



**Bayes-optimal inference for spreading processes on random networks**Davide Ghio <sup>1</sup>, Antoine L. M. Aragon,<sup>2</sup> Indaco Biazzo <sup>2</sup> and Lenka Zdeborová<sup>2</sup><sup>1</sup>*Ide PHICS Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rte Cantonale, 1015 Lausanne, Switzerland*<sup>2</sup>*SPOC Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rte Cantonale, 1015 Lausanne, Switzerland*

(Received 6 April 2023; accepted 25 September 2023; published 16 October 2023)

We consider a class of spreading processes on networks, which generalize commonly used epidemic models such as the SIR model or the SIS model with a bounded number of reinfections. We analyze the related problem of inference of the dynamics based on its partial observations. We analyze these inference problems on random networks via a message-passing inference algorithm derived from the belief propagation (BP) equations. We investigate whether said algorithm solves the problems in a Bayes-optimal way, i.e., no other algorithm can reach a better performance. For this, we leverage the so-called Nishimori conditions that must be satisfied by a Bayes-optimal algorithm. We also probe for phase transitions by considering the convergence time and by initializing the algorithm in both a random and an informed way and comparing the resulting fixed points. We present the corresponding phase diagrams. We find large regions of parameters where even for moderate system sizes the BP algorithm converges and satisfies closely the Nishimori conditions, and the problem is thus conjectured to be solved optimally in those regions. In other limited areas of the space of parameters, the Nishimori conditions are no longer satisfied and the BP algorithm struggles to converge. No sign of a phase transition is detected, however, and we attribute this failure of optimality to finite-size effects. The article is accompanied by a Python implementation of the algorithm that is easy to use or adapt.

DOI: [10.1103/PhysRevE.108.044308](https://doi.org/10.1103/PhysRevE.108.044308)**I. INTRODUCTION**

Spreading models on graphs are ubiquitous in many scientific domains. Examples range from modeling epidemic spreading processes [1], diffusion or relaxing out of equilibrium dynamics in physical systems [2], social dynamics processes, and spreading processes in technological systems [3].

In inference on spreading models, we typically want to recover some missing information, having a limited amount of observations on the system. Examples are the identification of the sources of a spreading process (the well-known patient-zero problem) [4,5], epidemic mitigation efforts [6–8], or the anonymization of Bitcoin transactions [9]. Inference problems on spreading models have a natural representation in a Bayesian framework. In the Bayesian approach, the spreading model is the prior, the partial observations on the system induce the likelihood, and the inference problem reduces to the computation of marginals of the posterior distribution. We will give a more precise formulation of the Bayesian approach to the problems and the assumptions in Sec. II.

A large volume of literature in different domains undertakes inference problems on spreading processes. Many are dedicated to the patient-zero problem, where the goal is to infer the nodes (patients) that start the spreading process. This problem was introduced in two different application areas [10,11]. In Ref. [12] the belief propagation (BP) equations to solve these inference problems were derived. Nevertheless, regarding their range of applicability and generality, we believe these results are not well shared among communities beyond the statistical physics ones.

This is, in particular, striking for two reasons. First, the BP-based inference algorithm obtained better results than competitors across various cases, and graph structures [6,13–15]. Second, on sparse random graphs the BP inference algorithms designed for other problems were argued to obtain an asymptotically exact estimation of the marginals of the posterior and thus the Bayes-optimal inference; see Refs. [16–18] for examples. One can thus expect that also the algorithm of [13] is Bayes-optimal on sparse random graphs and that there does not exist an algorithm that finds, on average, a more accurate solution.

In this paper, we investigate whether the BP for inference in spreading models [13] provides Bayes-optimal inference on large sparse random graphs. In particular, we rederive the algorithm, illustrating that it indeed is an instance of a standard BP algorithm applied in a setting where it is expected to be asymptotically optimal. We then investigate in detail its convergence from both random and informed initializations seeking a possible algorithmically hard phase, which we do not find. We study the phase diagram of inference problems, identifying regions of parameters where the optimal estimation error is large and others where it is rather small. We also evaluate the so-called Nishimori conditions [19,20] that must be satisfied for the Bayes-optimal inference. To facilitate the broad usage of these types of algorithms, we also share a basic Python code [21] to solve several inference problems that can be easily modified to deal with different spreading models.

This paper is structured as follows: in Sec. II we introduce the definition and the assumptions on the spreading and observations models; then the Bayesian approach to inference problems is presented. Furthermore, we define the observables and their optimal estimators to measure the performances of

the algorithm, the concept of Bayes-optimal inference, and the Nishimori conditions. In Sec. III the BP equations are derived and in Sec. IV we show the results of the algorithm on some specific inference problems and the corresponding phase diagrams.

Finally, in Sec. V we focus on some small regions of parameters for which the BP algorithm does not seem Bayes-optimal, in the sense that it does not converge and the Nishimori conditions are no longer respected. We conjecture that this behavior is due to finite-size effects.

## II. BAYESIAN INFERENCE FOR LOCAL SPREADING MODELS ON NETWORKS

### A. Local spreading models on networks

We consider a network or equivalently a graph  $G(V, E)$  with  $N = |V|$  nodes. Each node  $i$  carries a state variable  $x_i^t$  that may change at every time step  $t = 0, 1, 2, \dots, T$ . The variable  $x_i^t$  can take values from an ordered set of  $p + 1$  elements. We consider a class of models where at most  $p$  changes are possible within the considered time range and each change must be into the consecutive state, according to the ordering of the set.

For such a class of models (examples below) we represent the time trajectory of a node  $i$  by a set of times  $\vec{t}_i = \{t_i^1, \dots, t_i^p\}$  defined as the last time steps in which the node is in a certain state (i.e., it will be in a different state the next time step). The entire evolution of the system  $\vec{x}$ , of variables  $\{x_i^t\}_{i=1, \dots, N}^{t=0, \dots, T-1}$ , can thus be mapped to the variables  $\vec{t} = \{\vec{t}_i\}_{i=1, \dots, N}$ .

In this work, we consider all stochastic spreading processes for which the probability of the evolution can be written in the form

$$P(\vec{t}|\lambda) = \frac{1}{Z_{\text{prior}}} \prod_{i=1}^N \Psi_i(\vec{t}_i, \{\vec{t}_j\}_{j \in \partial i} | \lambda), \quad (1)$$

where  $Z_{\text{prior}}$  is the normalization,  $\lambda$  is a set of parameters and  $\partial i$  is the set of neighbors of the node  $i$  in the graph  $G(V, E)$ . The function  $\Psi_i$  can be an arbitrary nonnegative function. In this work, we assume the spreading model (1) that governs the process is known, and it depends on a set of parameters  $\lambda$  that is assumed known.

*Example 1: SIR.* A canonical model for epidemic spreading on a network is the SIR model, where each individual node can be in one of the 3 states: susceptible ( $S$ ), infectious ( $I$ ), and recovered ( $R$ ). The spreading process is stochastic and, more specifically, a Markov Chain: at time step  $t$ , each infectious node  $i$  may infect any one of its susceptible contacts  $\{j \in \partial i : x_j^t = S\}$  with probability  $\lambda_{ij}(t) \in [0, 1]$ . We call  $t_i$  the time of infection of node  $i$ , i.e.,  $x_i^{t_i} = S$  while  $x_i^{t_i+1} = I$ . Then, for each time  $t > t_i$ , node  $i$  can recover (i.e., change to the state  $R$ , earliest we can have  $x_i^{t_i+2} = R$ ) with probability  $\mu_i \in [0, 1]$ , and again we define the time of recovery  $r_i$  as the last time  $t$  in which  $x_i^t = I$ . Any given node can only undergo the transitions  $S \rightarrow I \rightarrow R$ . We can easily convince ourselves that this epidemic model can be described by the model (1), by taking  $\vec{t}_i = \{t_i, r_i\}$ , and as  $\lambda$  the set of all parameters  $\lambda_{ij}(t)$  and  $\mu_i$ . We also note that the SI model where there is no recovered state

is easily obtained as a special case of the SIR with  $\mu_i = 0$  for every node  $i$ .

*Example 2: deterministic-SIR.* More complicated models can be considered, for instance, non-Markovian models. As an example, in this work, we consider a slight variant of the SIR model, which we call *deterministic-SIR* (dSIR) or *SIR with deterministic recovery*. The model is no longer a Markov chain and is characterized by a single stochastic transition, i.e., the one between the  $S$  and the  $I$  states, which stays the same as the SIR model. The second transition, between the  $I$  and the  $R$  states, is modeled differently than in the standard SIR model. After being infected, the node  $i$  remains infectious for a fixed time  $\Delta_i$ , and then becomes recovered: in this case the infectivity of individual  $i$  switches to zero after  $\Delta_i$  time steps, and thus there is no stochasticity in the recovery process. In the following, we will consider the case in which  $\Delta_i = \Delta \forall i$ . In general, we can consider arbitrarily complex cases in which the infectivity  $\lambda_{ij}(t, t - t_i)$  (i.e., the probability of node  $i$  infecting node  $j$  at time  $t$ , if  $j$  is susceptible and  $i$  is infectious) changes both in time and with the time-delay from the time of infection  $t_i$ .

*Example 3: SIS model with up to  $\lceil \frac{p}{2} \rceil - 1$  reinfections.* The SIS model is another very commonly considered model for epidemic spreading, where the infected nodes return to a susceptible state after some time. The SIS model also falls under the framework considered in Eq. (1) if we restrict the number of possible reinfections. In particular, if we assume that only  $\lceil \frac{p}{2} \rceil - 1$  reinfections are possible, then we can again formulate the spreading in terms of the  $p$ -dimensional time trajectories  $\vec{t}_i$  and the probability over them given by Eq. (1). Such a model is suitable for infections where immunity is temporary or nonexistent, assuming that only a negligible fraction of the population gets reinfected more than  $\lceil \frac{p}{2} \rceil - 1$  times. For example, for COVID-19 this seems a reasonable assumption for say  $p \approx 20$ .

### B. The Bayesian inference framework

The high-level idea of Bayesian inference for spreading processes is that only some partial observations about the spreading are available on a given network, and the aim is to recover as much information about the spreading as possible.

We now define what kind of observations we consider. We assume site-dependent, factorized observations on the system  $\mathcal{O} = \{\mathcal{O}_i\}_{i=1, \dots, N}$ , that are also independent of the parameters  $\lambda$  and the other transition times  $\vec{T} \setminus \vec{t}_i$  when conditioned on the transition times of the node  $\vec{t}_i$ . Then the likelihood of a set of observations can be written as  $P(\mathcal{O}|\vec{t}, \lambda) = \prod_{i=1}^N P(\mathcal{O}_i|\vec{t}_i)$ , with  $P(\mathcal{O}_i|\vec{t}_i)$  a known probabilistic law. In this notation,  $\mathcal{O}_i$  includes all the partial information we get from the observations on the node  $i$ , and the case in which the node is not observed corresponds to  $\mathcal{O}_i = \emptyset$ , for which trivially  $P(\mathcal{O}_i = \emptyset|\vec{t}_i) = 1 \forall \vec{t}_i$ . Instead, if for instance, we observe that a node  $i$  has been infected at time  $T_{\text{obs}}$ , then we have  $P(\mathcal{O}_i|\vec{t}_i) = \mathbb{I}[t_i = T_{\text{obs}}]$ , where  $\mathbb{I}[\cdot]$  is the identity function that is equal to one when the condition in the argument is satisfied and zero otherwise.

In a Bayesian setting, considering  $P(\vec{t}|\lambda)$  as the prior and  $P(\mathcal{O}|\vec{t}, \lambda)$  as the likelihood, we can then recover the configuration through the posterior probability distribution that from

the Bayes rule reads

$$P(\vec{t}|\mathcal{O}, \lambda) = \frac{1}{P(\mathcal{O}|\lambda)} P(\vec{t}|\lambda) P(\mathcal{O}|\vec{t}, \lambda) \quad (2)$$

$$= \frac{1}{Z(\mathcal{O})} \prod_{i=1}^N \Psi_i(\vec{t}_i, \{\vec{t}_j\}_{j \in \partial_i} | \lambda) \prod_{i=1}^N P(\mathcal{O}_i | \vec{t}_i) \quad (3)$$

$$= \frac{1}{Z(\mathcal{O})} \prod_{i=1}^N \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_j\}_{j \in \partial_i}, \mathcal{O}_i | \lambda), \quad (4)$$

where we defined the normalization constant  $Z(\mathcal{O}) \equiv Z_{\text{prior}} P(\mathcal{O} | \lambda) = Z_{\text{prior}} \sum_{\vec{t}} \prod_{i=1}^N \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_j\}_{j \in \partial_i}, \mathcal{O}_i | \lambda)$  and  $\tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_j\}_{j \in \partial_i}, \mathcal{O}_i | \lambda) \equiv \Psi_i(\vec{t}_i, \{\vec{t}_j\}_{j \in \partial_i} | \lambda) P(\mathcal{O}_i | \vec{t}_i)$ .

Given the observations, there are different properties of the spreading that we may want to infer. For instance, one aim is to identify the sources of the infection (the patient(s) zero problem). This reduces to estimating the marginals over the individuals (variables) to be infectious at time step zero. Another possible goal is to assess the epidemic risk of being infectious at time  $t$ , for this, we need to compute the probability for each variable to be infectious at time  $t$ . A broad range of other possible goals reduces to computing marginals of the posterior, or averages of quantities over the posterior probability distribution.

The number of configurations of the systems grows exponentially with the system size, making the exhaustive computation of marginals and averages impossible for a size larger than a few dozen of nodes. In this work, we aim to show how the BP equations are used to estimate efficiently the marginals of this posterior probability distribution in a range of problems where the network of interactions among the variables is random and sparse.

### 1. The optimal overlap

Let us define the overlap  $O(\mathbf{x}, \mathbf{y})$  between two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , of discrete entries and length  $N$ , as the fraction of agreeing elements between the two:

$$O(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i, y_i}. \quad (5)$$

In particular, we are interested in the quantity  $O(\hat{\mathbf{x}}^t, \mathbf{x}^{*,t})$ , where  $\mathbf{x}^{*,t}$  is the ground-truth of the state of the system at time  $t$ , and  $\hat{\mathbf{x}}^t$  is a generic estimator of it. In general, we do not know the ground truth; assuming that it is distributed accordingly to the posterior, the best thing we can do (in a Bayesian framework) is to estimate the so-called mean overlap:

$$\text{MO}(\hat{\mathbf{x}}^t) = \frac{1}{N} \sum_{\mathbf{x}^t} P(\mathbf{x}^t | \mathcal{O}) \sum_{i=1}^N \delta_{x_i^t, \hat{x}_i^t}, \quad (6)$$

where  $P(\mathbf{x}^t | \mathcal{O})$  can be computed from  $P(\vec{t} | \mathcal{O}, \lambda)$ . Maximizing this quantity over the estimator  $\hat{\mathbf{x}}$  leads to the *maximum mean overlap estimator*

$$\hat{x}_i^{t, \text{MMO}} = \arg \max_{x_i^t} P_i(x_i^t | \mathcal{O}), \quad (7)$$

where  $P_i(x_i^t | \mathcal{O}) = \sum_{\mathbf{x}^t \setminus x_i^t} P(\mathbf{x}^t | \mathcal{O})$  is the marginal probability for the node  $i$  to be in a given state at time  $t$ . In the following,

we will refer to the performances of this estimator simply as  $O_t \equiv O(\hat{\mathbf{x}}^{t, \text{MMO}}, \mathbf{x}^{*,t})$  and  $\text{MO}_t \equiv \text{MO}(\hat{\mathbf{x}}^{t, \text{MMO}})$ , both implicitly depending on the marginal probability distributions.

