parameters through an exhaustive search in the space of parameters. This computation can be costly, wasting resources on uninteresting regions of the parameter space. Moreover, this type of experiment shows
that the log-likelihood landscape with the Bethe approximation is generally very simple, presenting, in most cases, a single local maximum. Here, we describe a different method to infer the parameters, together with the source of the epidemic outbreak. The idea is to perform an on-line log-likelihood maximization through gradient ascent in the Bethe approximation of the log-likelihood, by means of the following updates:

\[
\lambda \leftarrow \lambda + \epsilon \frac{\partial f}{\partial \lambda} \tag{24}
\]

\[
\mu \leftarrow \mu + \epsilon \frac{\partial f}{\partial \mu} \tag{25}
\]

with \(\epsilon\) as a free convergence parameter. The free energy of the system can be approximated with the Bethe free energy, which, in turn, can be expressed as a sum of local terms depending on BP messages. A detailed derivation of the expression of the derivatives \(\frac{\partial f}{\partial \lambda}\) and \(\frac{\partial f}{\partial \mu}\) of the Bethe free energy is reported in appendix B. In principle, the expressions obtained using the Bethe free energy are valid only at the BP fixed point and one should let BP updates converge before making a step of gradient ascent. In practice, we found that it is sufficient to interleave BP and gradient ascent updates in order to obtain equivalent results. A fixed point of the interleaved updates is both a critical point of the Bethe log-likelihood and a BP approximation for the marginals. We performed extensive simulations with a wide range of parameters and found that, for a reasonable fraction of infected nodes at the observation time, the inference can simultaneously identify the patient-zero perfectly and find good estimates of the epidemic parameters. Some examples of inferred parameters are shown in figure 9 for six different configurations of \((\lambda, \mu)\) parameters, with each pair of box plots referring to \(M = 1000\) samples.

The method can be extended to treat the non-uniform case, at the expense of a higher computational effort that would amount to computing local derivatives of the
The patient-zero problem with noisy observations

Figure 10. Inferred epidemic parameters for different observational noise rates $\nu$. Forward epidemic is simulated until observation time $T = 10$. Each box refers to $M = 1000$ instances of random regular graphs with $N = 1000$ nodes and degree $g = 4$. Box edges signal the 25th and 75th percentiles and the central red lines is the median. Whiskers extend to cover 99.3% of the data for a Gaussian distribution. Outliers are marked as red points outside the whiskers.

free-energy function for each edge in the graph with respect to edge-specific parameters. It should be clear that the number of parameters in the non-uniform case should not grow excessively for inference purposes. We could, nevertheless, account for age or gender-dependent differences in the probability of contracting the disease or in the recovery rate with the introduction of additional information attached to nodes and edges in the network. Notice that, even in this case, an exhaustive search in the parameter space could be computationally too expensive.

The inference of parameters can also be performed in the presence of observational noise. In figure 10, we show an example of inference for increasing levels of noise in the observation, as defined in the preceding section. Also in this case, the patient-zero is detected with probability 1 and the inferred parameters are good estimators of the true values even up to a significant fraction of noise.

We compared the ability of BP to infer the epidemic parameters with a simpler (though computationally expensive) procedure that does not require (nor provide) the simultaneous inference of the origin. The idea is to compare the statistical properties of the observation with the one of typical epidemics with given parameters $\lambda, \mu$ and choose those $\lambda, \mu$ that give properties that are closest (in a sense to be defined) to the ones observed. More precisely, for each $\lambda \in \{0.05, 0.1, \ldots, 0.95\}$ and $\mu \in \{0, 0.05, \ldots, 1\}$, we generate 1000 random epidemics and compute the mean of the number of infected $I_{\text{mean}}(\lambda, \mu)$ and recovered $R_{\text{mean}}(\lambda, \mu)$ individuals. Afterwards, given an observation with $I$ infected and $R$ recovered individuals, we find

$$ (\lambda^*, \mu^*) = \arg\min_{\lambda, \mu} (I - I_{\text{mean}}(\lambda, \mu))^2 + (R - R_{\text{mean}}(\lambda, \mu))^2. $$

