Epidemic inference through generative neural networks

Indaco Biazzo (✉ indaco.biazzo@polito.it)

Politecnico di Torino <https://orcid.org/0000-0002-9897-7543>

Alfredo Braunstein

Politecnico di Torino <https://orcid.org/0000-0002-4562-3610>

Luca Dall'Asta

Politecnico di Torino

Fabio Mazza

Politecnico di Torino

Article

Keywords:

Posted Date: February 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1200309/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Epidemic inference through generative neural networks

Biazzo Indaco^{1*}, Braunstein Alfredo^{1,2,3}, Luca Dall'Asta^{1,2,3}
and Mazza Fabio¹

¹Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10125, Italy.

²IIGM - Italian Institute for Genomic Medicine, c/o IRCSS, Candiolo, Turin, 10060, Italy.

³Collegio Carlo Alberto, P.za Arbarello 8, Turin, 10122, Italy.

*Corresponding author(s). E-mail(s): indaco.biazzo@polito.it;

Abstract

Reconstructing missing information in epidemic spreading on contact networks can be essential in prevention and containment strategies. For instance, identifying and warning infective but asymptomatic individuals (e.g., manual contact tracing) helped contain outbreaks in the COVID-19 pandemic. The number of possible epidemic cascades typically grows exponentially with the number of individuals involved. The challenge posed by inference problems in the epidemics processes originates from the difficulty of identifying the almost negligible subset of those compatible with the evidence (for instance, medical tests). Here we present a new generative neural networks framework that can sample the most probable infection cascades compatible with observations. Moreover, the framework can infer the parameters governing the spreading of infections. The proposed method obtains better or comparable results with existing methods on the patient zero problem, risk assessment, and inference of infectious parameters in synthetic and real case scenarios like spreading infections in workplaces and hospitals.

1 Introduction

Discrete-state stochastic compartmental models have been traditionally used to model infectious diseases [1–3] and provide a simple and unified mathematical

framework for a wide variety of spreading processes occurring in social and technological systems [4]. The time-forward simulations of most epidemic models, even those incorporating detailed demographic and mobility data, can be efficiently performed using Monte-Carlo based sampling techniques or, at least at the meta-population level, exploiting approximation methods, such as stochastic differential equations and moment closure schemes [5]. These computational methods have been largely applied to large-scale epidemic forecasting [6–9] and containment [10–12]. Their effectiveness crucially depends on the capacity to exploit the available information on the past behavior of the epidemic outbreak. At the meta-population level, such information, represented by temporal series of aggregate quantities (e.g. daily number of newly infected individuals inside a reference population) can be rather easily included within traditional Bayesian computational frameworks based on Monte Carlo sampling techniques [13–15]. The COVID-19 pandemic has motivated the interest for the large-scale adoption of epidemic surveillance techniques and digital contact tracing through smartphone applications [16, 17], which could make it possible to access/use a large amount of (possibly inaccurate) individual-based observational data, traditionally available only for case studies in rather small and controlled environments [18–20]. The availability of individual-based observational data unveils a crucial limitation of traditional Monte-Carlo based inferential techniques. While the number of possible epidemic realizations generated by a specific epidemic model on a given contact network scales exponentially with the systems size and the duration of the process, those compatible with individual-based observations are just an exponentially small fraction of them [21]. It follows that inferential methods based on the direct sampling of epidemic realizations on individual-based contact networks rapidly become inefficient as the size of the outbreak increases [21]. As an example, let us consider several simulated epidemic realizations (with the same initial condition, consisting of a single infected individual, and the same epidemic parameters) in a real graphs of temporal contacts between patients and staff members of a hospital [22] (see Fig. 1).

We define the daily configuration of the system as the daily epidemic state of the individuals (infected/not-infected) during the epidemic process. The plots in Fig. 1 indicate a fast divergence of the configurations of the simulated epidemics, even though they start from the same individual with the same epidemic parameters. Choosing, for instance, the final configuration of one epidemic cascade as the individual-based observation of the system, it is very unlikely to obtain the same configuration from a direct sampling of the epidemic model.

A step forward in this field is represented by the introduction of efficient algorithms for Bayesian inference based on Belief Propagation (BP) [23, 24]. In the Bayesian inference framework, the objective is to approximately compute the posterior probability of the system, assuming the epidemic model as a prior and the individual-based observations as the evidence. BP-based algorithms make it possible to obtain estimates of the local marginals of the posterior distribution and, as shown in [23, 25, 26], this approach outperforms competing methods on sparse contact networks on a variety of inference problems. In particular, the integration of such

Fig. 1 Simulated epidemic cascades in a hospital contact network. One thousand epidemic cascades simulated (with the same epidemic parameters) on a real contact graph measured in a hospital[22] (detailed information about epidemic models and contact networks are listed in the results section). The epidemics started from the same individual. Two samples (blue, and orange) of epidemic cascades are shown in the first and second rows of the figure. The third row represents the distance between them, where in this case the blue dots are the infected individuals present in the cascade 1 but not in cascade 2 and the orange ones are those present in cascade 2 but not in cascade 1. In the third row, the total number of blue and orange dots gives the Hamming distance between the two daily configurations. **Left-bottom plot.** Cumulative number of infected individuals for 1000 epidemic cascades started from the same individual. **Right-bottom plot.** Hamming distance ($\delta(1, i)$) between the cascade 1 and all the others $i \in [2, 3 \dots 1000]$.

algorithms in the framework of digital contact tracing for COVID-19 was recently shown to provide a better assessment of the individual risk and improve the mitigation impact of non-pharmaceutical interventions strategies [27].