Starting from these definitions of overlap and mean overlap, we now define the performance parameters which we are going to use in our numerical studies. The idea is to rescale (5) and (6) by comparing the performance of the MMO estimator to the one obtained by running BP without observations, i.e., on the prior, which in the following we will call *random estimator* and can be written as

$$\hat{x}_i^{\text{RND}, t} = \arg \max_{x_i^t} P_i(x_i^t), \quad (8)$$

where  $P_i(x_i^t) = \sum_{\mathbf{x}^t \setminus x_i^t} P(\mathbf{x}^t)$  is the marginal of the prior probability. We write the resulting estimate of the state of the system at time  $t$  as  $\hat{\mathbf{x}}^{\text{RND}, t}$  and we then define the *rescaled overlap* as

$$\tilde{O}_t = \frac{O_t - O(\hat{\mathbf{x}}^{\text{RND}, t}, \mathbf{x}^{*,t})}{1 - O(\hat{\mathbf{x}}^{\text{RND}, t}, \mathbf{x}^{*,t})}, \quad (9)$$

and the *rescaled mean overlap* as

$$\tilde{\text{MO}}_t = \frac{\text{MO}_t - \text{MO}(\hat{\mathbf{x}}^{\text{RND}, t})}{1 - \text{MO}(\hat{\mathbf{x}}^{\text{RND}, t})}. \quad (10)$$

Notice how these performance parameters are defined in such a way that they are negative or zero if and only if BP performs worse or equal to the random estimator, respectively, and are in (0,1] in any other case, with the value one indicating a perfect resolution of the problem.

### 2. The optimal mean-squared error

Another property we want to study is how well our algorithm can estimate the entire trajectory of each node. We will focus on models that feature a single transition time, such as the SI and dSIR models. However, one can extend our analysis to cover models with multiple transition times, such as the SIR and SIS models, by generalizing the following definitions. We consider the squared error between the ground-truth vector  $\mathbf{t}^*$  and an estimator  $\hat{\mathbf{t}}$ , defining

$$\text{SE}(\hat{\mathbf{t}}, \mathbf{t}^*) = \frac{1}{N} \sum_{i=1}^N (\hat{t}_i - t_i^*)^2. \quad (11)$$

As before, since we usually do not know the ground truth, we can assume that  $\mathbf{t}^*$  is distributed according to the posterior probability distribution and then define the *mean-squared error* (averaged on the posterior) as

$$\text{MSE}(\hat{\mathbf{t}}) = \frac{1}{N} \sum_{\mathbf{t}} P(\mathbf{t} | \mathcal{O}) \sum_{i=1}^N (\hat{t}_i - t_i)^2. \quad (12)$$

In this case, minimizing Eq. (12) with respect to the estimator, we obtain the *minimum mean-squared error estimator*

$$\hat{t}_i^{\text{MMSE}} = \sum_{\mathbf{t}} P(\mathbf{t} | \mathcal{O}) t_i = \sum_{t_i} P_i(t_i | \mathcal{O}) t_i, \quad (13)$$

where  $P_i(t_i | \mathcal{O}) = \sum_{\mathbf{t} \setminus t_i} P(\mathbf{t} | \mathcal{O})$  is the marginal probability for the node  $i$  to have a certain time of transition  $t_i$ . In the following, we will refer to the performances of this estimator simply

as  $SE = SE(\hat{\mathbf{t}}^{\text{MMSE}}, \mathbf{t}^*)$  and  $MSE = MSE(\hat{\mathbf{t}}^{\text{MMSE}})$ , both implicitly depending on the marginal probability distributions  $P_i(t_i|\mathcal{O})$ .

We can define the random estimator for the times of transition, as we did for the overlap where we averaged over the prior probability distribution, as

$$\hat{t}_i^{\text{RND}} = \sum_{t_i} P_i(t_i)t_i, \quad (14)$$

where  $P_i(t_i) = \sum_{\mathbf{t} \in \mathcal{T}_i} P(\mathbf{t})$  is the marginal of the prior probability. Then, our performance parameters will be, respectively,

$$R_{SE} = \frac{SE(\hat{\mathbf{t}}^{\text{RND}}, \mathbf{t}^*) - SE}{SE(\hat{\mathbf{t}}^{\text{RND}}, \mathbf{t}^*)} \quad (15)$$

and

$$R_{MSE} = \frac{MSE(\hat{\mathbf{t}}^{\text{RND}}) - MSE}{MSE(\hat{\mathbf{t}}^{\text{RND}})}. \quad (16)$$

Notice how, as for the rescaled overlaps, these performance parameters are defined in such a way that they are negative or zero if and only if BP performs worse or equal to the random estimator, respectively, and are in  $(0, 1]$  in any other case, with the value one indicating a perfect resolution of the problem.

### C. Nishimori conditions

In the Bayes-optimal setting, i.e., when the prior probability distribution and the probability function used to generate the observations are known, it is possible to derive consistency conditions on expectation over the posterior that are called the *Nishimori conditions* in statistical physics of disordered systems [19,20]. In words, the Nishimori conditions state that, under averages over the posterior, we cannot distinguish the ground truth we aim to infer and a random sample from the posterior. These conditions imply certain properties of the optimal estimators that we make explicit below. As a consequence, when we consider an algorithm that approximates the optimal estimators, the Nishimori conditions can serve as a necessary condition for the approximation to be close to the true optimal estimators. This is how we will use these conditions in the present work.

To state the Nishimori conditions, let us consider having a ground-truth configuration  $\mathbf{x}^*$  generated from the prior distribution  $P_{\text{g.t.}}(\mathbf{x}^*)$  and a measurement process leading to observations  $\mathcal{O}$ , generated through the likelihood function  $P_{\text{g.t.}}(\mathcal{O}|\mathbf{x}^*)$ . Let  $\mathbf{x}$  denote a sample from the posterior  $P(\mathbf{x}|\mathcal{O})$ . Then, given an observable  $f(\mathbf{x})$ , we can compute the two expectations:

$$\begin{aligned} \mathbb{E}_{\mathbf{x}^*}[f(\mathbf{x}^*)] &= \sum_{\mathbf{x}^*} f(\mathbf{x}^*)P_{\text{g.t.}}(\mathbf{x}^*) \\ &= \sum_{\mathbf{x}^*, \mathcal{O}} f(\mathbf{x}^*)P_{\text{g.t.}}(\mathbf{x}^*)P_{\text{g.t.}}(\mathcal{O}|\mathbf{x}^*), \end{aligned} \quad (17a)$$

$$\begin{aligned} \mathbb{E}_{\mathcal{O}}\mathbb{E}_{\mathbf{x}|\mathcal{O}}[f(\mathbf{x})] &= \sum_{\mathbf{x}, \mathcal{O}} f(\mathbf{x})P(\mathcal{O})P(\mathbf{x}|\mathcal{O}) \\ &= \sum_{\mathbf{x}, \mathcal{O}} f(\mathbf{x})P(\mathbf{x})P(\mathcal{O}|\mathbf{x}), \end{aligned} \quad (17b)$$

where in the first equation we used the fact that  $\sum_{\mathcal{O}} P_{\text{g.t.}}(\mathcal{O}|\mathbf{x}^*) = 1$  and in the second we applied Bayes theorem. Now we can notice that  $\mathbf{x}$  and  $\mathbf{x}^*$  are two dummy variables, such that when  $P(\mathbf{x}) = P_{\text{g.t.}}(\mathbf{x})$  and  $P(\mathcal{O}|\mathbf{x}) = P_{\text{g.t.}}(\mathcal{O}|\mathbf{x})$  the two expressions (17a) and (17b) coincide, and we say that the Nishimori conditions are satisfied. Therefore, checking that  $\mathbb{E}[f(\mathbf{x}^*)] = \mathbb{E}[f(\mathbf{x})]$  can give us solid evidence that we are in the *Bayes-optimal* case, and thus that the solution the algorithm finds is the best possible in a Bayesian framework.

For the overlaps defined above, the Nishimori conditions lead to checking that the two following quantities coincide:

$$\mathbb{E}[O_t] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[\mathbb{I}[\arg \max_{x'_i} P_i(x'_i|\mathcal{O}) = x_i^{*,t}]], \quad (18a)$$

$$\begin{aligned} \mathbb{E}[\text{MO}_t] &= \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \sum_{\tilde{x}'_i} P_i(\tilde{x}'_i|\mathcal{O}) \mathbb{I}[\arg \max_{x'_i} P_i(x'_i|\mathcal{O}) = \tilde{x}'_i] \right] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \max_{x'_i} P_i(x'_i|\mathcal{O}) \right]. \end{aligned} \quad (18b)$$

Analogously, for the squared errors:

$$\mathbb{E}[SE] = \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \left( \sum_{t_i} P_i(t_i|\mathcal{O})t_i - t_i^* \right)^2 \right], \quad (19a)$$

$$\mathbb{E}[MSE] = \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \sum_{t'_i} P_i(t'_i|\mathcal{O}) \left( \sum_{t_i} P_i(t_i|\mathcal{O})t_i - t'_i \right)^2 \right]. \quad (19b)$$

The above equalities tell us that, in the Bayes-optimal case, on average, the mean of the observable computed from the posterior must be equal to the actual observable calculated on the ground truth. In the following, we will show some cases where the BP algorithm, described in Sec. III, respects them, and we are confident that the solutions found are the best possible on average. Then in Sec. V, we will show other cases where the BP equations struggle to converge, the Nishimori conditions are no longer satisfied, and other algorithms should be used if one aims to reach Bayes optimality.

## III. INFERENCE WITH BELIEF PROPAGATION

The BP equations allow computing marginals exactly when the factor graph associated with a probability distribution is acyclic [22]. In this paper, we consider large random graphs that are locally treelike, i.e., the typical length of loops grows as  $\log N$ , with  $N$  the number of nodes. This is the case of Erdős–Rényi or random regular graphs (RRG) with an average degree of order one [23].

In random locally treelike factor graphs, the BP algorithm is expected to estimate the marginals asymptotically exactly, i.e., with an error vanishing in the limit of  $N \rightarrow \infty$ , as long as the system is in the so-called replica symmetric (RS) phase. As long as we are dealing with inference problems



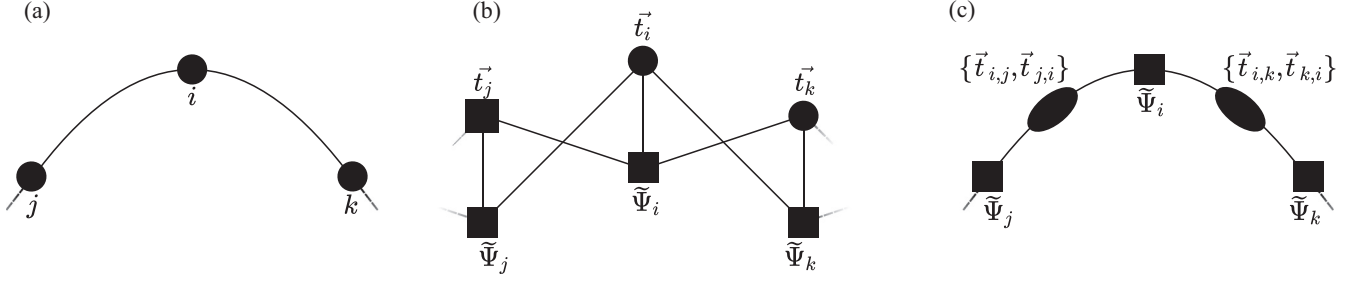


FIG. 1. Factor graph representation of the posterior (2). In panel (a) we show a possible path in a contact network, and in panel (b) its respective naive factor graph representation. In panel (c) we show the representation obtained by grouping the variables as explained in the main text.

in a Bayes-optimal setting, such as we are considering here, the equilibrium properties are in the replica symmetric phase. This follows from the Nishimori conditions. We thus expect the BP equations to give the correct marginals as long as their fixed point satisfies the Nishimori conditions and in the absence of a first-order phase transition that may prevent the BP algorithm from reaching the corresponding fixed point [20]. The presence of a first-order phase transition can be verified by running the BP algorithm both from random and from informed initializations and checking whether they converge to the same fixed point.