We also repeated the procedure using the median instead of the mean (and thus computing $I_{\text{median}}$ and $R_{\text{median}}$). In figure 11, we show the distributions of $\lambda^*$ and $\mu^*$ found by the above procedure based on 200 epidemic realizations with $\lambda = 0.6$ and $\mu = 0.5$, along the same distribution as found by the interleaved BP gradient ascent of
The patient-zero problem with noisy observations

Figure 11. Comparison of inference of epidemic parameters for 200 random realizations with $\lambda = 0.6$, $\mu = 0.5$ between BP and the naive method consisting of finding the couple $(\lambda^*, \mu^*)$ that are closest in terms of mean (resp. median) number of infected and recovered individuals in euclidean distance. The distributions for the inference with BP correspond to the fifth example reported in figure 9 and the first in figure 10.

the likelihood function. The results show that the BP-based procedure is able to infer the correct parameters $\lambda = 0.6$ and $\mu = 0.5$ with much higher accuracy.

6. Conclusions

BP was recently proposed as an efficient tool for the inference of the origins of an epidemic propagation on graphs from a snapshot of the system at a later time. In the present work, we generalized the analysis to more realistic cases in which observations are imperfect. Experimental results show that BP performs well even in the presence of strong sources of uncertainty such as observational noise, inability to distinguish between observed states and uncertainty of the intrinsic model parameter. We provided an exact solution on acyclic graphs for the first two problems and a variational solution to compute the gradient of the likelihood of the parameters. The latter can be employed to find local maxima of the likelihood function. We also characterized, by means of simulations, the amount of information that can be extracted on random graphs of various types, both on the past origin of the epidemic and on the missing bits on the present-time observation. Besides giving an excellent algorithmic answer to these questions, related to the past and the present of an observed epidemic, the scheme can be easily generalized to give accurate predictions about the future evolution of an outbreak from which only a partial observation (noisy and/or incomplete) of the current state is available. Work is in progress in this direction.

Acknowledgments

The authors acknowledge European Grant ERC No. 267915 and Italian FIRB Project No. RBFR10QUW4.

doi:10.1088/1742-5468/2014/10/P10016
Appendix A. Efficient BP updates

An efficient form for the update equations of the $\psi_i$ factor nodes is the following:

$$p_{\psi_i \rightarrow j} \left( \left( t_i^{(j)}, t_{ji}, g_i^{(j)} \right) \right) \propto \sum_{g_i, t_i} \sum_{t_i^{(k)}, j_{ki}, g_i^{(k)}} m_{i \rightarrow \psi_i} \left( t_i, g_i \right) \prod_{k \in \partial \backslash i} m_{k \rightarrow \psi_i} \left( t_i^{(k)}, t_{ki}, g_i^{(k)} \right)$$

$$\times \sum_{t_{ji}, k \in \partial \backslash j} m_{j \rightarrow \psi_i} \left( t_i, g_i \right)$$

$$\propto m_{i \rightarrow \psi_i} \left( t_i^{(j)}, g_i^{(j)} \right) \prod_{t_{ki}, k \in \partial \backslash j} m_{k \rightarrow \psi_i} \left( t_i^{(j)}, t_{ki}, g_i^{(j)} \right)$$

$$\times \left[ \delta \left( t_i^{(j)}, 0 \right) + \delta \left( t_i^{(j)}, 1 + \min_{k \in \partial i} \left\{ t_{ki} \right\} \right) \right]$$

$$\propto \delta \left( t_i^{(j)}, 0 \right) m_{i \rightarrow \psi_i} \left( 0, g_i^{(j)} \right) \prod_{t_{ki}, k \in \partial \backslash j} m_{k \rightarrow \psi_i} \left( 0, t_{ki}, g_i^{(j)} \right)$$

$$+ m_{i \rightarrow \psi_i} \left( t_i^{(j)}, g_i^{(j)} \right) \mathbb{I} \left( t_i^{(j)} \leq t_{ji} + 1 \right)$$