BP-based algorithms may experience non-convergence issues, for instance in dense and very structured contact networks, a phenomenon that calls for the search of alternative inference methods which could overcome such a limitation while maintaining comparable performances on sparse networks. Here we propose to use generative neural networks, specifically autoregressive neural networks (ANN), to learn the posterior probability of an epidemic process and efficiently sample from it. In practice, the autoregressive neural network can generate realizations of the epidemic process according to the stochastic dynamical rules of the prior model but compatible with the evidence. Deep autoregressive neural networks are used to generate samples according to a probability distribution learned from data, for instance for images [28], audio [29], text [30, 31] and protein sequences [32] generation tasks and, more generally, as a probability density estimator [33–35]. Autoregressive neural networks have recently been used to approximate the joint probability distributions of many (discrete) variables in statistical physics models [36], and applied in different physical contexts [37–40]. In this work, we show how to use a deep autoregressive neural network architecture to efficiently sample from a posterior distribution composed of a prior, given by the epidemic propagation model (even though

the parameters of such model can be contextually inferred), and from an evidence given by (time-scattered) observations of the state of a subset of individuals. Neural networks have already been applied to epidemic forecasting [41–43] but rarely to epidemic inference and reconstruction problems. Two recent preliminary works apply neural networks to epidemic inference problems: in [44] the patient zero problem is tackled using graph neural networks, while a similar technique is applied to epidemic risk assessment in [45]. The presented approach allows to address successfully a large class of epidemic inference problems, ranging from the patient-zero problem and individual risk assessment to the inference of the parameters of the propagation model under a unique neural network framework. We believe this to be a strong point in favour of this technique. In all such problems, the proposed autoregressive neural network architecture provides results that are at least as good as all other methods considered for comparison and outperforms them in most cases. The implementation of the algorithm and the instructions to reproduce the results are available at [46].

2 Methods

The posterior probability of the epidemic process

The dynamics of epidemic spreading in a contact network is commonly described by means of individual-based stochastic models in which individuals can be in a finite set of possible states, usually called epidemic compartments. For the sake of concreteness, consider the discrete-time SIR model, in which $x_i^t \in \mathcal{X} = \{S, I, R\}$ stands for the individual i being at time step t in the Susceptible (S), Infected (I) or Recovered (R) state. The infection of a susceptible individual due to a contact with an infected individual occurs with rate λ , while infected individuals recover in time with rate μ (heterogeneous epidemic parameters can be considered as well if necessary). In the epidemic propagation model, both the epidemic parameters and the temporal structure of the underlying contact network are assumed to be given and known, so that the individual transition probability of the corresponding Markov chain reads as follows

$$p(x_i^{t+1} = S | \mathbf{x}_{\partial i}^t, x_i^t) = \mathbb{1}[x_i^t = S] \prod_{j \in \partial i} (1 - \lambda \mathbb{1}[x_j^t = I]), \quad (1)$$

$$p(x_i^{t+1} = R | \mathbf{x}_{\partial i}^t, x_i^t) = \mathbb{1}[x_i^t = R] + \mathbb{1}[x_i^t = I] \mu, \quad (2)$$

and $p(x_i^{t+1} = I | \mathbf{x}_{\partial i}^t, x_i^t) = 1 - p(x_i^{t+1} = S | \mathbf{x}_{\partial i}^t, x_i^t) - p(x_i^{t+1} = R | \mathbf{x}_{\partial i}^t, x_i^t)$. Here $\mathbb{1}[x_i^t = X]$ is the indicator function that is equal to 1 when $x_i^t = X$ and zero otherwise, $\mathbf{x}_{\partial i}^t$ represents the set of individuals that are in contact with node i at time t . Defining the individual epidemic-state trajectory as $\underline{x}_i = \{x_i^0, \dots, x_i^T\}$, the probability of an epidemic realization $\underline{\mathbf{x}} = \{\underline{x}_1, \dots, \underline{x}_N\}$ between time 0 and time T is

$$p(\underline{\mathbf{x}}) = p(x_1^0, \dots, x_N^0) \prod_{i=1}^N \prod_{t=1}^T p(x_i^t | \mathbf{x}_{\partial i}^{t-1}, x_i^{t-1}), \quad (3)$$

where $p(x_1^0, \dots, x_N^0)$ is a prior distribution on the initial state, which is usually assumed to be factorized, i.e. $p(x_1^0, \dots, x_N^0) = \prod_i p(x_i^0)$.

Suppose that some information about the individual states at different times is available (e.g because individuals exhibit symptoms or undergo medical tests). We will assume that this information comes in the form of a set of independent observed variables O_r following known probabilistic laws $p_r(O_r|x_{i_r}^{t_r})$. For example, if an individual i_r has been observed in state X_r at time t_r , we will have $p_r(O_r|x_{i_r}^{t_r}) = \mathbb{1}[x_{i_r}^{t_r} = X_r]$. False negative and positive rates in tests can be easily represented generalizing the expression of p_r .

Under the assumption of independence between observations, the conditional probability of $\mathcal{O}$ given $\underline{x}$ is $p(\mathcal{O}|\underline{x}) = \prod_{r \in \mathcal{O}} p_r(O_r|x_{i_r}^{t_r})$, and the posterior probability of an epidemic cascade $\underline{x}$ given the observation $\mathcal{O}$ becomes

$$p(\underline{x}|\mathcal{O}) = \frac{1}{p(\mathcal{O})} p(\mathcal{O}|\underline{x}) p(\underline{x}) \quad (4)$$

$$= \frac{1}{Z} \prod_{i=1}^N p(x_i^0) \prod_{t=1}^T p(x_i^t | \mathbf{x}_{\partial i}^{t-1}, x_i^{t-1}) \prod_{r \in \mathcal{O}} p_r(O_r|x_{i_r}^{t_r}) \quad (5)$$