In our setting, however, the naive factor graph representation of the posterior probability distribution (2), where the nodes are the time trajectories of each variable and the factors are the set of  $\{\tilde{\Psi}_i\}$ , presents small loops even when the interaction graph among variables is acyclic. The reason is

that the factor  $\tilde{\Psi}_i$  couples  $i$  together with all its neighbors  $k \in \partial i$ , see Fig. 1(b). Instead, we use a different factor graph representation, introducing additional variables. We replicate the variables  $\vec{t}_i$  on each edge of the interaction graph, calling the replicas obtained  $\vec{t}_{i,j}$  with  $j \in \partial i$ , and we add local constraints imposing their equality. Calling  $\vec{\mathbf{t}}_\partial = \{\vec{t}_{i,j}\}_{i=1, \dots, N}^{j \in \partial i}$ , the new representation of the posterior probability distribution is

$$P(\vec{\mathbf{t}}, \vec{\mathbf{t}}_\partial | \mathcal{O}, \lambda) = \frac{1}{Z(\mathcal{O})} \prod_{i=1}^N \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_{j,i}, \vec{t}_{i,j}\}_{j \in \partial i} | \lambda) \times \prod_{j \in \partial i} \mathbb{I}[\vec{t}_{i,j} = \vec{t}_i], \quad (20)$$

and the respective factor graph is represented in Fig. 1(c). The BP equations associated with this posterior distribution are

$$m_{i \rightarrow j}(\vec{t}_{i,j}, \vec{t}_{j,i}) = \frac{1}{Z_{i \rightarrow j}} \sum_{\substack{\vec{t}_i, \\ \{\vec{t}_{i,k}, \vec{t}_{k,i}\}_{k \in \partial i \setminus j}}} \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_{k,i}, \vec{t}_{i,k}\}_{k \in \partial i} | \lambda) \mathbb{I}[\vec{t}_{i,j} = \vec{t}_i] \prod_{k \in \partial i \setminus j} \mathbb{I}[\vec{t}_{i,k} = \vec{t}_i] m_{k \rightarrow i}(\vec{t}_{i,k}, \vec{t}_{k,i}), \quad (21)$$

where the message  $m_{i \rightarrow j}$  goes from the factor node  $\tilde{\Psi}_i$  to  $\tilde{\Psi}_j$  and

$$Z_{i \rightarrow j} = \sum_{\substack{\vec{t}_i, \\ \{\vec{t}_{i,k}, \vec{t}_{k,i}\}_{k \in \partial i}}} \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_{k,i}, \vec{t}_{i,k}\}_{k \in \partial i} | \lambda) \mathbb{I}[\vec{t}_{i,j} = \vec{t}_i] \times \prod_{k \in \partial i \setminus j} \mathbb{I}[\vec{t}_{i,k} = \vec{t}_i] m_{k \rightarrow i}(\vec{t}_{i,k}, \vec{t}_{k,i})$$

is the normalization constant. By imposing the indicator functions and renaming the dummy variables we obtain the final form,

$$m_{i \rightarrow j}(\vec{t}_i, \vec{t}_j) = \frac{1}{Z_{i \rightarrow j}} \sum_{\{\vec{t}_k\}_{k \in \partial i \setminus j}} \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_k\}_{k \in \partial i} | \lambda) \prod_{k \in \partial i \setminus j} m_{k \rightarrow i}(\vec{t}_k, \vec{t}_i). \quad (22)$$

The messages are usually initialized randomly or uniformly and then updated according to Eq. (22) until convergence if they converge, or otherwise when they meet some stopping criterion. In this paper, we will also consider an informed initialization corresponding to the ground truth (that is of

course not available in practice) to check for first-order phase transitions appearing in the problem. The marginals of each variable are then computed [22] as

$$b_i(\vec{t}_i) = \frac{1}{Z_i} \sum_{\{\vec{t}_k\}_{k \in \partial i}} \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_k\}_{k \in \partial i} | \lambda) \prod_{k \in \partial i} m_{k \rightarrow i}(\vec{t}_k, \vec{t}_i), \quad (23)$$

and from the fixed point one can also compute the log-partition function as

$$\log Z = \sum_i \log Z_i - \sum_{(i,j)} \log Z_{(i,j)}, \quad (24)$$

where

$$Z_i = \sum_{\vec{t}_i} \sum_{\{\vec{t}_k\}_{k \in \partial i}} \tilde{\Psi}_i(\vec{t}_i, \{\vec{t}_k\}_{k \in \partial i} | \lambda) \prod_{k \in \partial i} m_{k \rightarrow i}(\vec{t}_k, \vec{t}_i), \quad (25a)$$

$$Z_{(i,j)} = \sum_{\vec{t}_i} \sum_{\vec{t}_j} m_{i \rightarrow j}(\vec{t}_i, \vec{t}_j) m_{j \rightarrow i}(\vec{t}_j, \vec{t}_i). \quad (25b)$$

The computational complexity of the brute-force calculation of the marginals is  $\mathcal{O}(T^{Np})$ , and thus scales exponentially with the system size and the number of possible transitions. Using

the BP equations, one update of all the messages reduces to  $\mathcal{O}(E T^{(d_{\max}-1)p})$  where  $E$  is the number of edges in the interaction graph considered, and  $d_{\max}$  is the maximum degree of the nodes in the interaction graph. Now the computation is no longer exponential in the number of variables, but still exponential in the degree of the graph. Depending on the particular spreading model, more efficient and smaller representations of the time trajectories can be adopted to avoid the exponential dependence on the degree. In Appendix A the BP equations are explicitly written for the deterministic-SIR model, where this simplification holds.

#### IV. BAYES-OPTIMAL PHASE DIAGRAMS

This section presents the results of applying the BP algorithm, described in Sec. III, to the dSIR model on epidemic inference problems. We use the definition of Nishimori identities, as provided in Sec. II, to investigate Bayes-optimality for various observables. The setting of the inference problems is:

*Network.* For simplicity, in the following, we will consider a single ensemble of graphs: the random regular graphs of degree  $d = 3$ , not varying in time (In Appendix B we report some additional results suggesting that considering more structured graphs does not influence considerably the performance of the algorithm). Furthermore, we fix a homogeneous probability of infection  $\lambda$  for each contact, equal for each edge of the interaction graph, and constant in time.

*Sources.* The sources are chosen uniformly at random, i.e., each individual could be infectious at time  $t = 0$  with probability  $\delta$ . Thus, the  $\delta$  parameter controls our simulations' average fraction of sources.

*Observations.* We will focus on two different types of observations, both typically used in the epidemic inference problems setting:

*Sensors.* A fraction  $\rho$  of individuals, called sensors, are chosen uniformly at random. These nodes reveal their entire trajectory in time, which for the dSIR model means just their time of infection  $t_i$ . Such observations were considered for instance in Refs. [11,24].

*Snapshot.* The state of all variables at a specific time  $T_{\text{obs}}$  is known. We get information on whether the node is still susceptible ( $S$ ) or not (either  $I$  or  $R$ ). Such observations were considered for instance in Refs. [12,25].

We will consider two observables, the overlap at time 0,  $O_{t=0}$ , and the squared error, SE. For the first one, in the numerical results, we compute the two following quantities

$$\overline{O_{t=0}} = \frac{1}{N_{\text{sim}}} \sum_{s=1}^{N_{\text{sim}}} \frac{1}{N} \sum_{i=1}^N \mathbb{I}[\arg \max_x b_{i,s}^{t=0}(x) = x_{i,s}^{*,t=0}], \quad (26a)$$

$$\overline{\text{MO}_{t=0}} = \frac{1}{N_{\text{sim}}} \sum_{s=1}^{N_{\text{sim}}} \frac{1}{N} \sum_{i=1}^N \max_x b_{i,s}^{t=0}(x), \quad (26b)$$

where we wrote the sample mean explicitly introducing the simulation index  $s$  and calling  $N_{\text{sim}}$  the total number of simulations. The  $b_i^t(x)$  is the BP estimation of  $P_i(x_i^t | \mathcal{O})$  that can be easily computed from the  $b_i(\vec{t}_i)$ . In the case of the squared error, restricting ourselves to the dSIR model, we need to

consider just a single transition time  $t_i$  for each node, and the marginals  $b_i$  are just one-dimensional vectors.

Analogous to the previous case, we compute

$$\overline{\text{SE}} = \frac{1}{N_{\text{sim}}} \sum_{s=1}^{N_{\text{sim}}} \frac{1}{N} \sum_{i=1}^N \left( \sum_t b_{i,s}(t) t - t_{i,s}^* \right)^2, \quad (27a)$$

$$\overline{\text{MSE}} = \frac{1}{N_{\text{sim}}} \sum_{s=1}^{N_{\text{sim}}} \frac{1}{N} \sum_{i=1}^N \sum_{t'} b_{i,s}(t') \left( \sum_t b_{i,s}(t) t - t' \right)^2. \quad (27b)$$

Then, suppose we are in the Bayes-optimal case. In that case, we must observe that (for  $N_{\text{sim}}$  sufficiently large) these two couples of average values are equal up to their statistical error, ( $\overline{O_{t=0}} = \overline{\text{MO}_{t=0}}$ ) and ( $\overline{\text{SE}} = \overline{\text{MSE}}$ ). In practice, in the following, we show the results for the set of rescaled observables ( $\widetilde{O}_{t=0}, \widetilde{\text{MO}}_{t=0}, R_{\text{SE}}, R_{\text{MSE}}$ ) defined in Sec. II.

Two different initializations of the BP messages are used:

*rnd.* We let the BP algorithm converge at a fixed point in the case of no observations. Then we use the resulting values of the messages as the initial point,

*inf.* We let the BP algorithm converge at a fixed point where the times of infection of all individuals are observed. Then we use the resulting values of the messages as the initial point.

If the algorithm converges to the same fixed point for both of the two initializations, then it indicates an absence of a first-order phase transition and asymptotic optimality of the obtained results, as in, e.g., Ref. [17]. Furthermore, checking the validity of the Nishimori conditions that hold on average in the large size limit provides another evidence that the BP algorithm reached the Bayes-optimal marginals. In the following, we will analyze phase diagrams of various inference settings and for each point, we will check the two initializations and the validity of the Nishimori conditions.

In Fig. 2 we illustrate such a check that is later done for each point in the phase diagrams:

(i) In the left panels, we probe the Nishimori conditions through the rescaled observables defined in Sec. II, namely computing  $R_{\text{SE}} - R_{\text{MSE}}$  and  $\widetilde{O}_t - \widetilde{\text{MO}}_t$  on the single instances and then averaging the values found. By doing this for different sizes  $N$ , we see how, as expected, increasing the size decreases the variance of the differences and concentrates the single-instance values to zero.

(ii) In the right panels, we track the BP algorithm through its iterations for some random instances of the problem, initializing first the messages randomly and then repeating the experiment initializing the messages informatively (see details above). We can see how in each instance showed the two initializations start converging exponentially to the same value after just a dozen of iterations until the difference becomes of the order of the tolerance imposed to check the convergence of BP.

#### A. Phase diagrams for observations via sensors

##### 1. Finding the sources

In this section, the performance in finding the sources of infection on random regular graphs is analysed. The results of rescaled overlap  $\widetilde{O}_{t=0}$  computed with the BP algorithm

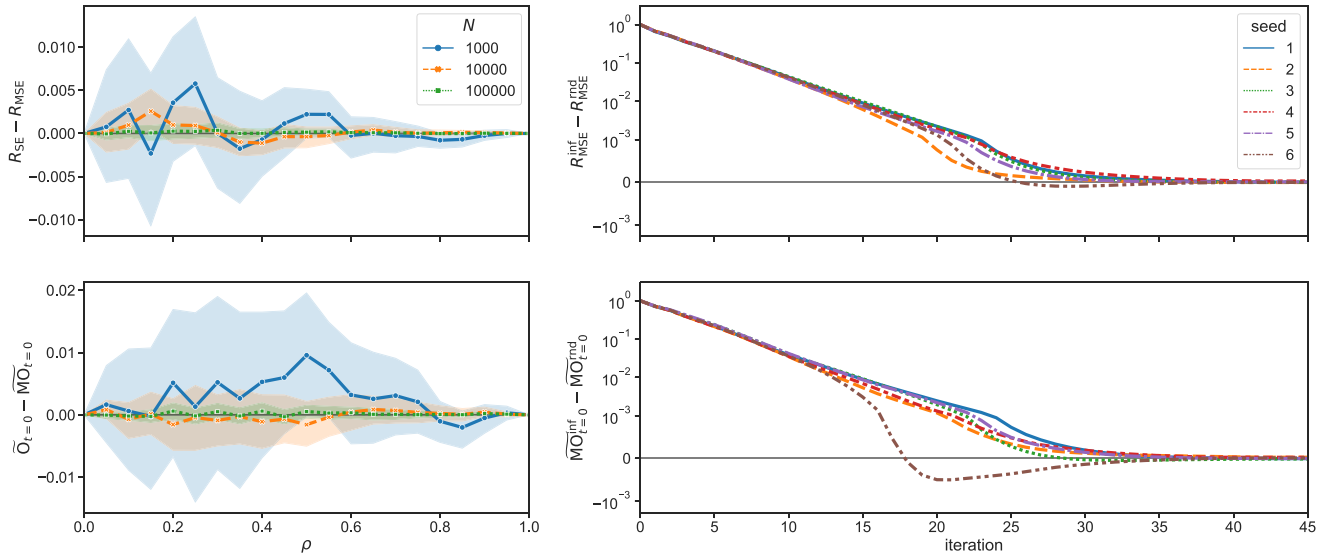


FIG. 2. Nishimori conditions and initializations difference. The setting is the following: we consider SI epidemics with sensors on an ensemble of 3-RRGs with  $\lambda = 0.8$  and  $\delta = 0.1$ . In the left panels, we furthermore vary  $N$  in  $[10^3, 10^5]$  and  $\rho$  in  $[0,1]$ , plotting the Nishimori differences  $R_{SE} - R_{MSE}$  in the upper panel and  $\tilde{O}_{t=0} - \tilde{M}O_{t=0}$  in the lower panel, both computed with a *rnd* initialization. The averages are done over 50 instances. Instead, in the right panels we fix  $N = 10^4$  and  $\rho = 0.2$  and we track BP through its iterations until convergence for six randomly generated instances of the problem, for each starting both with *rnd* and *inf* initializations. Using as reference the observables defined in Eqs. (15) and (9), we plot the difference between the two initializations, and we show that they lead, for each instance, to the same result. The scale of the y-axis of these plots is set as *symlog*, meaning it is logarithmic outside the interval  $[-10^{-3}, 10^{-3}]$  around zero, in which it is linear. The damping parameter  $\eta$  was fixed to 0.4 for all iterations (see Appendix A 2 a for details).

are shown in Fig. 3. The plots show the phase diagram of the problem for different systems settings. We consider both an SI model and a dSIR model with delay  $\Delta = 1$ , and, as stopping criteria of the simulation, when all the individuals are infected and when there are no more infectious individuals, respectively. Fixing  $\lambda = 0.4, 0.6, 0.8$ , we vary  $\delta$  and  $\rho$  inside an interesting interval.

In general, we can see that the system presents a transition between a phase (at low  $\delta$  and high  $\rho$ ) where the inference is successful, and we can find (most of) the sources, to another one (at high  $\delta$  and low  $\rho$ ) where the inference is unfeasible since we reach the performance of the random estimator. Varying  $\lambda$  and  $\Delta$  we observe several interesting phenomena:

(i) Decreasing the probability of transmission  $\lambda$ , starting from  $\lambda = 0.8$ , we see that the zone we infer as the random estimator remains more or less the same. At the same time, the zone of perfect inference shrinks. In particular, it moves more and more to the right, and we need more and more observations to achieve the same performance. Furthermore, it is interesting to notice that this effect is visible, especially at low values of the parameter  $\delta$ , where the level lines (in which the rescaled overlap has a fixed value) become more and more equally distributed, which in other terms tells us that the transition becomes less and less sharp. Instead, this effect disappears when the fraction of sources increases and the performance parameter becomes less dependent on the value of  $\lambda$ .

(ii) The inference on the sources depends just slightly on the parameter  $\Delta$ . We notice that the two cases (on the lower and upper panels of Fig. 3, respectively) are very similar, especially for high values of the parameter  $\delta$ .

(iii) There is a small region of parameters, at small values of  $\rho$  and  $\delta$ , where the BP equations struggle to converge, the Nishimori identities are no longer respected, and we observe a dependence of the results on the initialization of the BP messages. In this region, we can claim that the BP equations do not solve the problems in a Bayes-optimal way. In Sec. V, a detailed analysis of this and other regions where the BP algorithm poorly perform is shown.