$$\times \prod_{k \in \partial \backslash j, t_{ki} \geq t_i^{(j)} + 1} m_{k \rightarrow \psi_i} \left( t_i^{(j)}, t_{ki}, g_i^{(j)} \right)$$

$$- m_{i \rightarrow \psi_i} \left( t_i^{(j)}, g_i^{(j)} \right) \mathbb{I} \left( t_i^{(j)} < t_{ji} + 1 \right)$$

$$\times \prod_{k \in \partial \backslash j, t_{ki} > t_i^{(j)} + 1} m_{k \rightarrow \psi_i} \left( t_i^{(j)}, t_{ki}, g_i^{(j)} \right)$$

where in (A.3) we use the fact that

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

To switch to the simplified representation with $\sigma_{ji}, \sigma_{ij}$ variables defined in (19) instead of $t_{ji}, t_{ij}$ ones, we proceed as follows. In equation (A.3), we can easily group the sums over different configurations of $t_{ki}, t_i^{(j)}$ and write:

$$p_{\psi_i \rightarrow j} \left( \left( t_i^{(j)}, \sigma_{ji}, g_i^{(j)} \right) \right) \propto \delta \left( t_i^{(j)}, 0 \right) m_{i \rightarrow \psi_i} \left( 0, g_i^{(j)} \right) \prod_{k \in \partial \backslash j} m_{k \rightarrow \psi_i} \left( 0, \sigma_{ki}, g_i^{(j)} \right)$$

$$+ m_{i \rightarrow \psi_i} \left( t_i^{(j)}, g_i^{(j)} \right) \prod_{\sigma_{ji} = 1, 2} \sum_{k \in \partial \backslash j, \sigma_{ki} = 1, 2} m_{k \rightarrow \psi_i} \left( t_i^{(j)}, \sigma_{ki}, g_i^{(j)} \right)$$

$$- m_{i \rightarrow \psi_i} \left( t_i^{(j)}, g_i^{(j)} \right) \prod_{\sigma_{ji} = 2} \sum_{k \in \partial \backslash j} m_{k \rightarrow \psi_i} \left( t_i^{(j)}, 2, g_i^{(j)} \right).$$
Similarly, the outgoing message to the \((t_i, g_i)\) variable node is:

\[
p_{\psi_i \rightarrow t}(t_i, g_i) \propto \delta(t_i, 0) \prod_{k \in \partial t} m_{k \rightarrow \psi_i}(0, \sigma_{ki}, g_i) + \prod_{k \in \partial t} \sum_{\sigma_{ki}=1,2} m_{k \rightarrow \psi_i}(t_i, \sigma_{ki}, g_i)
- \prod_{k \in \partial t} m_{k \rightarrow \psi_i}(t_i, 2, g_i)
\]

(A.5)

In the simplified \((t, \sigma, g)\) representation for the messages, the update equation for the \(\phi_{ij}\) nodes reads:

\[
p_{\phi_{ij} \rightarrow j}(t_j, \sigma_{ij}, g_j) \propto \sum_{t_i, \sigma_{ji}, g_i} \Omega(t_i, t_j, \sigma_{ij}, \sigma_{ji}, g_i) m_{i \rightarrow \phi_{ij}}(t_i, \sigma_{ji}, g_i) \]

(A.6)

where

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

(A.7)

and

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

(A.8)

Simple algebra and precalculation of terms in (A.6)–(A.8) bring a significant optimization for updates involving the factor node \(\phi_{ij}\) down to \(O(TG^2)\) operations per update. We show in algorithm 1 a possible pseudocode implementation of this optimization.