$$= \frac{1}{Z} \prod_i \Psi_i(x_i, \mathbf{x}_{\partial i}) \quad (6)$$

where $\Psi_i(x_i, \mathbf{x}_{\partial i}) = p(x_i^0) \prod_{t=1}^T p(x_i^t | \mathbf{x}_{\partial i}^{t-1}, x_i^{t-1}) \prod_{r: i_r=i} p_r(O_r|x_{i_r}^{t_r})$. Here, $\mathbf{x}_{\partial i}$ represents the set of individuals that have been in contact with i at least one time during the interval $[0, T]$. The quantity $Z = \sum_{\underline{x}} \prod_i \Psi_i(x_i, \mathbf{x}_{\partial i})$ is the normalization constant of the posterior probability or model evidence. Several quantities of interest can be computed from the posterior distribution. For instance, the problem of identifying the initial source of an outbreak in a population (the so-called patient-zero problem) requires to estimate the marginal probability $p(x_i^0 = I|\mathcal{O}) = \sum_{\underline{x}} \mathbb{1}(x_i^0 = I) p(\underline{x}|\mathcal{O})$ for every individual i . On the other hand, if the present time is T , the marginal probability $p(x_i^T = I|\mathcal{O})$ provides a measure of the current epidemic risk for every individual i . The exact computation of probability marginals requires a sum over all admissible realizations of the process, which becomes unfeasible for more than a few dozens of individuals because their number typically grows exponentially with the number of individuals. In the next subsection, we describe a method to approximately compute the marginals of the posterior distribution in Eq.6 using Autoregressive Neural Networks (ANNs).

Learning the posterior probability using autoregressive neural networks

Given a realization $\underline{x}$ of the epidemic process and a permutation $\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ of the individuals of the system, which imposes a specific ordering to the variables $\{x_{\pi_i}\}$, the probability of the realization $\underline{x}$ can be written as the product

of conditional probabilities (chain rule) in the form

$$p(\underline{\mathbf{x}}) = \prod_{i=1}^N p(\underline{x}_i | \underline{\mathbf{x}}_{<i}) \quad (7)$$

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

$$q^\theta(\underline{\mathbf{x}}) = \prod_i q_i^{\theta_i}(\underline{x}_i | \underline{\mathbf{x}}_{<i}), \quad (8)$$

which can be interpreted as a (possibly deep) autoregressive neural network depending on a set of parameters $\theta = \{\theta_i\}$. From the analytical expression of the probabilistic model $p(\underline{\mathbf{x}})$, and thus that of the posterior distribution $p(\underline{\mathbf{x}} | \mathcal{O})$ defined in Eq. 6, the operation of parameters learning can be performed using a variational approach proposed in Ref. [36], in which the (reversed) Kullback-Liebler (KL) divergence

$$D_{KL}(q^\theta || p) = \sum_{\underline{\mathbf{x}}} q^\theta(\underline{\mathbf{x}}) \ln \frac{q^\theta(\underline{\mathbf{x}})}{p(\underline{\mathbf{x}} | \mathcal{O})} \quad (9)$$

is minimized with respect to the parameters θ of the trial distribution $q^\theta(\underline{\mathbf{x}})$. The minimization of the KL divergence can be performed using standard gradient descent algorithms in which the gradient of the KL divergence with respect to the parameters of the trial distribution reads

$$\nabla_\theta D_{KL}(q || p) = \nabla_\theta \sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) [-\log p(\underline{\mathbf{x}} | \mathcal{O}) + \log q_\theta(\underline{\mathbf{x}})] \quad (10)$$

$$= \sum_{\underline{\mathbf{x}}} \nabla_\theta q_\theta(\underline{\mathbf{x}}) \left[-\log \left(\prod_i \Psi_i(\underline{x}_i, \underline{\mathbf{x}}_{\partial i}) \right) + \log q_\theta(\underline{\mathbf{x}}) \right] + \sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \nabla_\theta \log q_\theta(\underline{\mathbf{x}}) \quad (11)$$

$$= \sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \left[-\log \left(\prod_i \Psi_i(\underline{x}_i, \underline{\mathbf{x}}_{\partial i}) \right) + \log q_\theta(\underline{\mathbf{x}}) \right] \nabla_\theta \log q_\theta(\underline{\mathbf{x}}), \quad (12)$$

where we used that $\sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) \nabla_\theta \log q_\theta(\underline{\mathbf{x}}) = \sum_{\underline{\mathbf{x}}} \nabla_\theta q_\theta(\underline{\mathbf{x}}) = 0$ and $\log(Z) \sum_{\underline{\mathbf{x}}} \nabla_\theta q_\theta(\underline{\mathbf{x}}) = 0$ due to the normalization $\sum_{\underline{\mathbf{x}}} q_\theta(\underline{\mathbf{x}}) = 1$. The computational bottleneck of this calculation is that in the equations 9 and 12, the sum runs over all possible epidemic realizations, a set that grows exponentially with the size of the system. This issue is avoided by exploiting the generative power of autoregressive neural networks by training them using generated sample data through ancestral sampling. This means that the averages over the autoregressive probability distribution can be approximated as a sum over a large number of independent samples extracted from the autoregressive probability distribution q^θ , in which the conditional structure

of the autoregressive neural network allows to use the ancestral sampling procedure [47], see Fig.2.

Fig. 2 Ancestral sampling of epidemic cascades. **Left.** Ancestral sampling of epidemic cascades using artificial neural networks. For each individual i there is a neural network ANN_i that computes the probability $q(\underline{x}_i | \underline{x}_{i-1} \dots \underline{x}_1)$ of its time trajectory $\underline{x}_i$ given the time trajectory of previous individuals. The time trajectory $\underline{x}_i$ is extracted from the conditional probability $q(\underline{x}_i | \underline{x}_{i-1} \dots \underline{x}_1)$ and passed to the following neural networks. **Right.** Each neural network is composed of several fully connected layers (see SM for details).

A common way to represent the conditional probabilities in Eq.8 is by means of feed-forward deep neural networks with sharing schemes architectures [35, 48] to reduce the number of parameters. Due to the possible high variability in the dependence of $p(\underline{x}_i | \underline{x}_{<i})$ on $\underline{x}_{<i}$ [40], instead of adopting a sharing parameters scheme we reduce the number of parameters by limiting the dependency of the conditional probability to a subset of $\underline{x}_{<i}$. The subset considered is formed by all $x_j \in \underline{x}_{<i}$ such that x_j is at most a second-order neighbor of i in the graph induced by the contact network, i.e. the one in which there is an edge between two individuals if they had at least one contact during the epidemic process. The approximation employed turns out to be exact when the interaction graph is acyclic (see SM for a proof). The function to minimize, the Kullback-Leibler divergence in Eq.9, could attain large (or even infinite) negative values, causing convergence issues in the parameter learning process. As illustrated in the SI, this is avoided by introducing a regularization parameter, a fictitious temperature, and an annealing procedure to improve the convergence.