### 2. Inference on the times of infection

Another property we are interested in characterizing is the ability to infer the single times of infection for each individual in the network. Having defined the rescaled squared error in Eq. (16), in Fig. 4 we show its behavior for the same set of parameters of previous results. As we can see from the figure, the behavior as a function of  $\lambda$  is qualitatively very similar to the one shown in Fig. 3 for the overlap, even if the two quantities describe two very different aspects of inference. In this case, the MSE is influenced much less than the overlap at time 0 by the same change in the infectivity parameter  $\lambda$ . Furthermore, in contrast to the previous plot, now we see that the parameter  $\Delta$  plays a major role in determining the value of the performance parameter: comparing the upper and lower panels of Fig. 4, we see that, again especially for low values of  $\delta$ , is much easier (in terms of needed observations) to have a high value of the performance parameter in the case of the SI model compared to the dSIR model with  $\Delta = 1$ .

The great variability of the  $R_{SE}$  between the case of dSIR with  $\Delta = 1$  and  $\Delta = \infty$  (the SI model) shown in Fig. 4 leads us to study in detail how the performances change, varying

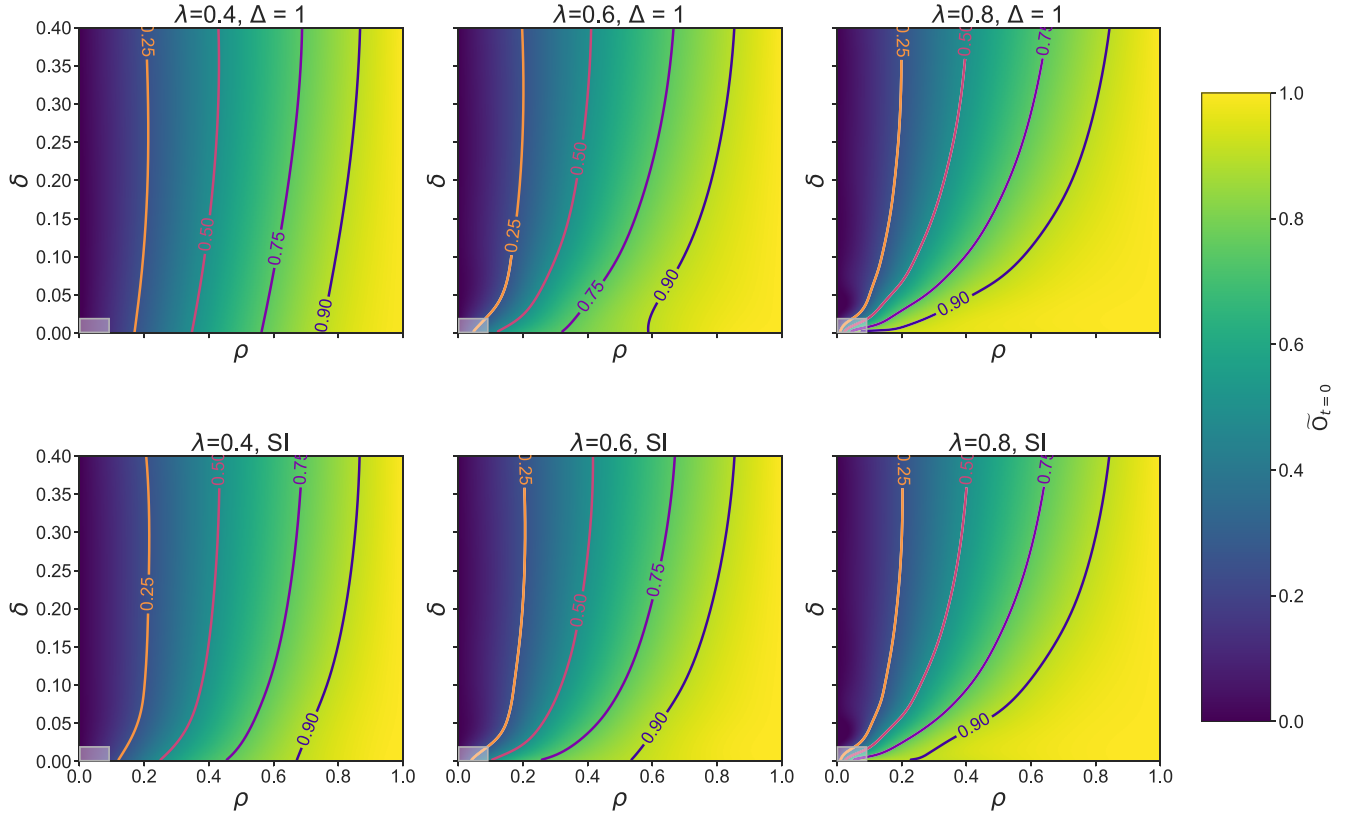


FIG. 3. Phase diagrams for source-finding via sensors. We show the rescaled overlap defined in Eq. (9) for 3-RRGs with  $N = 10^4$ , computed at time  $t = 0$ . We juxtapose the SI model to the dSIR model with  $\Delta = 1$ , and for both, we compare three different values of  $\lambda$ , and we show how the performance parameter varies with the fraction of sensors  $\rho \in [0, 1]$  and the fraction of sources  $\delta \in [2.5 \times 10^{-3}, 0.4]$ . The plots show the results for the case of random initialization since they are practically indistinguishable from the informed ones. For all the points outside the grey rectangles, the Nishimori conditions were verified, and thus also  $\text{MO}_{t=0}$  has the same behavior. Each contour plot was obtained by fitting the value of the rescaled overlap over a  $21 \times 24$  grid of data, averaging 25 simulations for each point on the grid. In the regions at small values of  $\delta$  and  $\rho$  enclosed in the grey rectangles, we observe that the BP equations no longer converge, the Nishimori identities do not hold, and the results are no longer Bayes optimal. More insights about these regions are covered in Sec. V.

the recovery time  $\Delta$ . The results are shown in Fig. 10 in the Appendix B.

### B. Phase diagrams for observations via a snapshot

In this section, we present the phase diagrams of the rescaled overlap and squared error for snapshot inference problems. In this case, the state (susceptible or infected) of all individuals at a given time of observation ( $T_{\text{obs}}$ ) is probed. To link with previous literature using snapshot observations (see, for instance, Refs. [13,26]), in the following, we consider tests that are not able to distinguish between infectious ( $I$ ) and recovered individuals ( $R$ ), but just to signal that the time of infection happened before the time of observation. In practice, this observation excludes all values of  $t_i \geq T_{\text{obs}}$ . Of course, considering tests capable of distinguishing between the  $I$  and the  $R$  states could improve the algorithm's performance on both tasks we considered in our work.

Furthermore, a second choice we made for the analysis was to focus on the case of “backward inference,” neglecting the study on the ability of our algorithm to infer what happens to the system after we take the snapshot. In practice, this is done

simply by fixing the final time  $T$  (as defined in Appendix A) to be equal to  $T_{\text{obs}}$ .

#### 1. Mean overlap at time zero

Let us start, as before, with the task of inferring the sources of epidemics, considering again for simplicity an ensemble of 3-RRGs and comparing performances on the SI model to the ones on the dSIR model with delay  $\Delta = 1$ . Fixing three different values of  $\lambda$ , we vary the fraction of sources  $\delta$  and the time of observation  $T_{\text{obs}}$  inside an interesting interval. The results for the rescaled mean overlap, computed using the BP algorithm and expressed by Eq. (9), are presented in Fig. 5. Upon comparison with the results for sensors shown in Fig. 3, it is evident that, in general, the inference is much more challenging in the snapshot case. This is because snapshots provide less information about the dynamics than the sensor case, which reveals a finite fraction of nodes with precisely known infection times.

The results in Fig. 5 show that increasing the observation time  $T_{\text{obs}}$  significantly reduces the algorithm's ability to retrieve sources. Moreover, it is possible to achieve a



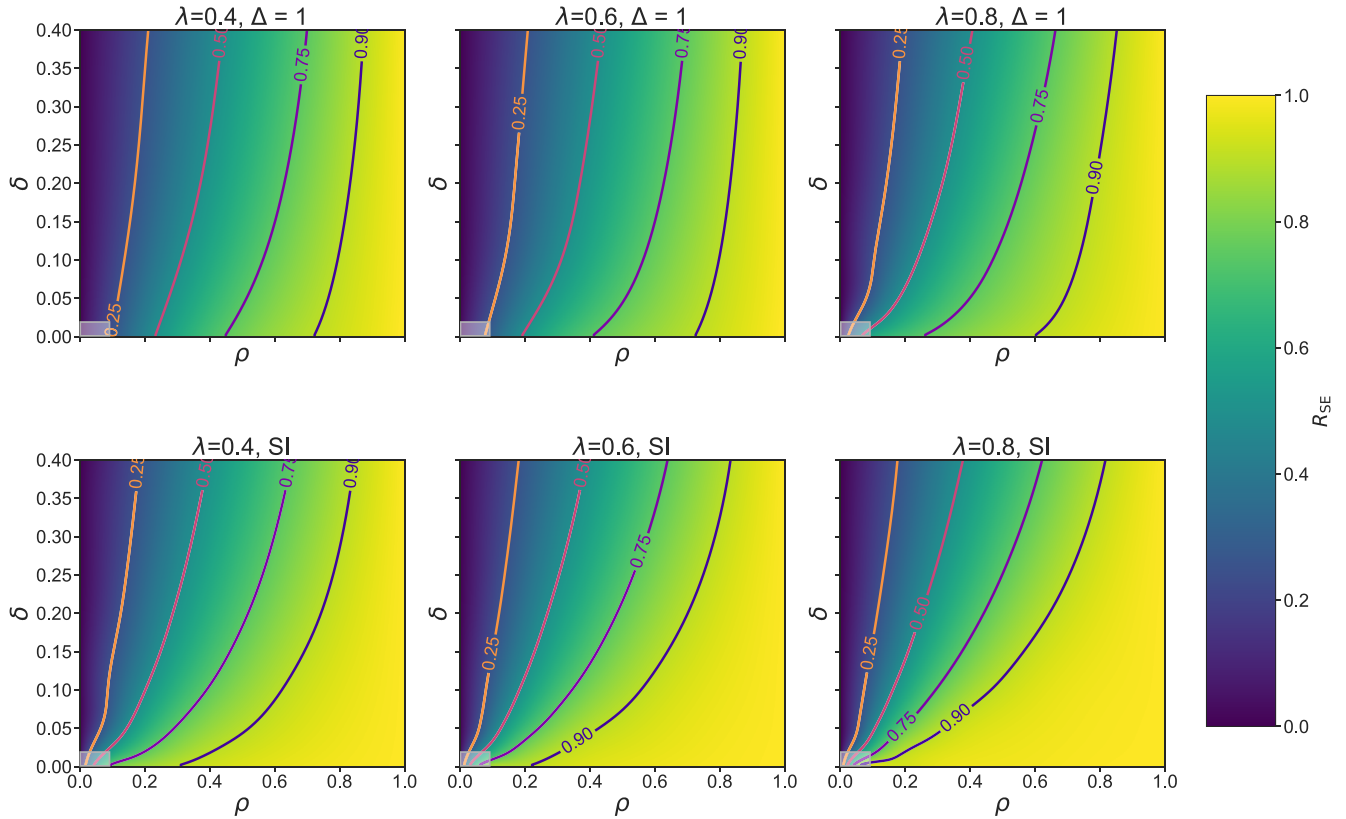


FIG. 4. Phase diagrams for inferring infection times via sensors. We show the rescaled squared error defined in Eq. (15) for 3-RRGs with  $N = 10^4$ . We juxtapose the SI model to the dSIR model with  $\Delta = 1$ . For both, we compare three different values of  $\lambda$ , showing how the performance parameter varies with the fraction of sensors  $\rho \in [0, 1]$  and the fraction of sources  $\delta \in [2.5 \times 10^{-3}, 0.4]$ . As described in Fig. 3, we show the results of random initialization, and the grey region of each panel corresponds to the violation of the Nishimori condition and the failure, in that region, of the BP equations to obtain Bayes-optimal results. The plots were obtained by fitting the value of the rescaled SE over a  $21 \times 24$  data grid, averaging over 25 simulations for each point on the grid.

better-than-random performance only when the observation time is sufficiently small and  $\delta$  is low.

As for the spreading parameter  $\lambda$  and the recovery delay  $\Delta$ , the qualitative behavior is similar to that observed in the sensor case. Specifically, the inference of patient zero depends only slightly on  $\Delta$ , whereas the infectivity parameter has a more significant impact. For instance, at  $\lambda = 0.4$ , even at small values of  $T_{\text{obs}}$  and  $\delta$ , the task is extremely challenging.

Similar to the case of observation via sensors, we identified a region in which BP has difficulty converging, and the Nishimori conditions are not met. This region is marked in a lighter color. Once again, this occurs for small values of  $\delta$  and coincides with the transition between as-random and better-than-random performance. In this case, the transition occurs when  $T_{\text{obs}}$  approaches the time  $T$  at which the epidemic “stops” (see Sec. V for more information).

### 2. Mean-squared error on the times of infection

In the following, we analyze the performances to infer the entire time trajectory of the individuals, measured in terms of the rescaled mean-squared error  $R_{\text{MSE}}$ , Eq. (15). The results are presented in Fig. 6, where we consider both the SI model and the dSIR model with  $\Delta = 1$ .

Interestingly, we achieve much better performance in inferring the time trajectory of the nodes than those obtained for the source-finding task, with the same values of epidemic parameters. Specifically, we can always find a region in which  $R_{\text{SE}} > 0.9$  for  $\lambda \geq 0.4$ .

As one would expect, for the SI model, we see that, for any arbitrary  $\delta$ , increasing  $T_{\text{obs}}$  implies a decrease in performance. This is explained by the fact that a snapshot done later in time brings very little information on the epidemic compared to snapshots performed earlier in time. The limit case is doing the snapshot after the epidemic has stopped when all nodes are infected, and we cannot do better than the random estimator, resulting in  $R_{\text{MSE}} = 0$ .