**Appendix B. Gradient ascent method for the inference of the epidemic parameters**

The free energy of the system can be approximated with the Bethe free energy, which can be expressed as a sum of local terms depending on the BP messages.

\[
-f = \sum_a f_a + \sum_i f_i - \sum_{(ia)} f_{(ia)}
\]

(B.1)

where

\[
f_a = \log \left( \sum_{\{z_i\}_{i \in \partial a}} F_a(\{z_i\}_{i \in \partial a}) \prod_{i \in \partial a} m_{i \rightarrow a}(z_i) \right)
\]

(B.2)

\[
f_{(ia)} = \log \left( \sum_{z_i} m_{i \rightarrow a}(z_i) p_{F_{a \rightarrow i}}(z_i) \right)
\]

(B.3)

\[
f_i = \log \left( \sum_{z_i} \prod_{b \in \partial i} p_{F_{b \rightarrow i}}(z_i) \right).
\]

(B.4)

\[\text{doi:10.1088/1742-5468/2014/10/P10016}\]
Algorithm 1. Update of factor $\phi_{ij}$. This routine computes the output message $p_{\phi_{ij}}(t_j, \sigma_{ij}, g_j)$ from the input message $m_{i\to \phi_{ij}}(t_i, \sigma_{ji}, g_i)$.

For simplicity, the value $t = \infty$ is stored in position $T_{inf} = T + 1$ on the messages. This update can be used with a transmission probability $\lambda_t$ that can eventually depend on time.

for $\sigma_i = 0$ to 2 do
  for $g_i = 0$ to G do
    for $t_i = 0$ to $T_{inf}$ do
      $H(t_i, \sigma_i) += m_{i\to \phi_{ij}}(t_i, \sigma_i, g_i)$
    end for
  end for
  $R(T_{inf}, \sigma_i) \leftarrow H(T_{inf}, \sigma_i)$
end for

for $t = T$ to 0 do
  $R(t, \sigma_i) += H(t, \sigma_i)$
end for

for $g_j = 0$ to G do
  for $t_j = 0$ to $T_{inf}$ do
    $t_n \leftarrow \min(t_j + g_j, T)$
    clear($g0, q0$)
    $p \leftarrow 1$
    for $t = t_j$ to $t_n$ do
      $q0(t) \leftarrow p$
      $g0(t) \leftarrow g0(t - 1) + p \cdot \lambda_t$
      $p \leftarrow p \cdot (1 - \lambda_t)$
    end for
    $q0(t_n + 1) \leftarrow p$
    for $t_i = t_j + 1$ to $t_n + 1$ do
      $z0 \leftarrow g0(t_i - 2) - g0(t_j - 1)$
      $z1 \leftarrow q0(t_i - 1) \cdot \lambda_{t_i-1}$
      $z2 \leftarrow q0(t_n + 1) + g0(t_n) - g0(t_i - 1)$
      $p_{\phi_{ij}}(t_j, 2, g_j) += z0 \cdot H(t_i, 0) + z1 \cdot H(t_i, 1) + z2 \cdot H(t_i, 2)$
      $U(t_i, 0) += z0 \cdot m_{i\to \phi_{ij}}(t_j, 2, g_j)$
      $U(t_i, 1) += z1 \cdot m_{i\to \phi_{ij}}(t_j, 2, g_j)$
      $U(t_i, 2) += z2 \cdot m_{i\to \phi_{ij}}(t_j, 2, g_j)$
    end for
    if $t_n + 2 \leq T_{inf}$ then
      $z0 \leftarrow g0(t_n)$
      $z2 \leftarrow q0(t_n + 1)$
      $p_{\phi_{ij}}(t_j, 2, g_j) += z0 \cdot R(t_n + 2, 0) + z2 \cdot R(t_n + 2, 2)$
      $S(t_n + 2, 0) += z0 \cdot m_{i\to \phi_{ij}}(t_j, 2, g_j)$
      $S(t_n + 2, 2) += z2 \cdot m_{i\to \phi_{ij}}(t_j, 2, g_j)$
    end if
  end for
end for