Inferring the parameters of the propagation model

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

$$p(\mathcal{O} | \Lambda) = \sum_{\underline{x}} p(\mathcal{O} | \underline{x}) p(\underline{x} | \Lambda) \quad (13)$$

$$= \sum_{\underline{\mathbf{x}}} \prod_i \Psi_i(\underline{x}_i) \quad (14)$$

$$= Z(\Lambda). \quad (15)$$

The quantity Z is the same normalization constant introduced in Eq.6, where the dependence on the parameters was dropped. Formally,

$$\Lambda^* = \arg \max_{\Lambda} Z(\Lambda) = \arg \max_{\Lambda} \log Z(\Lambda). \quad (16)$$

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

$$\log Z(\Lambda) = \sum_{\underline{\mathbf{x}}} p(\underline{\mathbf{x}}|\mathcal{O}) \log \prod_i \Psi_i(\underline{x}_i, \Lambda) - \sum_{\underline{\mathbf{x}}} p(\underline{\mathbf{x}}|\mathcal{O}) \log p(\underline{\mathbf{x}}|\mathcal{O}) \quad (17)$$

$$\geq \sum_{\underline{\mathbf{x}}} q^\theta(\underline{\mathbf{x}}) \log \prod_i \Psi_i(\underline{x}_i, \Lambda) - \sum_{\underline{\mathbf{x}}} q^\theta(\underline{\mathbf{x}}) \log q^\theta(\underline{\mathbf{x}}) \quad (18)$$

$$= -\langle H \rangle_q + S_q \quad (19)$$

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

$$\langle H \rangle_q = - \sum_{\underline{\mathbf{x}}} q^\theta(\underline{\mathbf{x}}) \log \prod_i \Psi_i(\underline{x}_i, \Lambda) \quad (20)$$

$$S_q = - \sum_{\underline{\mathbf{x}}} q^\theta(\underline{\mathbf{x}}) \log q^\theta(\underline{\mathbf{x}}). \quad (21)$$

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

1. Update the parameters $\{\theta_i\}$ of the autoregressive neural network to minimize the KL divergence in Eq.9.
2. Update the parameters Λ to maximize the quantity $\langle H \rangle_q$.

These steps are repeated until the end of the learning process.

Results

As a preliminary illustration of the ability of the proposed Autoregressive Neural Network (ANN) to sample epidemic realizations from a given posterior distribution,

we reconsider the example in Fig. 1, focusing on the blue epidemic cascade. We train the ANN to learn the posterior probability composed by the prior, i.e. the epidemic model that generates the blue cascade, and the evidence, i.e. its final configuration at day 12. The result is shown in Fig. 3. The epidemic cascades generated by the ANN have Hamming distances from the reference one that reduce to zero at day 12 (central-bottom plot) and a fraction of them have prior probabilities larger than the probability of the (blue) epidemic cascade taken as reference, right-bottom plot in Fig. 3. This example suggests that the ANN approach can generate epidemic cascades compatible with the observations and sampled according to the prior epidemic model.

Fig. 3 Epidemics cascades generated by the ANN. Epidemic cascades generated by the ANN trained on a posterior probability composed by a prior, the epidemic model that generate the blue cascade, and the evidence, its final configuration at day 12. The contact network is a real contact graph measured in a hospital[22]. An example of the epidemic cascade generated is shown in second row (1_{ANN} , orange). The third row represent the hamming distance between them (see caption Fig. 1.) **Left-bottom plot.** Cumulative number of infected individuals for epidemic cascade simulated (blue curve) and generated by the ANN ($i_{ANN} \in [1, 2 \dots 1000]$). **Central-bottom plot.** Hamming distance ($\delta(1, i_{ANN})$) between blue epidemic cascade and those generated by the ANN $i_{ANN} \in [2, 3 \dots 1000]$. **Right-bottom plot.** Distribution of the values of the prior probability of the generated epidemic cascades (P_{ANN}). The blue vertical line is the value of the prior probability of the blue cascade ($\log(P_1)$).

In the following, we exploit the ability of the ANN to sample epidemic cascades from a posterior distribution to tackle three challenging epidemic inference problems: *i*) the patient-zero problem, in which the unique source of a partially observed epidemic outbreak has to be identified, *ii*) the risk assessment problem, in which the epidemic risk of each individual has to be estimated from partial information during the evolution of the epidemic process, and *iii*) the inference of the epidemic parameters. Results are compared with those obtained using already existing methods in the field of epidemic inference. We also evaluate how the efficiency of the ANN algorithm depends on the size of the epidemic outbreak, measuring the number of

generated epidemic samples necessary to obtain nearly optimal results. The comparison between different inference techniques, the Autoregressive Neural Network (ANN), a Belief Propagation based approach (SIB) (17, 33), together with the Soft Margin estimator (SM), is carried out on both random graphs and real-world contact networks. The Soft Margin (SM) estimator [21] is based on Monte Carlo methods in which samples are weighted according to the overlap between the observations and the generated epidemic cascade (see SM for details). The Belief Propagation approach [23], implemented in the SIB software [27, 49] provides exact inference on acyclic contact networks and performs very effectively on sparse network structures. In the present work, we focus instead on real-world contact networks, which turns out to have relatively dense interaction patterns. The first contact network, taken from the dataset *InVS13* [50], is related to a work environment (**work**), while the second one was collected in a hospital (**hospital**) [22]. In both cases, the dataset used is the temporal list of contacts, respectively between 95 and 330 individuals, for a period of two weeks. Since the real duration $\delta_{i,j}^t$ of each contact is known, the probability of infections between individuals i, j at time t is computed as $\lambda_{i,j}^t = 1 - e^{-\gamma \delta_{i,j}^t}$, where γ is the rate of infection. For comparison, we also consider synthetic contact networks: a random regular graph (**rrg**) with $N = 100$ individuals and degree equal to 10, and a random geometric graph (**proximity**), in which $N = 100$ individuals are randomly placed on a square of linear size $\sqrt{N}$. In the latter, the probability that individuals i and j are in contact is $e^{-d_{ij}/l}$, where d_{ij} is the distance between i and j and l is a parameter (set to $l = 10$) that controls the density of contacts. For both synthetic contact networks, we generated epidemic processes (SIR epidemic cascades, see Material and Methods for details) with a duration of 15 days.