The dSIR model displays a different behavior, as is evident by looking at the upper panels of Fig. 6. At fixed arbitrary  $\delta$  and increasing  $T_{\text{obs}}$ , we observe, in the beginning, a decrease in performance (as in the SI model). But after reaching a minimum, the  $R_{\text{MSE}}$  increases. This is because for the dSIR model, at the end of the epidemic, a fraction of the nodes are recovered ( $R$ ), and the rest are still susceptible ( $S$ ). This implies that by doing a snapshot after the epidemic is finished, we still retain the information on who is still susceptible. This explains why the value of  $R_{\text{MSE}}$  does not go to zero as  $T_{\text{obs}}$  increases. To explain the nonmonotonicity, we must look at the definition of the  $R_{\text{MSE}}$  in Eq. (16), where the MSE is

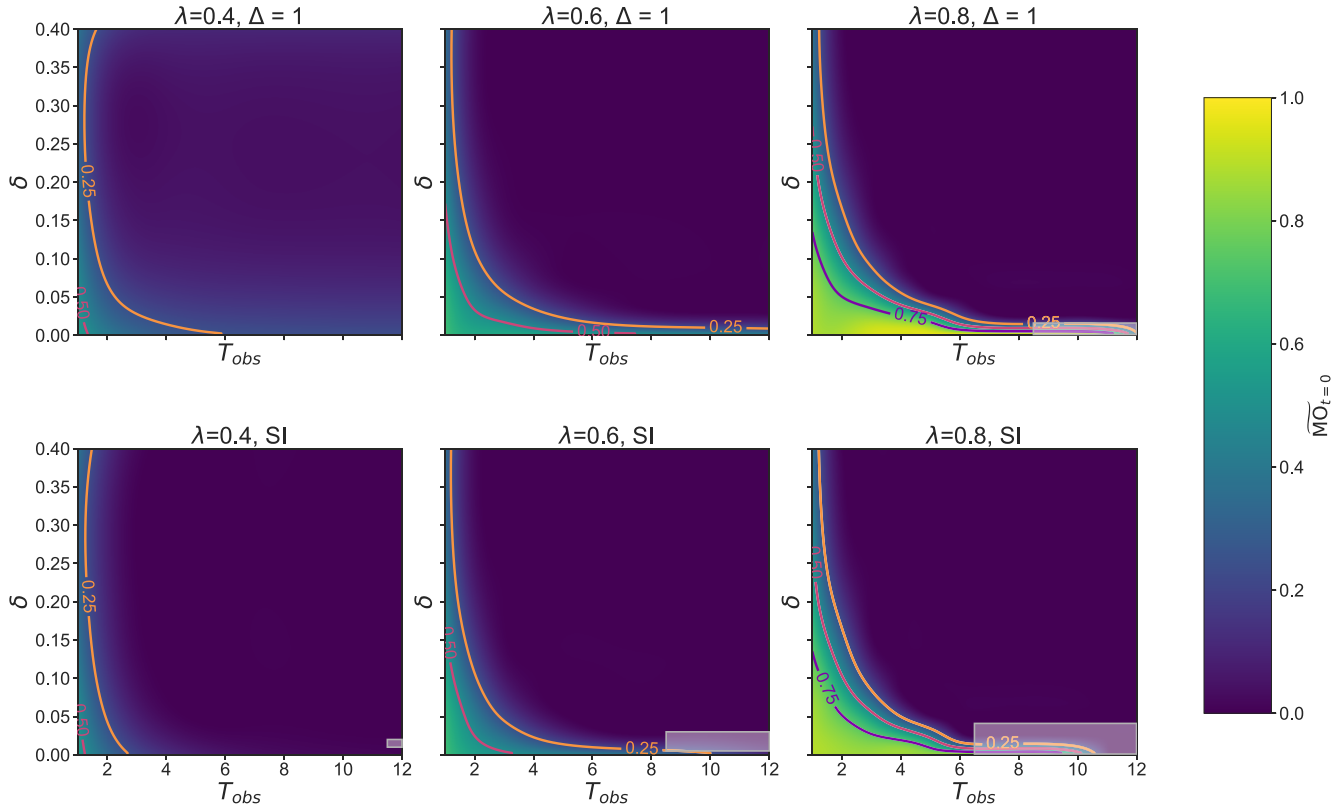


FIG. 5. Phase diagrams for source-finding in the snapshot framework. We show the rescaled mean overlap defined in Eq. (10) for 3-RRGs with  $N = 10^4$ , computed at time  $t = 0$ . We juxtapose the SI model to the dSIR model with  $\Delta = 1$  and for both, we compare three different values of  $\lambda$  and we show how the performance parameter varies with the time of observation  $T_{\text{obs}} \in [1, 12]$  and the fraction of sources  $\delta \in [2.5 \times 10^{-3}, 0.4]$ . As for previous phase diagrams, we show just the case of random initialization. For all the points outside the grey rectangles, the Nishimori conditions were verified, and thus also  $\hat{O}_{t=0}$  has the same behavior. Each contour plot was obtained by fitting the value of the rescaled mean overlap over a  $12 \times 20$  grid of data, doing 25 simulations for each point on the grid.

rescaled with the MSE( $\hat{\mathbf{t}}^{\text{RND}}$ ), the random estimator. This last observable gets worse as we increase  $T_{\text{obs}}$  because the possible times of infections increase as well. The result is the observed nonmonotonicity of the  $R_{\text{MSE}}$  as a function of the time of observations  $T_{\text{obs}}$ . As an illustration of this phenomenon, in Fig. 11 in the Appendix B we fix  $\lambda = 0.6$ , and some values of  $\delta$  for which this behavior is clearly visible by looking at SE and SE( $\hat{\mathbf{t}}^{\text{RND}}$ ) for the case  $\Delta = 1$  and  $\Delta = \infty$ .

Similar arguments allow us to explain the impact of the infectivity parameter  $\lambda$  on the performance of BP. We can notice that decreasing  $\lambda$  increases the number of susceptible individuals at the end of the simulation. In turn, their times of infection will be inferred perfectly thanks to the information obtained from the snapshot.

## V. REGIONS WITH FAILURE OF BAYES OPTIMALITY

The previous section presented the performance of the BP algorithm in two different settings of inference. We showed large regions of parameters where the BP equations converge and satisfy the Nishimori conditions. We thus conjecture that in these regions, the solutions found by the BP equations are very close to the Bayes-optimal performance.

It should be noted that observing such behavior at relatively small sizes is rather surprising. The mathematical arguments

about the exactness of BP rely on the fact that loops in the graph are longer than the correlation length. But since the length of the loops only grows logarithmically with size, this condition seems unreasonable to graphs of size as we treat in our experiments. Yet, we observe that BP converges and the Nishimori conditions are satisfied. Similar surprisingly small finite-size effects have been observed in many previous works using BP. Sometimes already at the size of several thousand, we see behavior that is very close to the predicted thermodynamic limit. It is not clear why this is so.

As pointed out in Sec. IV, we observed narrow regions of parameters (marked in clear color in the phase diagrams) where the BP equations do not converge, and the Nishimori conditions are no longer satisfied. For the two observation scenarios considered, the sensor and snapshot ones, we observe that the regions where BP fails are always at small values of  $\delta$  (the fractions of sources) and during the transition from regions where inference is impossible, i.e., the optimal solutions found by BP equals those of the random estimator, and regions where a close to perfect recovery of the missing information is achieved. As an example, we will focus on the case of snapshot observations, even though a similar behavior can be observed for sensors.

We investigate these regions closely and conclude that what is observed is not a sign of a critical region (where one

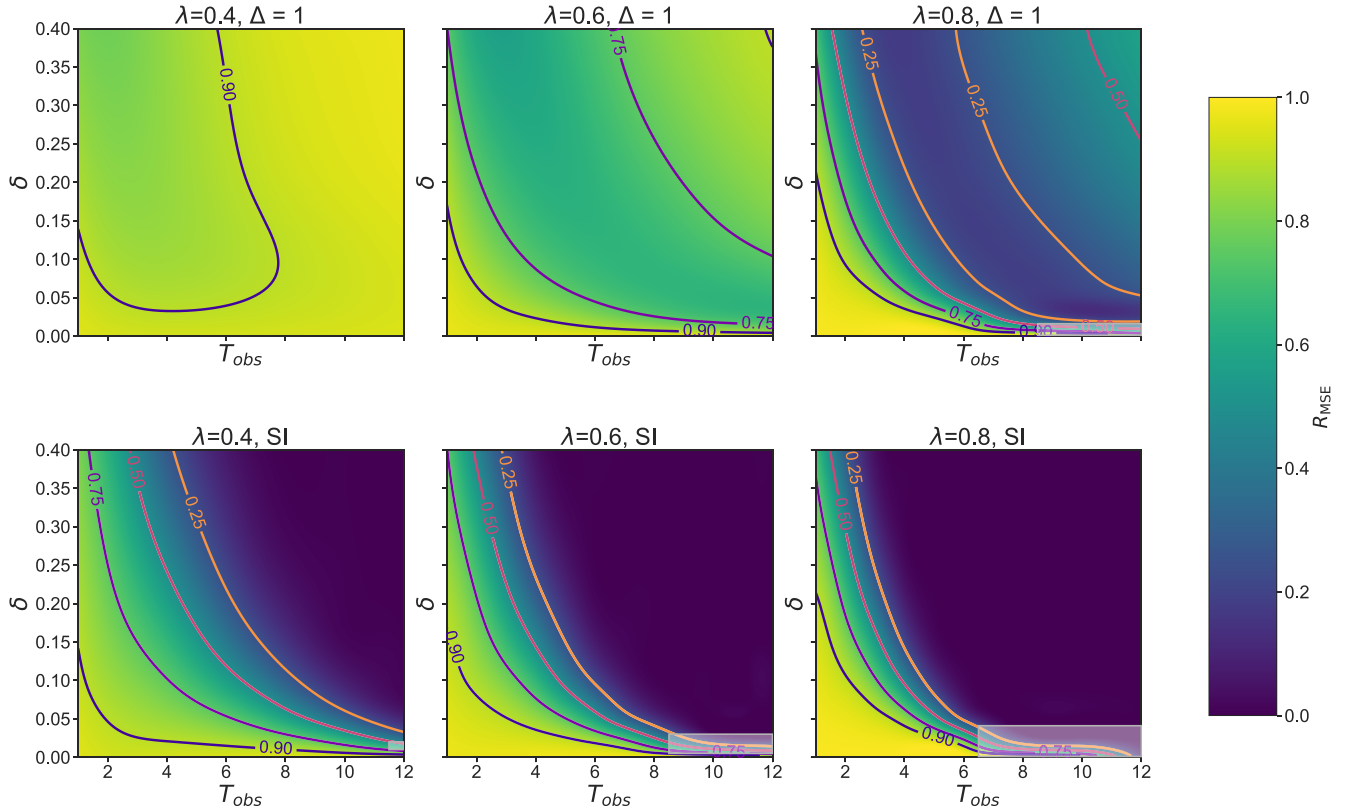


FIG. 6. Phase diagrams for the inference of infection times in the snapshot framework. We show the rescaled mean-squared error defined in Eq. (16) for 3-RRGs with  $N = 10^4$ . We juxtapose the SI model to the dSIR model with  $\Delta = 1$  and for both, we compare three different values of  $\lambda$  and we show how the performance parameter varies with the time of observation  $T_{\text{obs}} \in [1, 12]$  and the fraction of sources  $\delta \in [2.5 \times 10^{-3}, 0.4]$ . As for previous phase diagrams, we show just the case of random initialization. For all the points outside the grey rectangles, the Nishimori conditions were verified, and thus also  $R_{\text{SE}}$  has the same behavior. Each contour plot was obtained by fitting the value of the rescaled MSE over a  $12 \times 20$  grid of data, doing 25 simulations for each point on the grid.

would get a diverging length scale) and, thus, must be a finite-size effect. However, for the sizes we can simulate, this trouble does not go away. The existence of this region should perhaps not be surprising; what should be surprising is the region presented previously, where even for very moderate sizes, the thermodynamic limit is already effectively reached. We did not identify anything crisp, but we do hope that pointing to the existence of a region where BP should be asymptotically optimal but for the considered rather large sizes it is not will help to shed some light on the fact that BP is often able to perform well and converge even when the loops are still very short.

In Fig. 7, we investigate the inference from a snapshot at a certain time  $T_{\text{obs}}$ , and then we infer the infection times before  $T_{\text{obs}}$ , disregarding the subsequent epidemic evolution. We consider different values of  $\lambda \in [0.5, 1]$  and we rescale the snapshot time as

$$T_{\text{obs}}(\lambda) = \frac{T_{\text{obs}}(1)}{\lambda}, \tag{28}$$

to ensure that we observe (on average) the same “epidemic time,” i.e., keeping roughly constant the fraction of susceptible nodes at the time of observation. The left panel of Fig. 7 shows the values of the  $R_{\text{MSE}}$  varying  $\lambda$  and  $\delta$ . The central and right panels, where we investigate the Nishimori conditions

and the number of iterations needed for BP to converge, show that during the transition regions of the  $R_{\text{MSE}}$ , the algorithm struggles to converge, and the Nishimori conditions are no longer satisfied.

Figure 8 shows a finite-size effect study of the nonconvergence region, again in the snapshot case. The values of the observation time  $T_{\text{obs}}$ , the infectivity parameter  $\lambda$ , and the fraction of sources  $\delta$  are chosen to observe the transition where close to perfect inference is possible to regions where we obtain the same results of a random estimator. First, we observe that the values of the observable  $R_{\text{MSE}}$  (first row of Fig. 8) basically do not change with  $N$ , the number of nodes considered. The curve at  $N = 10^4$  and  $N = 10^5$  are almost indistinguishable, comforting the choice of  $N = 10^4$  as the primary size used to characterize the phase space of the inference problems analysed. The small values of  $\delta$  chosen allow us to observe, during the transition on the  $R_{\text{MSE}}$  values, that the BP equations stop to converge, the second row of Fig. 8. Observing the violation of the Nishimori conditions, the third row of Fig. 8, we can note that the regions where they are violated shrink very mildly. This fact confirms our hypothesis that the nonconvergence of the BP equations and the violation of the Nishimori conditions may be due to finite-size effects problems; nevertheless, the improvement of the Nishimori condition is very slow with  $N$ , making the BP equation not

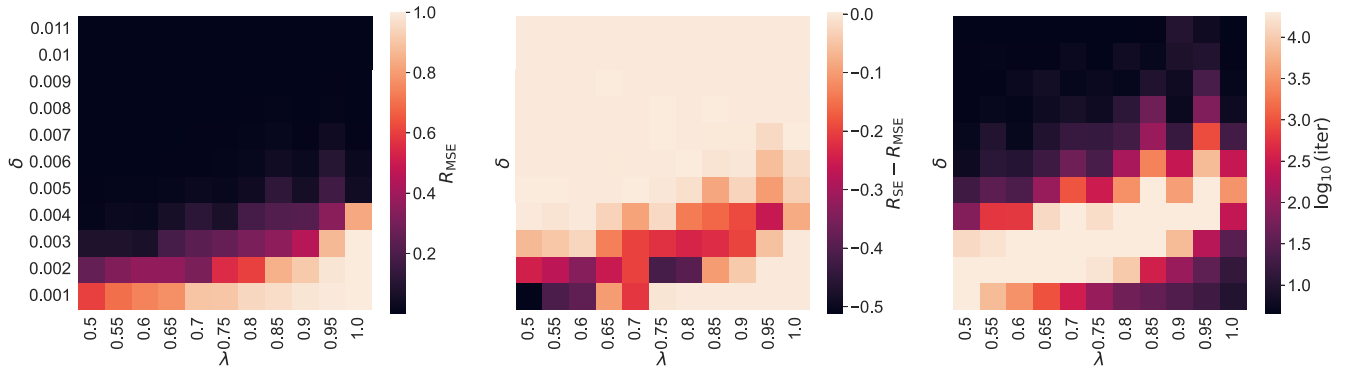


FIG. 7. Failure of the Nishimori conditions. Example of inference for snapshot observations in a region of parameters where Bayes optimality fails. The plots were generated using an ensemble of 3-regular random graphs (3-RRGs) with  $N = 10^4$  and varying values of  $\lambda$  and  $\delta$ . We consider an SI model where the snapshot time is fixed at  $T_{\text{obs}} = 9/\lambda$ . In the left panel, we show the performance of BP through the rescaled MSE defined in equation (16). The central panel probes the Nishimori conditions by studying how  $\delta R_{\text{SE}} \equiv R_{\text{SE}} - R_{\text{MSE}}$  behaves in the same range of parameters. The right panel shows the number of iterations for BP to converge. We set the maximum number of iterations to  $2 \times 10^4$ , so that the bright points indicate instances where BP did not converge. Each point was averaged over ten instances.