(Continued)
The computation of the gradient of the free energy deserves some special attention. As $f$ is a function of all the BP messages, one could argue that this message also depends on the model parameters at every step in the BP algorithm. Actually, there is no need to consider this implicit $(\lambda, \mu)$ dependence if BP has reached its fixed point, which is when BP equations are satisfied and the messages are nothing but Lagrange multipliers with respect to the constraint minimization of the Bethe free energy functional [30]. In our scheme, the only explicit dependence of free energy on epidemic parameters is in the factor node term $f_a$’s involving compatibility functions $\phi_{ij} = \omega_{ij} (t_{ij} - t_i | g_i) \omega_{ji} (t_{ji} - t_j | g_j)$ and $G_i (g_i) = \mu_i (1 - \mu_i)^{a_i}$ and the gradient can be computed very easily. For the $\phi_{ij}$ nodes we have:

$$\frac{\partial f_{\phi_{ij}}}{\partial \lambda} = \frac{\partial \phi_{\lambda}}{\partial \lambda} + \sum_{t_i, t_{ji}, g_i, t_{ij}, g_j} \frac{\partial \phi_{ij}}{\partial \lambda}(t_i, t_{ji}, g_i, t_{ij}, g_j) m_i \to \phi_{ij}(t_i, t_{ji}, g_i) m_j \to \phi_{ij}(t_j, t_{ij}, g_j)$$

where

$$\frac{\partial \phi_{ij}}{\partial \lambda} = \begin{cases} 
1 & t_i < t_j \text{ and } t_i = t_j < t_i + g_i \\
- (g_i - t_i) \lambda (1 - \lambda)^{g_i - t_i - 1} & t_i < t_j \text{ and } t_i < t_j = t_i + g_i \\
(1 - \lambda)^{t_i - t_j} - (t_{ij} - t_i) \lambda (1 - \lambda)^{t_{ij} - t_i - 1} & t_j < t_i \text{ and } t_j < t_i < t_i + g_i \\
- (g_j - t_j) \lambda (1 - \lambda)^{g_j - t_j - 1} & t_j < t_i \text{ and } t_j < t_i = t_j + g_j \\
(1 - \lambda)^{t_i - t_j} - (t_{ji} - t_j) \lambda (1 - \lambda)^{t_{ji} - t_j - 1} & t_j < t_i \text{ and } t_j < t_{ji} = t_j + g_j \\
0 & \text{else}
\end{cases}$$

And the gradient with respect to the messages, equation (B.6) takes the form:

$$\frac{\partial \phi_{ij}}{\partial \lambda} = \begin{cases} 
\chi(t_i, t_j, \sigma_{ij}, g_i) & t_i < t_j \text{ and } \sigma_{ij} = 2, \sigma_{ij} \neq 2 \\
\chi(t_i, t_j, \sigma_{ij}, g_i) - (g_i + 1) (1 - \lambda)^{g_i} & t_i < t_j, \sigma_{ij} = 2, \sigma_{ij} = 2 \\
\chi(t_j, t_i, \sigma_{ij}, g_j) & t_i < t_j, \sigma_{ij} = 2, \sigma_{ij} \neq 2 \\
\chi(t_j, t_i, \sigma_{ij}, g_j) - (g_j + 1) (1 - \lambda)^{g_j} & t_i < t_j, \sigma_{ij} = 2, \sigma_{ij} = 2 \\
0 & \text{otherwise}
\end{cases}$$
The patient-zero problem with noisy observations

where

$$\chi(t_1, t_2, \sigma, g) = \sum_{t=t_1}^{t_1+g} \delta(\sigma(t_2, t), \sigma) (1-\lambda)^{t-t_1} - (t-t_1) \lambda (1-\lambda)^{t-t_1-1}. \quad (B.8)$$

For the $G_i$ nodes we have:

$$\frac{\partial f_{G_i}}{\partial \mu} = \sum_{g_i} \tilde{G}_i(g_i) m_{i-G}(g_i) \sum_{g_i} G_i(g_i) m_{i-G}(g_i) \quad (B.9)$$

where

$$\tilde{G}_i(g_i) = \begin{cases} (1-\mu)^{g_i} - g_i \mu (1-\mu)^{g_i-1} & : g_i < G \\ G - G (1-\mu)^{G-1} & : g_i = G. \end{cases} \quad (B.10)$$