The patient zero problem

Given the exact knowledge of the final state of the epidemics at time T , the patient-zero problem consists in identifying the (possibly unique) source of the epidemics. In a Bayesian framework, this problem can be tackled by computing for each individual the marginal probability of being infected at time $t = 0$ given a set of observations $\mathcal{O}$. This quantity can be estimated from the posterior distribution (Eq. 6 in the Material and Methods) with all three algorithms (ANN, SIB and SM) considered in this work. For each contact network (**rrg**, **proximity**, **work**, **hospital**), results were computed averaging over 100 realizations of the epidemic model. The three algorithms were used to rank infected and recovered individuals in decreasing order according to the estimated probability of being infected at time zero for each epidemic realization. Fig. 4 displays, for each algorithm, the normalized cumulative number (over 100 instances) of patient-zero individuals actually found as a function of the accordingly ranked fraction of infected or recovered individuals. The ANN algorithm outperforms all the other methods as indicated by the larger values of the area under the curve (AUC) obtained in all cases considered. The improvement is also evident when analyzing the fraction of true patients zero correctly identified by each algorithm (left bar plots in Fig. 4). As an example, in the hospital case, ANN identifies correctly the patient zero in the 74% of the instances, SIB in the 54% and

SM in the 35% of them. In all cases, the performances the ANN algorithm are comparable or better than those of the other approaches. The results on the patient zero problem reveal the ability of the ANN algorithm to efficiently generate epidemic cascades according to the posterior probability defined in Eq. 6.

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

Scaling properties with the size of the epidemic outbreak

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

algorithm based on a Monte Carlo technique, it can give estimates of marginal probabilities with arbitrary accuracy when a sufficiently large sample of epidemic cascades is generated.

In the case of complete observation of the final state, the larger the epidemic size (i.e., the total number of infected individuals at time T), the larger is the number of epidemic cascades that are compatible with the observation. For instance, in an epidemic realization of duration T time steps in which n_I individuals are tested infected and $N - n_I$ tested susceptible at time T , the number of epidemic configurations compatible with the observations scales as T^{n_I} . Both ANN and SM rely on sampling procedures, so their performances could suffer from convergence issues when the epidemic size (n_I) increases. We compute the total number of samples generated by the ANN during the learning process and the number of samples of epidemic cascades generated by SM in the Monte Carlo procedure. In both cases, we assume that convergence is reached when $\sum_i |P_{\text{algo}}(x_i^0 = I|\mathcal{O}) - P_{\text{sib}}(x_i^0 = I|\mathcal{O})| < 0.1$ with $\text{algo} \in \{\text{ANN}, \text{SM}\}$, where $P_{\text{algo}}(x_i^0 = I|\mathcal{O})$ is the estimated marginal probability that individual i is infected at the initial time according to each method. The results on the scaling properties of the ANN and SM as a function of the epidemic size on a tree contact network with degree and depth both equal to 6 (**tree**) are shown in Figure 5. Here we set the duration of the epidemic cascades to $T = 15$ days. The ANN algorithm has a quasi-linear dependence with the epidemic size; conversely, the SM algorithm exhibits a very sharp increase in the number of simulations necessary for good estimates of the marginals, and already for epidemic sizes of order ten individuals, good estimates are difficult to obtain.

Epidemic risk assessment

The risk assessment problem consists in finding the individuals who have the highest probability of being infected at a specific time given a partial observation $\mathcal{O}$. In particular, we consider here a realization of the SIR model with $\mu = 0$ (i.e. only the states S and I are available) where half of the infected individuals are observed with certainty at the final time T . The results of the risk-assessment analysis obtained by means of the ANN algorithm are compared with those provided by the SIB algorithm and two other methods recently proposed in [27]. The Simple-Mean-Field (SMF) algorithm is based on a mean-field description of the epidemic process in which information about the observed individuals is heuristically included. The Contact Tracing (CT) algorithm computes the individual risks according to the number of contacts with observed infected individuals in the last $\tau = 5$ time steps (days).

A measure of the ability to correctly identify the unobserved infected individuals at the final time T is represented by the area under the Receiving Operating Characteristic (ROC) curve. This quantity, averaged over 100 instances of the epidemic realizations, is reported, for the methods considered above, in Table 1, for different contact networks (**rrg**, **proximity** and **work**). All algorithms perform similarly on random graphs, whereas ANN and SIB outperform the other two methods in the case of the **work** contact network.

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

	rrg 1 src	rrg 2 src	proximity 1 src	work 1 src
ANN	0.710 ± 0.010	0.670 ± 0.009	0.734 ± 0.010	0.889 ± 0.005
SIB	0.710 ± 0.010	0.671 ± 0.009	0.732 ± 0.010	0.886 ± 0.005
SMF	0.704 ± 0.010	0.671 ± 0.009	0.724 ± 0.009	0.796 ± 0.007
CT	0.685 ± 0.009	0.659 ± 0.008	0.711 ± 0.008	0.790 ± 0.006

Table 1 Epidemic risk assessment results. Area under the Receiving Operating Characteristic (ROC) curves for the risk assessment problem on random regular graphs (**rrg**) with 1 and 2 sources, on the **proximity** random graphs and **work** real-world contact network. The results are averaged over 100 different epidemic cascades generated with the same epidemic parameters. For each case, the ROC curve for the classification of the unobserved infected individuals at the final time is computed. In the **rrg** case, the epidemic parameters are $\lambda_{rrg}^{(1)} = 0.035$ for the single source and $\lambda_{rrg}^{(2)} = 0.03$ for the double source case. For the **proximity** random graphs, $\lambda_{prox} = 0.03$. In the case of the **work** network, the model has rate of infection $\gamma_{work} = 10^{-3}$.