Bayes optimal in practical ranges of system sizes. For instance, this is the case of the very-well known patient-zero problem, where only one source infection is supposed to start the diffusion process.

We also investigate whether the observed issue cannot be a sign of proximity to a critical point. In Fig. 9, we consider again the SI model with snapshot observations, fixing  $\lambda = 0.5$  and  $\delta = 0.005$  and considering values of  $T_{\text{obs}}$  around the problematic region, which in this case is observed to be around  $T_{\text{obs}} = 14$ . In the presence of a critical point, one

should observe what in statistical physics is called “critical slowing down,” i.e., a power-law increase in convergence time when approaching the critical point. Since the algorithm does not converge in our case, we track the two initializations separately. We study the iteration time for both initializations to reach approximately the same value. Figure 9 shows the  $R_{\text{MSE}}$  values in the intervals from  $T_{\text{obs}} = 12$ , where we take the snapshot early enough to make a good inference, and  $T_{\text{obs}} = 17$ , where conversely we do it late enough to have no added information with respect to a random estimator. The

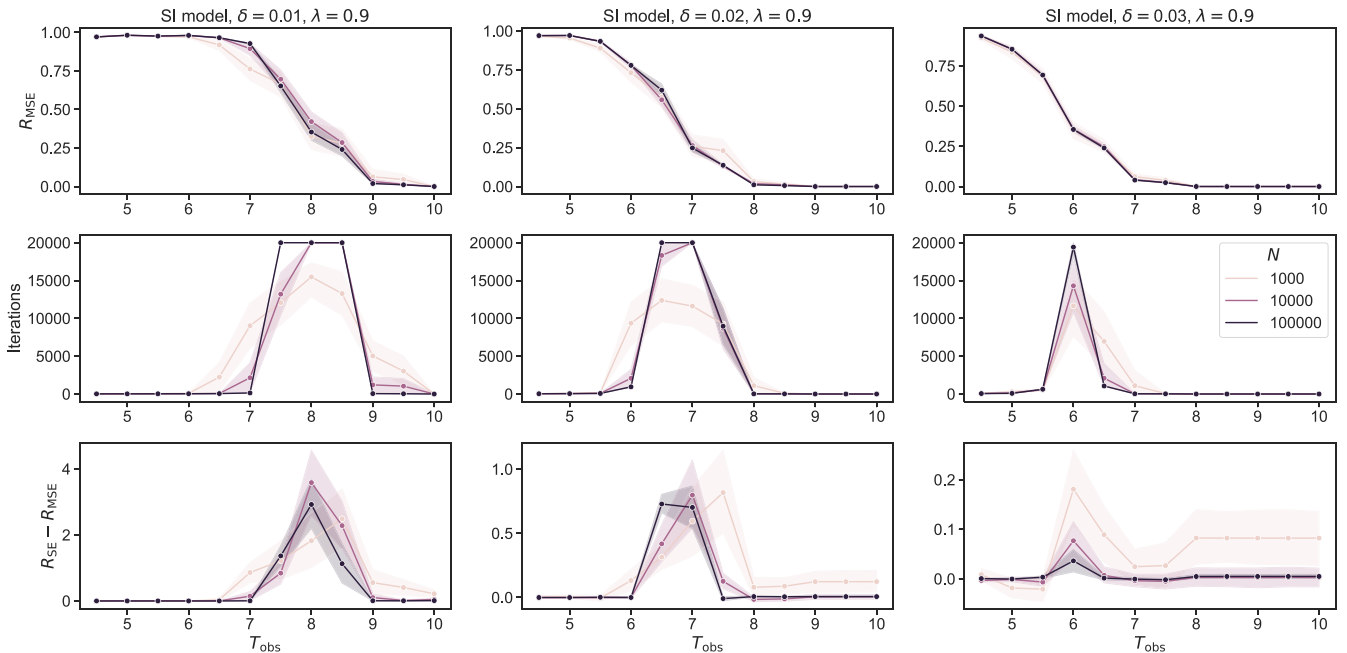


FIG. 8. Finite-size scaling. The plots were generated simulating an SI model on 3-regular random graphs (3-RRGs) of varying size, using snapshots at times  $T_{\text{obs}}$  as observations and starting BP with uninformed messages. The fraction of source nodes varies  $\delta = 0.01, 0.02, 0.03$ . The first row shows the values of  $R_{\text{MSE}}$  computed with the BP equations and averaged over 20 instances. The plots in the second row display the number of iterations to reach convergence, with a maximum of allowed iterations of  $2 \times 10^4$ . The plots in the third row show the difference  $R_{\text{SE}} - R_{\text{MSE}}$ , which certifies the violation of the Nishimori conditions when different from zero. We observe that the region of violation of the Nishimori condition may shrink very slowly with the system size.



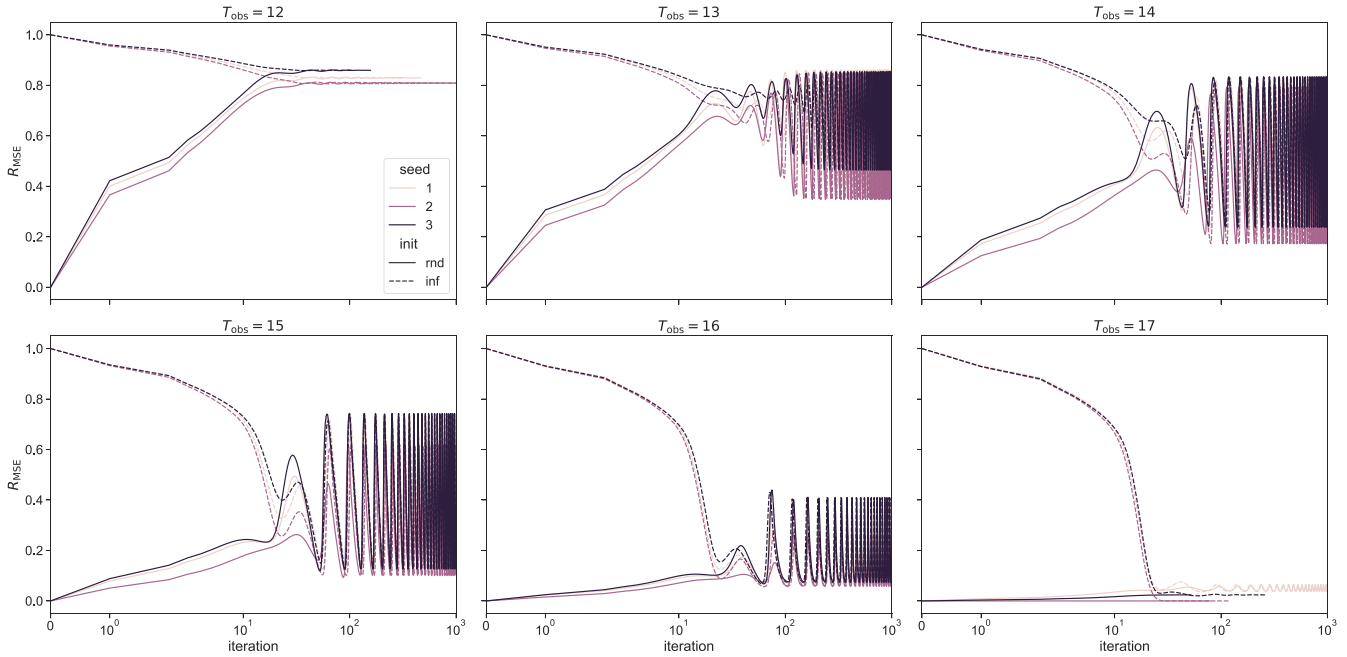


FIG. 9. Absence of critical slowing down. The plots were generated using an ensemble of 3-regular random graphs (3-RRGs) with  $N = 10^4$ ,  $\lambda = 0.5$ , and  $\delta = 0.005$ . We consider an SI model where the snapshot time is changed in each panel between  $T_{\text{obs}} = 12$  (top left) and  $T_{\text{obs}} = 17$  (bottom right). For every panel, we examine three instances of the problem generated randomly, and we monitor the rescaled mean-squared error (16) during the BP iterations. We compare the results of the same realizations when we initialize the messages, whether informed or not. The damping parameter in the BP algorithm was fixed to  $\eta = 0.4$ .

time needed for the two initialization to have similar values does not present a significant change. The same behavior can be observed by looking directly at the difference between BP messages, as shown in Fig. 12 in Appendix B.

Finally, after reaching the point where the two initializations predict almost the same observable value, both oscillate. The effect is present even with random updates of the BP equations and several different values of the damping parameter; see Appendix A for details about the damping parameter definition.

VI. CONCLUSIONS

In this paper, we present a detailed analysis of the performance of the belief propagation algorithm for spreading models defined on random networks. Compared to previous work on this topic, we have focused here on investigating whether such an algorithm is Bayes optimal, and on the possible presence of phase transitions in the relative phase diagrams. This is done by analyzing the convergence of BP from both the random and informed initial conditions, looking for first-order phase transitions. We find no such phase transitions. However, we analyze consistency conditions known as Nishimori conditions in the domain of disordered systems, which allowed us to heuristically investigate Bayes optimality in our simulations. When these conditions are satisfied, we conjecture that belief propagation asymptotically achieves the Bayes-optimal performance. We have provided phase diagrams both for the classical task of finding the sources of an epidemic (given different types of information) and for characterizing the entire trajectory of the nodes' state.

Furthermore, we dedicate the last section to describe our efforts in understanding the algorithm's behavior in the particular regime when the failure of the Nishimori conditions and the lack of convergence tells us we are not Bayes optimal. Although not reaching a firm conclusion, we feel confident in ruling out the hypothesis of the presence of some kind of phase transition, and we attribute this kind of behavior to finite-size effects leading to spurious correlations in the graphs that would disappear very slowly (e.g., logarithmically) with the system size.

ACKNOWLEDGMENT

We thank Patrick Thiran, Florent Krzakala, Alejandro Lage Castellanos, and Laurent Massoulié for useful discussions.

APPENDIX A: BPEPI

1. BP implementation for the dSIR model

The dynamical process of a generic dSIR model can be described using a single time of infection  $\{t_i\}_{i=1}^N$  for each variable, where  $t_i \in \{-1, 0, 1, \dots, T - 1, T\}$  such that:

- (i)  $t = -1$  is the (fictitious) time in which the sources get infected,
- (ii)  $t = 0$  is the first "simulation time," i.e., the first time in which the rest of the individuals can get the infection from the sources,
- (iii)  $t = T - 1$  is the last simulation time, i.e., the last time an individual can get the infection from a neighbor,
- (iv)  $t = T$  is the (fictitious) time in which, conventionally, all the individuals that have not been infected until (and at)

time  $t = T - 1$ , get infected. Notice that these nodes are still susceptible at time  $T$ , thus this does not affect observables computed at  $t = T$ .

It is not hard to see that the model just defined can be described with the formalism introduced in Sec. II. We define the probability of the susceptible node  $i$  not to be infected

at time  $s$  as  $p_{SS}(s, \{t_k\}_{k \in \partial_i}) = \prod_{k \in \partial_i} [1 - \lambda_{ki}(s - t_k)] \mathbb{I}(s > t_k)$ , and in the following, we will focus on the case in which the probability to be a source of the epidemic is uniform among the individuals, such that  $P(t_i = -1) = \delta \forall i$ . The factor of the posterior probability entering in the BP equations (22) can be written as

$$\begin{aligned} \tilde{\Psi}_i(t_i, \{t_k\}_{k \in \partial_i} | \boldsymbol{\lambda}) = & \left( (1 - \delta) \prod_{s=0}^{t_i-1} p_{SS}(s, \{t_k\}_{k \in \partial_i}) [1 - p_{SS}(t_i, \{t_k\}_{k \in \partial_i})] \mathbb{I}[-1 < t_i < T] \right. \\ & \left. + (1 - \delta) \prod_{s=0}^{T-1} p_{SS}(s, \{t_k\}_{k \in \partial_i}) \mathbb{I}[t_i = T] + \delta \mathbb{I}[t_i = -1] \right) P(\mathcal{O}_i | t_i). \end{aligned} \quad (\text{A1})$$

The first term in the square brackets represents the probability of the node  $i$  to be infected at time  $t_i$  by one or more of its neighbors, while the second term is the probability to remain susceptible during the whole epidemic process, and the third one is the probability to be a source of the epidemic. Inserting the above factor in Eq. (22) we observe easily that the sum over the times of infections of neighbors can be factorized, and we now show that the time complexity of the algorithm is reduced to  $\mathcal{O}(E * d_{\max} T^2)$ , thus no more exponential in the maximum degree of the interaction network and only quadratic in the maximum time  $T$ .

In this Appendix, we go into the details of the belief propagation equations, explaining how the code [21] is implemented for the deterministic-SIR model.