References

[1] Bailey N T J 1975 *The Mathematical Theory of Infectious Diseases and Its Applications* (London: Griffin)
[2] Kermack W O and McKendrick A G 1927 *Proc. R. Soc. Lond. A* **115** 700
[3] Isella L, Stehlé J , Barrat A, Cattuto C, Pinton J-F and Van den Broeck W 2011 *J. Theor. Biol.* **271** 166
[4] Rocha L E C, Liljeros F and Holme P 2010 *Proc. Natl Acad. Sci.* **107** 5706
[5] Danon L, Ford A P, House T, Jewell C P, Keeling M J, Roberts G O, Ross J V and Vernon M C 2011 *Interdiscip. Perspect. Infect. Dis.* **2011** 284909
[6] Marathe M and Vallikanti A K S 2013 *Commun. ACM* **56** 88–96
[7] Shah D and Zaman T 2010 *ACM Sigmetrics Perform. Eval. Rev.* **38** 203
[8] Comin C H and da Fontoura Costa L 2011 *Phys. Rev. E* **84** 056105
[9] Shah D and Zaman T 2011 *IEEE Trans. Inform. Theory* **57** 5163–81
[10] Fioriti V and Chinnici M 2012 arXiv:1211.2333
[11] Pinto P C, Thiran P and Vetterli M 2012 *Phys. Rev. Lett.* **109** 068702
[12] Antulov-Fantulin N, Lancic A, Stefanic H, Sikic M and Smuc T 2013 Statistical inference framework for source detection of contagion processes on arbitrary network structures
[13] Dong W, Zhang W and Tan C W 2013 *IEEE Int. Symp. on Information Theory Proc.* (Istanbul, 7–12 July 2013) pp 2671–5
[14] Lokhov A Y, Mézard M, Ohta H and Zdeborová L 2014 *Phys. Rev. E* **90** 012801
[15] Luo W, Tay W P and Leng M 2013 *IEEE Trans. Signal Process.* **61** 2850
[16] Zhu K and Ying L 2013 *Information Theory and Applications Workshop* (San Diego, CA, 10–15 February 2013) pp 1–9
[17] Karamchandani N and Franceschetti M 2013 *IEEE Int. Symp. on Information Theory Proc.* (Istanbul, 7–12 July 2013) pp 2184–8
[18] Altarelli F, Braunstein A, Dall’Asta L, Lage-Castellanos A and Zecchina R 2014 *Phys. Rev. Lett.* **112** 118701
[19] Karrer B and Newman M E J 2010 *Phys. Rev. E* **82** 016101
[20] Milling C, Caramanis C, Mannor S and Shakkottai S 2013 *Proc. 14th ACM Int. Symp. on Mobile ad Hoc Networking and Computing* (Bangalore, India, 29 July–1 August 2013) pp 177–86
[21] Meiron E A, Milling C, Caramanis C, Mannor S, Orda A and Shakkottai S 2014 arXiv:1402.1263
[22] Kermack W O and McKendrick A G 1932 *Proc. R. Soc. A* **138** 55
[23] Bayati M, Shah D and Sharma M 2008 *IEEE Trans. Inform. Theory* **54** 1241
[24] Bayati M, Braunstein A and Zecchina R 2008 *J. Math. Phys.* **49** 125206
[25] Gamarnik D, Shah D and Wei Y 2010 *Belief Propagation for Min-Cost Network Flow: Convergence and Correctness* (Society for Industrial and Applied Mathematics) National Science Foundation (US) (Project CMMI-0726734)
[26] Weiss Y and Freeman W T 2001 *Neural Comput.* **13** 2173
[27] Bayati M and Nair C 2006 A rigorous proof of the cavity method for counting matchings *Allerton Conf. on Communication, Control and Computing* (Allerton)

doi:10.1088/1742-5468/2014/10/P10016
[28] Altarelli F, Braunstein A, Dall’Asta L and Zecchina R 2013 Phys. Rev. E 87 062115
[29] Altarelli F, Braunstein A, Dall’Asta L and Zecchina R 2013 J. Stat. Mech. P09011
[30] Yedidia J S, Freeman W T and Weiss Y 2000 NIPS 13 689–95
[31] Mézard M and Montanari A 2009 Information, Physics and Computation (New York: Oxford University Press)
[32] Barabási A-L and Albert R 1999 Science 286 509