Epidemic Parameters inference

The parameters Λ governing the epidemic process can be simultaneously inferred during the learning process of the ANN algorithm using a heuristic method inspired by Expectation Maximization (see Materials and Methods). Other iterative algorithms, such as SIB, can incorporate such a parameter likelihood climbing step during

their convergence [51]. A comparison between the performances of the two algorithms in learning the infectiousness parameter governing the spreading process on different contact networks (**tree**, **rrg**, **proximity** and **work**) is displayed in Fig.6 (left plot), in which we adopt the same setting of the patient-zero problem where the states of all individuals are known at the final time T . The ANN algorithm largely outperforms SIB in **rrg** and **proximity** graphs, obtaining comparable results for the **tree** and **work** instances. We also test the performance of parameters inference in a more challenging scenario where the population is split into two classes, with two different rates of infections γ_1, γ_2 (which could correspond, for instance, to a simplified scenario of vaccinated/not-vaccinated individuals). The states of all individuals at final time $T = 14$ are observed for ten epidemic cascades on the **hospital** contact network. Then we infer the parameters with two different epidemic models: in the first one, the population is correctly divided (we call this the *true* model); in the second, we split the population randomly (*null* model). The goal is to verify whether the *true* model has a larger likelihood than the *null* model, that is it can better explain the observations. In the central plot of Fig.6, we observe how well the *true* model can infer the correct values of the infections rate of the two sub-populations. As expected, the two values of γ inferred with the *null* model are similar to each other but different from the correct ones. From the rightmost plot of Fig.6, we observe that the log-likelihood of the *true* model is much larger of the one of the *null* model, indicating the former better explains the observations. This example shows how the ANN approach can therefore be used to select the epidemic model that best explains the observations based on the estimate of their log-likelihood.

Fig. 6 Inference of epidemic infectiousness parameters. **Left plot.** Average relative error in the inference of the infectiousness parameters over ten epidemic cascade per interaction graph. On **tree**, **rrg** and **proximity** networks, the discrete-time SIR model has infection probability respectively equal to $\lambda_{\text{tree}} = 0.35$, $\lambda_{\text{rrg}} = 0.04$, $\lambda_{\text{proximity}} = 0.03$. The **work** case has rate of infection $\gamma_{\text{work}} = 10^{-3}$. The initial conditions for the parameter learning process were set to $\lambda_{\text{init}} = 0.5$ for **tree**, $\lambda_{\text{init}} = 0.1$ for **RRG** and **proximity** networks and to $\gamma_{\text{init}} = 10^{-2}$ for the **work** network. **Central plot.** Box plot for the case of two classes of individuals with different rate of infection γ_1, γ_2 inferred by the ANN. We consider two inference model where the population is divided according the propagation model (*true* model) and randomly (*null* model), see the text for details. The *true* model is able to correctly infer the parameters with only ten different epidemic cascade. **Right plot.** Box blot of the log-likelihood difference between the *true* and *null* model.

3 Discussion

This work shows how significant individual-based epidemiological inference problems defined on contacts networks can be successfully addressed using autoregressive neural networks. In problems such as patient zero detection and epidemic risk assessment, the proposed method exploits the generative power of autoregressive neural networks, in order to learn to generate epidemic realizations that are sampled according to the epidemic model and, at the same time, are compatible with the observations. When the parameters of the model are unknown, it is also capable of inferring them during the learning process. The approach is flexible enough to be easily applied to other epidemic inference problems and with different propagation models. The proposed architectures for the autoregressive networks significantly reduce the number of necessary parameters with respect to the vanilla implementation. Moreover, convergence properties are improved by means of a regularization method which exploits the introduction of a fictitious temperature and an associated annealing process.

According to the results obtained on three different problems (patient zero, risk assessment, and parameters inference) on both synthetic and real contact networks, the proposed method equals the currently best methods in the literature on epidemic inference, outperforming them in several cases. In particular, the ANN approach is computationally less demanding than standard Monte Carlo methods and very effective also on dense contact networks, where more efficient algorithms based on message-passing methods might experience convergence issues. For this reason, the method seems very promising for epidemic inference problems defined in small communities such as hospitals, workplaces, schools, cruises, where contact data could be available. In such contexts it could be used to detect the source of an outbreak, measure the risk of individuals being infected to improve contact tracing, or to estimate the channels of contagion as well as the infectivity of classes of people thanks to the possibility of inferring the propagation parameters.

Even though showing suitable scaling property with the system size, the proposed approach, reasonably, needs new architectural schemes and learning processes to be applied in epidemic inference problems regarding more than several thousand individuals. Work is in progress in this direction, possibly guided by symmetries and regularity of the prior epidemic models.

Acknowledgments. This work has been supported by the SmartData@PoliTO center on Big Data and Data Science. We acknowledge computational resources by HPC@POLITO (<http://www.hpc.polito.it>).

Author contributions

I.B. conceived the original idea, I.B and F.M. implemented the algorithms, I.B., A.B., L.D.A., F.M. contributed to the development of the project, the analysis of the results, and wrote the manuscript.

Declarations

The authors declare that they have no competing interests. All data and code needed to evaluate the conclusions are released on GitHub with DOI: [released after acceptance].