Let us start by considering again the factor contribution (A1) and noticing that we can rewrite the product

$$\prod_{s=0}^{t_i-1} p_{SS}(s, \{t_k\}_{k \in \partial_i}) [1 - p_{SS}(t_i, \{t_k\}_{k \in \partial_i})]$$

as

$$\prod_{k \in \partial_i} \prod_{s=0}^{t_i-1} [1 - \lambda_{ki}(s - t_k)]^{\theta(s-t_k)} - \prod_{k \in \partial_i} \prod_{s=0}^{t_i} [1 - \lambda_{ki}(s - t_k)]^{\theta(s-t_k)}. \quad (\text{A2})$$

Then, putting it back into Eq. (22), we get, apart from normalization,

$$\begin{aligned} m_{i \rightarrow j}(t_i, t_j) \propto & (1 - \delta) P(\mathcal{O}_i | t_i) \mathbb{I}[-1 < t_i < T] \\ & \times \left[ \left( \prod_{s=0}^{t_i-1} [1 - \lambda_{ji}(s - t_j)]^{\theta(s-t_j)} \right) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \left[ \left( \prod_{s=0}^{t_i-1} [1 - \lambda_{ki}(s - t_k)]^{\theta(s-t_k)} \right) m_{k \rightarrow i}(t_k, t_i) \right] \right. \\ & \left. - \left( \prod_{s=0}^{t_i} [1 - \lambda_{ji}(s - t_j)]^{\theta(s-t_j)} \right) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \left[ \left( \prod_{s=0}^{t_i} [1 - \lambda_{ki}(s - t_k)]^{\theta(s-t_k)} \right) m_{k \rightarrow i}(t_k, t_i) \right] \right] \\ & + (1 - \delta) \mathbb{I}[t_i = T] P(\mathcal{O}_i | t_i = T) \left( \prod_{s=0}^{T-1} [1 - \lambda_{ji}(s - t_j)]^{\theta(s-t_j)} \right) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \left( \prod_{s=0}^{T-1} [1 - \lambda_{ki}(s - t_k)]^{\theta(s-t_k)} \right) \\ & \times m_{k \rightarrow i}(t_k, T) + \delta \mathbb{I}[t_i = -1] P(\mathcal{O}_i | t_i = -1) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T m_{k \rightarrow i}(t_k, t_i = -1). \end{aligned} \quad (\text{A3})$$

## 2. Practical implementation

We define the following matrices for each directed edge  $i, j$ :

$$\begin{aligned} \Lambda_{ji}^1(t_j, t_i) & \equiv \prod_{s=0}^{t_i-1} [1 - \lambda_{ji}(s - t_j)]^{\theta(s-t_j)}, \\ \Lambda_{ji}^0(t_j, t_i) & \equiv \prod_{s=0}^{t_i} [1 - \lambda_{ji}(s - t_j)]^{\theta(s-t_j)}, \end{aligned}$$

where one can easily see that  $\Lambda_{ij}^0(t_i, t_j) = \Lambda_{ij}^1(t_i, t_j)[1 - \lambda_{ji}(t_i - t_j)]^{\theta(t_i - t_j)}$ . The messages read

$$\begin{aligned} m_{i \rightarrow j}(t_i, t_j) \propto & (1 - \delta)P(\mathcal{O}_i|t_i)\mathbb{I}[-1 < t_i < T] \times \left( \Lambda_{ji}^1(t_j, t_i) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \Lambda_{ki}^1(t_k, t_i) m_{k \rightarrow i}(t_k, t_i) \right. \\ & \left. - \Lambda_{ji}^0(t_j, t_i) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \Lambda_{ki}^0(t_k, t_i) m_{k \rightarrow i}(t_k, t_i) \right) + (1 - \delta)\mathbb{I}[t_i = T]P(\mathcal{O}_i|t_i = T)\Lambda_{ji}^1(t_j, T) \\ & \times \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \Lambda_{ki}^1(t_k, T) m_{k \rightarrow i}(t_k, T) + \delta\mathbb{I}[t_i = -1]P(\mathcal{O}_i|t_i = -1) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T m_{k \rightarrow i}(t_k, t_i = -1), \end{aligned} \quad (\text{A4})$$

which can also be rewritten as

$$\begin{aligned} m_{i \rightarrow j}(t_i, t_j) \propto & (1 - \delta)\mathbb{I}[-1 < t_i < T]P(\mathcal{O}_i|t_i) \times \left( \Lambda_{ji}^1(t_j, t_i) \prod_{k \in \partial_i \setminus j} \gamma_{ki}^1(t_i) - \Lambda_{ji}^0(t_j, t_i) \prod_{k \in \partial_i \setminus j} \gamma_{ki}^0(t_i) \right) + (1 - \delta)\mathbb{I}[t_i = T] \\ & \times P(\mathcal{O}_i|t_i = T)\Lambda_{ji}^1(t_j, T) \prod_{k \in \partial_i \setminus j} \gamma_{ki}^1(T) + \delta\mathbb{I}[t_i = -1]P(\mathcal{O}_i|t_i = -1) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T m_{k \rightarrow i}(t_k, t_i = -1), \end{aligned} \quad (\text{A5})$$

where we have defined the quantities:

$$\begin{aligned} \gamma_{ki}^1(t_i) & \equiv \sum_{t_k=-1}^T \Lambda_{k,i}^1(t_k, t_i) m_{k \rightarrow i}(t_k, t_i), \\ \gamma_{ki}^0(t_i) & \equiv \sum_{t_k=-1}^T \Lambda_{k,i}^0(t_k, t_i) m_{k \rightarrow i}(t_k, t_i). \end{aligned}$$

The marginals for the infection times  $t_i$  can be obtained, after fixing as  $j$  one neighbor of  $i$ , through the following equation:

$$b_i(t_i) \propto \sum_{t_j=-1}^T m_{i \rightarrow j}(t_i, t_j) m_{j \rightarrow i}(t_j, t_i). \quad (\text{A6})$$

The belief propagation algorithm is implemented in Python and can be found in Ref. [21]. Additionally, in Ref. [27], there is the code to reproduce every figure in the paper.

### a. Damping

To improve the convergence of the belief propagation algorithm, as the one following from Eq. (A4), a damping updating scheme is typically used. In practical terms, the new messages are a linear combination of the old and new ones. In mathematical terms, writing as  $m^n$  the message at iteration  $n$ , we use

$$m^{n+1} = \eta m^n + (1 - \eta) f_{\text{BP}}(m^n), \quad (\text{A7})$$

where we write as  $f_{\text{BP}}(m^n)$  the right-hand side of Eq. (A4), and the parameter  $\eta$  is introduced to control the intensity of the damping.

If not specified otherwise, in the simulations in the main text, then the damping was set to 0 for the first 200 iterations, to 0.2 for the next 200, and then 0.4 until the end.

### b. Notes about implementation

The messages can be stored in a NumPy tensor of shape  $2|E| \times (T+2) \times (T+2)$ , as well as  $\Lambda^0$  and  $\Lambda^1$  that can be

precomputed from the beginning. These data are stored in a Python class called *FactorGraph*. We implement a class called *SparseTensor* that represents the tensor of shape  $2|E| \times T \times T$ , and includes a transparent mapping between the pair of nodes  $(i, j)$  and the corresponding labeled edge.

*Efficient implementation of Eq. (A5).* For each node  $i$  we update, at the same time, all messages that come out. We start by computing the following quantities:

- (i)  $\gamma_{ki}^1(t_i)$  and  $\gamma_{ki}^0(t_i)$  for all  $k \in \partial_i$ ,
- (ii)  $\gamma_i^1(t_i) = \prod_{k \in \partial_i} \gamma_{ki}^1(t_i)$  and  $\gamma_i^0(t_i) = \prod_{k \in \partial_i} \gamma_{ki}^0(t_i)$ .

Calling  $\widehat{M}_{ij}$  the matrix containing the message  $m_{i \rightarrow j}(t_i, t_j)$  for each  $t_i$  and  $t_j$ , the update is described by the following matrix equation:

$$\begin{aligned} \widehat{M}_{ij}[1 : T+1, :] &= (1 - \delta) \underline{O}_i \cdot [\widehat{\Lambda}_{ji}^1[:, 1 : T+1]] \cdot \underline{\gamma}_i^1 \\ &\quad \cdot (\underline{\gamma}_{ji}^1)^{-1} - \widehat{\Lambda}_{ji}^0[:, 1 : T+1] \cdot \underline{\gamma}_i^0 \\ &\quad \cdot (\underline{\gamma}_{ji}^0)^{-1}], \\ \widehat{M}_{ij}[T+1, :] &= (1 - \delta) \underline{O}_i(T) \widehat{\Lambda}_{ji}^1[:, T+1] \cdot \frac{\gamma_i^1(T)}{\gamma_{ji}^1(T)}, \\ \widehat{M}_{ij}[0, :] &= \delta \underline{O}_i(-1) \prod_{k \in \partial_i \setminus j} \sum_{t_k=-1}^T \widehat{M}_{ki}[t_k, 0]. \end{aligned}$$

The  $\cdot$  are element-wise products, which are handled using the indexing rule of NumPy [28].  $\underline{O}_i$ ,  $\gamma_i^{1/0}$ , and  $\gamma_{ij}^{1/0}$  are  $(T+2)$ -dimensional vectors.  $\widehat{M}_{ij}$ ,  $\widehat{\Lambda}_{ij}^1$ , and  $\widehat{\Lambda}_{ij}^0$  are  $(T+2) \times (T+2)$  matrices, where the first entry is  $t_i$  and the second one is  $t_j$ .

## APPENDIX B: ADDITIONAL IMAGES

### 1. Inference of the final time

We can consider the case of what happens when we try to infer the state of the system at final time  $T$ , which is trivial for

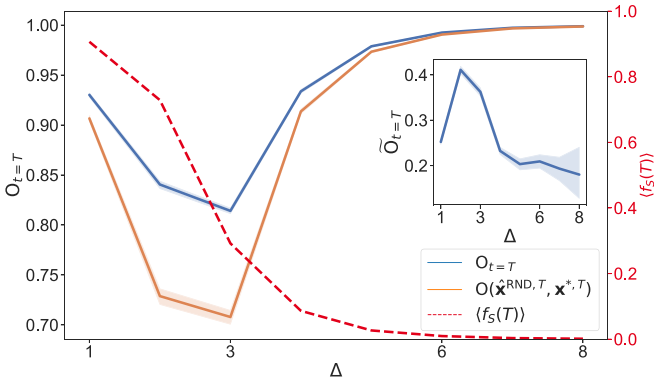


FIG. 10. Inference of the state at the final time as a function of  $\Delta$ . The plots were generated using an ensemble of 3-regular random graphs (3-RRGs), fixing  $\lambda = 0.25$  and  $\delta = 0.04$  for the spreading. Furthermore, we consider sensor observations, with a fraction of observed nodes  $\rho = 0.2$ . We compare the overlap (5) at  $t = T$  computed with  $\hat{\mathbf{x}}^{\text{MMO},T}$  and with  $\hat{\mathbf{x}}^{\text{RND},T}$  varying  $\Delta$  in a dSIR model. The dashed red line represents the average fraction of susceptible nodes at  $t = T$ . The inset plot represents the behavior of the rescaled overlap defined in Eq. (9) for the same range of parameters. We verified that the Nishimori conditions were satisfied for all the points displayed in the plots and that the informed initialization led to the same result as the uninformed one.

the SI model since all nodes are infected, but become more interesting for the dSIR model, where a fraction of nodes remains susceptible at the end of the epidemic. We consider the dSIR model with delay parameter  $\Delta$ .

In Fig. 10 the ability of the algorithm to infer the state of the system at the final time  $T$  is shown. The average value of the fraction of susceptible nodes at the end of the epidemic is represented as the dashed red line in the figure, and we see that, as expected, as we increase  $\Delta$  we go back to the SI model, where we perform similar to the random estimator. If, instead, we take a lower value of  $\Delta$ , then we can have less trivial situations, with the hardest one being the case in which the fraction of susceptible nodes  $f_S(T) \approx 1/2$ , in which the random estimator performs worse. Nevertheless, it is interesting to notice that this is also the regime where the algorithm performs best, as can be seen from the inset plot in Fig. 10, in which the behavior of the rescaled overlap with  $\Delta$  is displayed.

## 2. Rescaled squared error for snapshot observations in the dSIR model

In this section, we present an example to explain the behavior described in Sec. IV B 2 where the results of the rescaled squared error for snapshot observations were presented. In particular, concerning Fig. 6, we additionally fix  $\lambda = 0.6$ , but the same behavior is visible for every value of the infectivity parameter. We then compare what happens for the SI model ( $\Delta = \infty$ ) and for the dSIR model with  $\Delta = 1$ , looking separately at the SE computed with the optimal estimator and with the random estimator. Let us start by looking at the SI model case, displayed in the right panels of Fig. 11. We can see that both observables, when increasing  $T_{\text{obs}}$ , reach some plateau at some value which depends on the fraction of sources  $\delta$ . Since the value of the plateau is the same for both estimators,

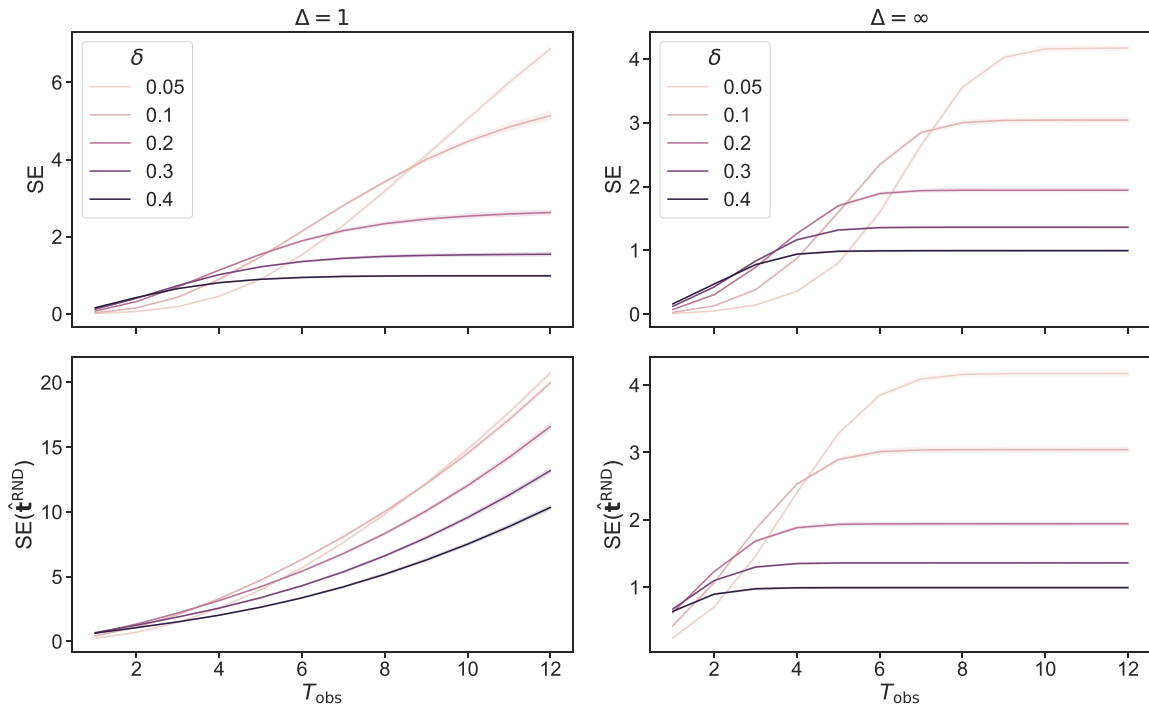


FIG. 11. Compendium of Fig. 6. Remaining in the same framework as Fig. 6, we fix  $\lambda = 0.6$  and study both the SI model and the dSIR model with delay  $\Delta = 1$ . For both scenarios, in the upper panel, we plot the squared error (11) as a function of the snapshot time, and in the lower panel we do the same for the squared error computed with the random estimator defined in Sec. II.



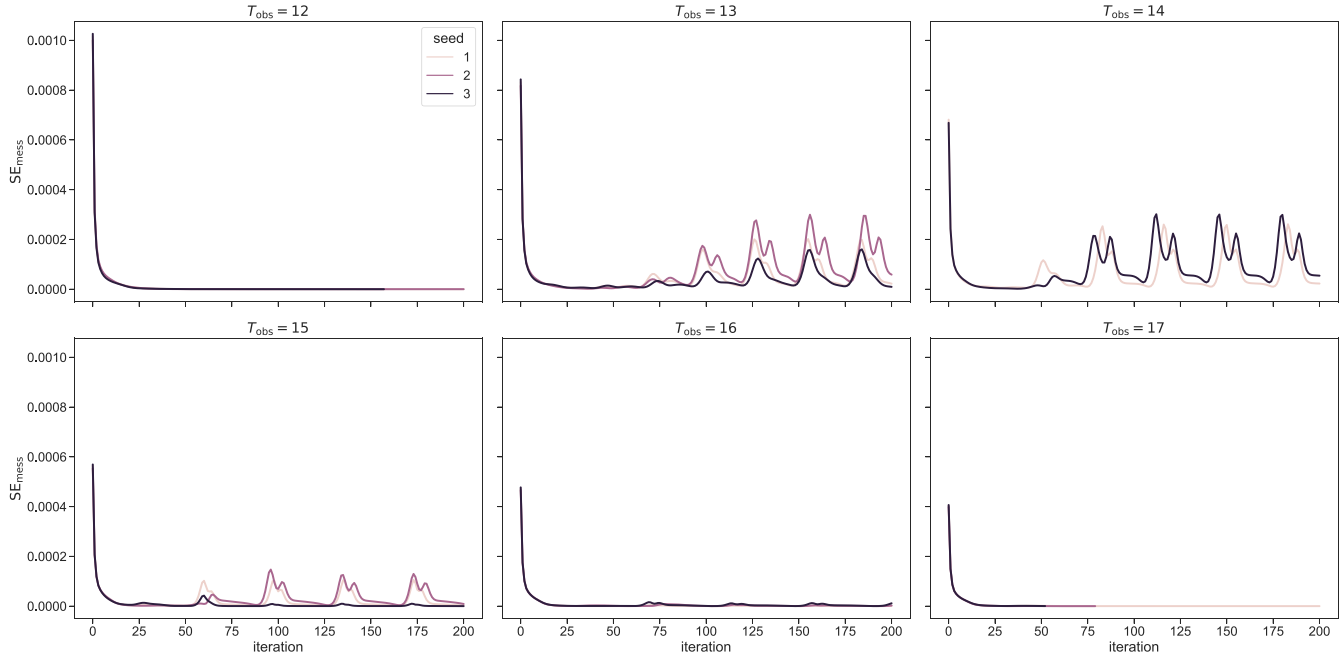


FIG. 12. Absence of critical slowing down in the convergence of the BP messages. The plots were generated using an ensemble of 3-regular random graphs (3-RRGs) with  $N = 10^4$ ,  $\lambda = 0.5$  and  $\delta = 0.005$ . We consider observations via a snapshot on an SI model, where the snapshot time is changed in each panel between  $T_{\text{obs}} = 12$  (top left), and  $T_{\text{obs}} = 17$  (bottom right). For each panel, we examine three instances of the problem generated randomly, and we monitor the squared error defined in Eq. (B1) during the BP iterations. The damping parameter  $\eta$  was fixed to 0.4 for all iterations. We observe that the BP messages of two initializations quickly converge to almost the same values in every case, without a critical slowing down in the nonconvergence region, where the oscillations are observed.

it means that for high values of  $T_{\text{obs}}$  the rescaled squared error defined in Eq. (15) will go to zero eventually.

Conversely, for the dSIR model, we see that the same behavior is valid for the SE computed with the optimal estimator but not for the one using the random estimator. Indeed, as discussed in the main text, the latter estimator performs quadratically worse when increasing  $T_{\text{obs}}$  in the case of a fraction of nodes remaining susceptible at the end of the epidemic. This, in turn, makes the rescaled squared error (15) grow when increasing  $T_{\text{obs}}$ , even if the SE is constant, and finally leads to the nonmonotonic behaviours showed in Fig. 6.

### 3. Squared error between BP messages

In this section, we show how the absence of the critical slowing down on the convergence of the BP equations that we have presented in Fig. 9 can be also seen directly by looking at the BP messages. Specifically, we generate an instance of the problem, i.e., we fix the graph, the sources, the epidemic process, and the observations, and then we run the BP equations with *rnd* and *inf* initialization of the messages. Then, at each iteration time, we compare the two sets of messages by computing the following observable:

$$SE_{\text{mess}} = \frac{1}{2|E|(T+2)^2} \sum_{\{i \rightarrow j\}} \sum_{t_i=-1}^T \sum_{t_j=-1}^T [m_{i \rightarrow j}^{\text{inf}}(t_i, t_j) - m_{i \rightarrow j}^{\text{rnd}}(t_i, t_j)]^2, \tag{B1}$$

where the first sum is done over all the ordered pairs in the graph. The results for the setting of Fig. 9 are presented in

Fig. 12. One can clearly see that the behavior for the first few dozen iterations is the same, with the squared error going quickly to zero, whatever the value of  $T_{\text{obs}}$ . Then, we see that while for high and low values of  $T_{\text{obs}}$  the squared error remains low and the algorithm converges, for intermediate values at some point it becomes bigger again, and we can see the typical oscillations that characterize the non-Bayes-optimal regime.

### 4. Comparison with other graphs ensembles

In this section, we alter the underlying graph structure through which our process propagates to check the results' robustness. We explore two variations with respect to the framework previously examined in the main text, i.e., random regular graphs with a degree of  $d = 3$ . The first variation considers the Erdos-Renyi graph ensemble [29], where the nodes are randomly connected, with probability  $p$ .

Second, we modify the average node degree  $d$ . In the context of Erdos-Renyi graphs, we set  $p$  to  $\frac{d}{N-1}$  to match the same average degree as that of random regular graphs. This ensures that, in both scenarios, the average number of edges in the graph will be the same,  $dN/2$  on average.

The Erdos-Renyi graphs, with  $Np > 1$ , will almost surely have a unique giant component containing a finite fraction of the vertices, and no other component will contain more than  $\mathcal{O}(\log N)$  vertices [29]. In our framework, being interested in how a process spreads on the graph, we neglect all the disconnected components (in which the spreading is trivial) and consider only the graph's giant component. The results of this study are presented in Fig. 13, in which we fix a particular

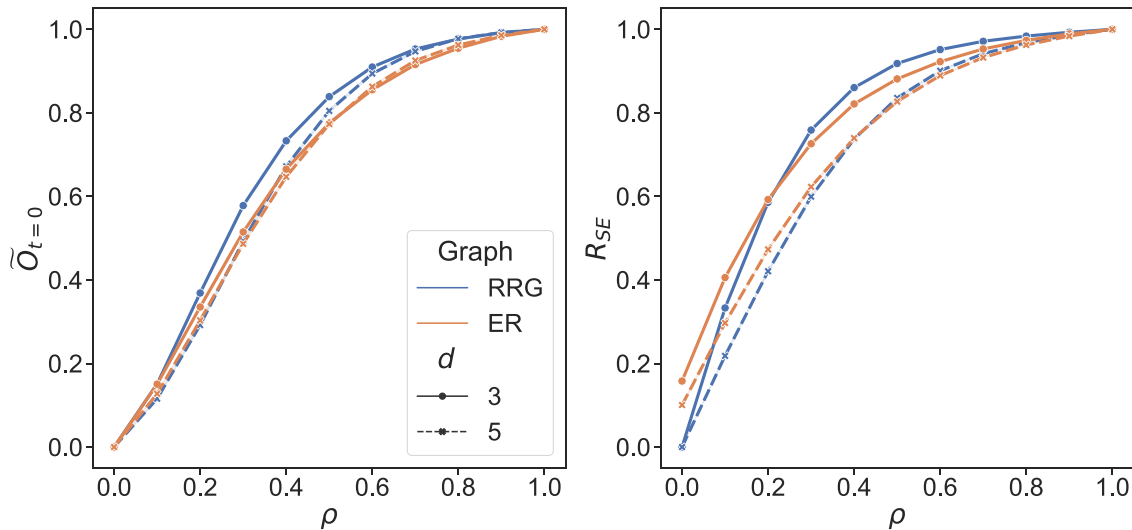


FIG. 13. Comparison between Erdos-Renyi graphs and Random Regular graphs. The plots were generated using both an ensemble of regular random graphs (RRG) and random Erdos-Renyi graphs (ER). For both, we consider two values of average degree  $d = 3, 5$  (as explained in the text). We furthermore fix  $\lambda = 0.8$  and  $\delta = 0.1$  for the spreading in such a way to be always in the Bayes-optimal setting. Moreover, we consider sensor observations with a fraction of observed nodes  $\rho$  varying between 0 and 1. We look at how the overlap (9) at  $t = 0$  and the  $R_{SE}$  defined in Eq. (15) behave when we change the type of graph ensemble considered. In general, we see that the performance parameters depend only slightly on the choice of the ensemble, except the  $R_{SE}$ , which seems always to decrease when considering higher average degrees  $d$ , for both RRGs and ER graphs.

value of the parameters  $\lambda$  and  $\delta$  in such a way to always be in the Bayes-optimal region, and we vary the fraction of observed nodes  $\rho$  in a sensors framework. At fixed  $d$ , we observe slight changes in the algorithm's performances on

RRGs or ER graphs in the inference problems considered. Instead, the change of the average degree  $d$  seems to affect both graph ensembles, generally decreasing the algorithm's performance when one increases  $d$ .

- 
- [1] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, *Rev. Mod. Phys.* **87**, 925 (2015).
- [2] M. Henkel, H. Hinrichsen, S. Lübeck, and M. Pleimling, *Non-equilibrium Phase Transitions* (Springer, Berlin, 2008), Vol. 1.
- [3] A. Vespignani, Modelling dynamical processes in complex socio-technical systems, *Nat. Phys.* **8**, 32 (2012).
- [4] S. Shelke and V. Attar, Source detection of rumor in social network—A review, *Online Soc. Netw. Media* **9**, 30 (2019).
- [5] E. Ortega, A. Braunstein, and A. Lage-Castellanos, Contamination source detection in water distribution networks using belief propagation, *Stochastic Environ. Res. Risk Assess.* **34**, 493 (2020).
- [6] A. Baker, I. Biazzo, A. Braunstein, G. Catania, L. Dall'Asta, A. Ingrassio, F. Krzakala, F. Mazza, M. Mézard, A. P. Muntoni, M. Refinetti, S. S. Mannelli, and L. Zdeborová, Epidemic mitigation by statistical inference from contact tracing data, *Proc. Natl. Acad. Sci. USA* **118**, 32 (2021).
- [7] N. M. Ferguson, D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke, Strategies for mitigating an influenza pandemic, *Nature (London)* **442**, 448 (2006).
- [8] E. Brooks-Pollock, G. O. Roberts, and M. J. Keeling, A dynamic model of bovine tuberculosis spread and control in Great Britain, *Nature (London)* **511**, 228 (2014).
- [9] G. Fanti and P. Viswanath, Deanonymization in the bitcoin p2p network, *Adv. Neural Info. Process. Syst.* **30**, 1 (2017).
- [10] D. Shah and T. Zaman, Rumors in a network: Who's the culprit? *IEEE Trans. Inf. Theory* **57**, 5163 (2011).
- [11] P. C. Pinto, P. Thiran, and M. Vetterli, Locating the source of diffusion in large-scale networks, *Phys. Rev. Lett.* **109**, 068702 (2012).
- [12] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Ingrassio, and R. Zecchina, The patient-zero problem with noisy observations, *J. Stat. Mech.* (2014) P10016.
- [13] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina, Bayesian inference of epidemics on networks via belief propagation, *Phys. Rev. Lett.* **112**, 118701 (2014).
- [14] A. Braunstein, A. Ingrassio, and A. P. Muntoni, Network reconstruction from infection cascades, *J. R. Soc. Interface.* **16**, 20180844 (2019).
- [15] I. Biazzo, A. Braunstein, L. Dall'Asta, and F. Mazza, A Bayesian generative neural network framework for epidemic inference problems, *Sci. Rep.* **12**, 19673 (2022).
- [16] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, UK, 2009).
- [17] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, *Phys. Rev. E* **84**, 066106 (2011).
- [18] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová, Information-theoretic thresholds from the cavity method, in

- Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* (ACM Press, New York, NY, 2017), pp. 146–157.
- [19] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, Oxford, UK, 2001).
- [20] F. Krzakala and L. Zdeborová, Statistical physics of inference: Thresholds and algorithms, *Adv. Phys.* **65**, 453 (2016).
- [21] A. Aragon, I. Biazzo, D. Ghio, and L. Zdeborová, BPEPI: Belief propagation implementation for deterministic sir models on random graphs (2022), <https://github.com/SPOC-group/bpepi>.
- [22] J. S. Yedidia, W. T. Freeman, and Y. Weiss, Understanding belief propagation and its generalizations, *Explor. Artif. Intel. New Millen.* **8**, 0018 (2003).
- [23] B. Bollobás, Random graphs, in *Modern Graph Theory* (Springer, Berlin, 1998), pp. 215–252.
- [24] B. Spinelli, L. E. Celis, and P. Thiran, A general framework for sensor placement in source localization, *IEEE Trans. Netw. Sci. Eng.* **6**, 86 (2019).
- [25] A. Y. Lokhov, M. Mézard, H. Ohta, and L. Zdeborová, Inferring the origin of an epidemic with a dynamic message-passing algorithm, *Phys. Rev. E* **90**, 012801 (2014).
- [26] K. Zhu, Z. Chen, and L. Ying, Catch'em all: Locating multiple diffusion sources in networks with partial observations, in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (AAAI Press, Washington, 2017), Vol. 31.
- [27] A. Aragon, I. Biazzo, D. Ghio, and L. Zdeborová, Repository for reproducing the figures in the paper (2022), <https://github.com/IdePHICS/BPEPl-results>.
- [28] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Array programming with NumPy, *Nature (London)* **585**, 357 (2020).
- [29] P. Erdős and A. Rényi, On random graphs I, *Public. Math. Debre.* **6**, 290 (1959).