References

- [1] Kermack, W., Mckendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **115**, 700–721 (1927). <https://doi.org/10.1098/rspa.1927.0118>
- [2] Keeling, M.J., Rohani, P.: Modeling Infectious Diseases in Humans and Animals. Princeton university press, ??? (2011)
- [3] Pastor-Satorras, R., Castellano, C., Mieghem, P.V., Vespignani, A.: Epidemic processes in complex networks. Reviews of Modern Physics **87**, 925 (2015). <https://doi.org/10.1103/RevModPhys.87.925>
- [4] Vespignani, A.: Modelling dynamical processes in complex socio-technical systems. Nature Physics **8**, 32–39 (2012). <https://doi.org/10.1038/nphys2160>
- [5] Brauer, F., Driessche, P.d., Wu, J.: Lecture notes in mathematical epidemiology. Berlin, Germany. Springer **75**(1), 3–22 (2008)
- [6] Eubank, S., Guclu, H., Kumar, V.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. Nature **429**(6988), 180–184 (2004)
- [7] Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A.: Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences **106**(51), 21484–21489 (2009)
- [8] Merler, S., Ajelli, M., Pugliese, A., Ferguson, N.M.: Determinants of the spatiotemporal dynamics of the 2009 h1n1 pandemic in europe: implications for real-time modelling. PLoS Comput Biol **7**(9), 1002205 (2011)
- [9] Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A.P., Mu, K., Rossi, L., Sun, K., *et al.*: The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science **368**(6489), 395–400 (2020)
- [10] Ferguson, N.M., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., Burke, D.S.: Strategies for containing an emerging influenza pandemic in southeast asia. Nature **437**, 209–214 (2005). <https://doi.org/10.1038/nature01621>

- [//doi.org/10.1038/nature04017](https://doi.org/10.1038/nature04017) 737
- [11] Halloran, M.E., Ferguson, N.M., Eubank, S., Longini, I.M., Cummings, D.A.T., 738
Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T.C., Wagener, D., 739
Beckman, R., Kadau, K., Barrett, C., Macken, C.A., Burke, D.S., Cooley, P.: 740
Modeling targeted layered containment of an influenza pandemic in the united 741
states. *Proceedings of the National Academy of Sciences of the United States* 742
of America **105**, 4639–4644 (2008). <https://doi.org/10.1073/pnas.0706849105> 743
744
- [12] Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, 745
X., Zhao, S., Huang, Y., et al.: Transmission heterogeneities, kinetics, and 746
controllability of sars-cov-2. *Science* **371**(6526) (2021) 747
748
- [13] Sisson, S.A., Fan, Y., Tanaka, M.M.: Sequential monte carlo without likeli- 749
hoods. *Proceedings of the National Academy of Sciences* **104**(6), 1760–1765 750
(2007) 751
752
- [14] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.: Approximate 753
bayesian computation scheme for parameter inference and model selection in 754
dynamical systems. *Journal of the Royal Society Interface* **6**(31), 187–202 755
(2009) 756
757
- [15] Kypraios, T., Neal, P., Prangle, D.: A tutorial introduction to bayesian infer- 758
ence for stochastic epidemic models using approximate bayesian computation. 759
Mathematical biosciences **287**, 42–53 (2017) 760
761
- [16] Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, 762
L., Parker, M., Bonsall, D., Fraser, C.: Quantifying sars-cov-2 transmission 763
suggests epidemic control with digital contact tracing. *Science* **368** (2020). 764
<https://doi.org/10.1126/science.abb6936> 765
766
- [17] Wymant, C., Ferretti, L., Tsallis, D., Charalambides, M., Abeler-Dörner, L., 767
Bonsall, D., Hinch, R., Kendall, M., Milsom, L., Ayres, M., Holmes, C., Briers, 768
M., Fraser, C.: The epidemiological impact of the nhs covid-19 app. *Nature* 769
(2021). <https://doi.org/10.1038/s41586-021-03606-z> 770
771
- [18] Mastrandrea, R., Fournet, J., Barrat, A.: Contact patterns in a high school: a 772
comparison between data collected using wearable sensors, contact diaries and 773
friendship surveys. *PloS one* **10**(9), 0136497 (2015) 774
775
- [19] Smieszek, T., Castell, S., Barrat, A., Cattuto, C., White, P.J., Krause, G.: Con- 776
tact diaries versus wearable proximity sensors in measuring contact patterns at 777
a conference: method comparison and participants’ attitudes. *BMC infectious* 778
diseases **16**(1), 1–14 (2016) 779
780
- [20] Sapiezynski, P., Stopczynski, A., Lassen, D.D., Lehmann, S.: Interaction data 781
from the copenhagen networks study. *Scientific Data* **6**(1), 1–10 (2019) 782

- [21] Antulov-Fantulin, N., Lančić, A., Šmuc, T., Štefančić, H., Šikić, M.: Identification of Patient Zero in Static and Temporal Networks: Robustness and Limitations. *Physical Review Letters* **114**(24), 248701 (2015). <https://doi.org/10.1103/PhysRevLett.114.248701>. Publisher: American Physical Society. Accessed 2021-02-02
- [22] Obadia, T., Silhol, R., Opatowski, L., Temime, L., Legrand, J., Thiébaud, A.C.M., Herrmann, J.L., Éric Fleury, Guillemot, D., Boëlle, P.Y.: Detailed contact data and the dissemination of staphylococcus aureus in hospitals. *PLoS Computational Biology* **11**, 1004170 (2015). <https://doi.org/10.1371/journal.pcbi.1004170>
- [23] Altarelli, F., Braunstein, A., Dall'Asta, L., Lage-Castellanos, A., Zecchina, R.: Bayesian inference of epidemics on networks via belief propagation. *Physical Review Letters* **112**(11), 118701 (2014). <https://doi.org/10.1103/PhysRevLett.112.118701>. Accessed 2014-05-06
- [24] Lokhov, A.Y., Mézard, M., Ohta, H., Zdeborová, L.: Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Phys. Rev. E* **90**, 012801 (2014). <https://doi.org/10.1103/PhysRevE.90.012801>
- [25] Altarelli, F., Braunstein, A., Dall'Asta, L., Zecchina, R.: Optimizing spread dynamics on graphs by message passing. *Journal of Statistical Mechanics: Theory and Experiment* **2013**(09), 09011 (2013)
- [26] Altarelli, F., Braunstein, A., Dall'Asta, L., Wakeling, J.R., Zecchina, R.: Containing epidemic outbreaks by message-passing techniques. *Phys. Rev. X* **4**, 021024 (2014). <https://doi.org/10.1103/PhysRevX.4.021024>
- [27] Baker, A., Biazzo, I., Braunstein, A., Catania, G., Dall'Asta, L., Ingrosso, A., Krzakala, F., Mazza, F., Mézard, M., Muntoni, A.P., Refinetti, M., Sarao Mannelli, S., Zdeborová, L.: Epidemic mitigation by statistical inference from contact tracing data. *Proceedings of the National Academy of Sciences* **118**(32) (2021) <https://arxiv.org/abs/https://www.pnas.org/content/118/32/e2106548118.full.pdf>. <https://doi.org/10.1073/pnas.2106548118>
- [28] Van Oord, A., Kalchbrenner, N., Kavukcuoglu, K.: Pixel recurrent neural networks. In: *International Conference on Machine Learning*, pp. 1747–1756 (2016). PMLR
- [29] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.W., Kavukcuoglu, K.: Wavenet: A generative model for raw audio. *CoRR* **abs/1609.03499** (2016) <https://arxiv.org/abs/1609.03499>
- [30] Bengio, Y., Ducharme, R., Vincent, P., Janvin, C.: A neural probabilistic language model. *The journal of machine learning research* **3**, 1137–1155

- (2003) 829
830
- [31] Keskar, N.S., McCann, B., Varshney, L.R., Xiong, C., Socher, R.: Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858* (2019) 831
832
833
834
- [32] Trinquier, J., Uguzzoni, G., Pagnani, A., Zamponi, F., Weigt, M.: Efficient generative modeling of protein sequences using simple autoregressive models. *arXiv preprint arXiv:2103.03292* (2021) 835
836
837
838
- [33] Larochelle, H., Murray, I.: The neural autoregressive distribution estimator. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 29–37 (2011). *JMLR Workshop and Conference Proceedings* 839
840
841
842
- [34] Germain, M., Gregor, K., Murray, I., Larochelle, H.: Made: Masked autoencoder for distribution estimation. In: *International Conference on Machine Learning*, pp. 881–889 (2015). *PMLR* 843
844
845
846
- [35] Uribe, B., Côté, M.-A., Gregor, K., Murray, I., Larochelle, H.: Neural autoregressive distribution estimation. *The Journal of Machine Learning Research* **17**(1), 7184–7220 (2016) 847
848
849
850
- [36] Wu, D., Wang, L., Zhang, P.: Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.* **122**, 080602 (2019). <https://doi.org/10.1103/PhysRevLett.122.080602> 851
852
853
854
- [37] Sharir, O., Levine, Y., Wies, N., Carleo, G., Shashua, A.: Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Phys. Rev. Lett.* **124**, 020503 (2020). <https://doi.org/10.1103/PhysRevLett.124.020503> 855
856
857
858
- [38] Pan, F., Zhou, P., Zhou, H.-J., Zhang, P.: Solving statistical mechanics on sparse graphs with feedback-set variational autoregressive networks. *Physical Review E* **103**(1), 012103 (2021) 859
860
861
862
- [39] McNaughton, B., Milošević, M.V., Perali, A., Pilati, S.: Boosting monte carlo simulations of spin glasses using autoregressive neural networks. *Phys. Rev. E* **101**, 053312 (2020). <https://doi.org/10.1103/PhysRevE.101.053312> 863
864
865
866
- [40] Hibat-Allah, M., Inack, E.M., Wiersema, R., Melko, R.G., Carrasquilla, J.: Variational neural annealing. *Nature Machine Intelligence* (2021). <https://doi.org/10.1038/s42256-021-00401-3> 867
868
869
870
- [41] Pal, R., Sekh, A.A., Kar, S., Prasad, D.K.: Neural network based country wise risk prediction of covid-19. *Applied Sciences* **10**(18) (2020). <https://doi.org/10.3390/app10186448> 871
872
873
874

- [42] Philemon, M.D., Ismail, Z., Dare, J.: A review of epidemic forecasting using artificial neural networks. *International Journal of Epidemiologic Research* **6**(3), 132–143 (2019)
- [43] Chimmula, V.K.R., Zhang, L.: Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals* **135**, 109864 (2020). <https://doi.org/10.1016/j.chaos.2020.109864>
- [44] Shah, C., Dehmamy, N., Perra, N., Chinazzi, M., Barabási, A.-L., Vespignani, A., Yu, R.: Finding patient zero: Learning contagion source with graph neural networks (2020)
- [45] Tomy, A., Razzanelli, M., Di Lauro, F., Rus, D., Della Santina, C.: Estimating the state of epidemics spreading with graph neural networks. *arXiv preprint arXiv:2105.05060* (2021)
- [46] to release after acceptance. to release
- [47] Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press, ??? (2016)
- [48] Oord, A.v.d., Kalchbrenner, N., Vinyals, O., Espeholt, L., Graves, A., Kavukcuoglu, K.: Conditional image generation with pixelcnn decoders. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS' 16*, pp. 4797–4805. Curran Associates Inc., Red Hook, NY, USA (2016)
- [49] Sib: [s]tatistical [i]nference in epidemics via [b]elief propagation. <https://github.com/sibyl-team/sib>
- [50] G'enois, M., Barrat, A.: Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science* **7**(1), 11 (2018). <https://doi.org/10.1140/epjds/s13688-018-0140-1>
- [51] Braunstein, A., Ingrosso, A., Muntoni, A.P.: Network reconstruction from infection cascades. *Journal of The Royal Society Interface* **16**(151), 20180844 (2019). <https://doi.org/10.1098/rsif.2018.0844>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [sm.pdf](#)