DATI, AI E ROBOTICA @POLITO

RICERCA, TRASFERIMENTO TECNOLOGICO E SUPPORTO ALLE AZIENDE SUI TEMI FONDAMENTALI DEI BIG DATA, INTELLIGENZA ARTIFICIALE, ROBOTICA E RIVOLUZIONE DIGITALE

RECONSTRUCTION OF EPIDEMIC CASCADES WITH GENERATIVE NEURAL NETWORKS

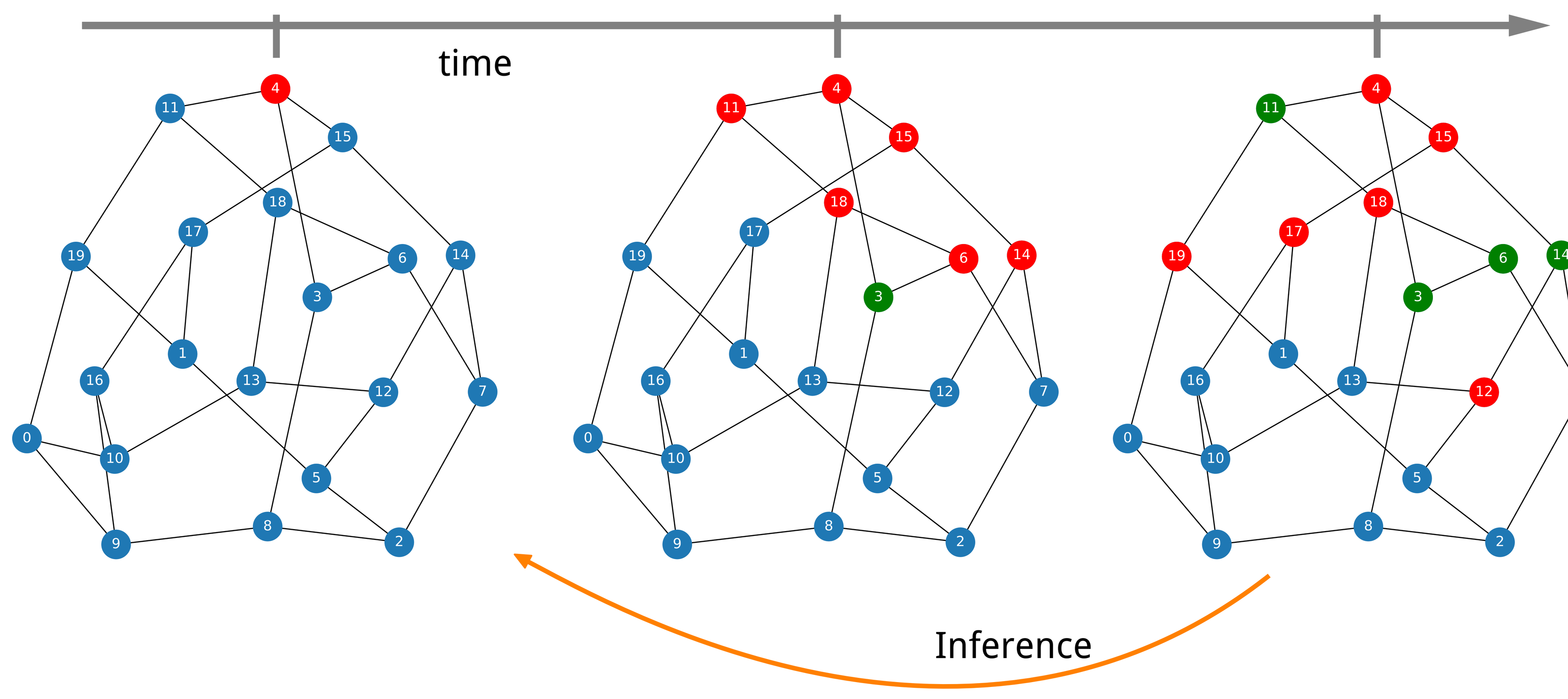
Authors: Indaco Biazzo, Fabio Mazza, Luca Dall'Asta, Alfredo Braunstein

Politecnico di Torino

SmartData@PoliTO

MOTIVATION AND BACKGROUND

Discrete-state stochastic compartmental models have been traditionally used to model infectious diseases, and provide a simple framework for a variety of spreading processes in social and technological systems.
The COVID-19 pandemic has highlighted the need to track epidemics on an individual level, and has pushed for the large scale collection of contact data in order to contain the spread. The inference at the individual level, however, becomes much more challenging as the number of possible epidemic realizations increases exponentially with the number of individuals or the duration.



Different strategies have been devised to efficiently solve this problem from statistical inference. Algorithms based on Belief Propagation (BP) show good results, but they may experience convergence issues on some dense and very structured contact networks.

MATERIALS AND METHODS

In the Bayesian formulation of the inference, we have a posterior distribution from which we need to compute marginal probabilities.

$$P(\mathbf{x}|\mathcal{O}) = \frac{1}{P(\mathcal{O})} \prod_i \gamma(x_i^0) \prod_{t=1}^{T} P(x_i^t|\mathbf{x}^{t-1}) \prod_r P(\mathcal{O}_r|\mathbf{x})$$

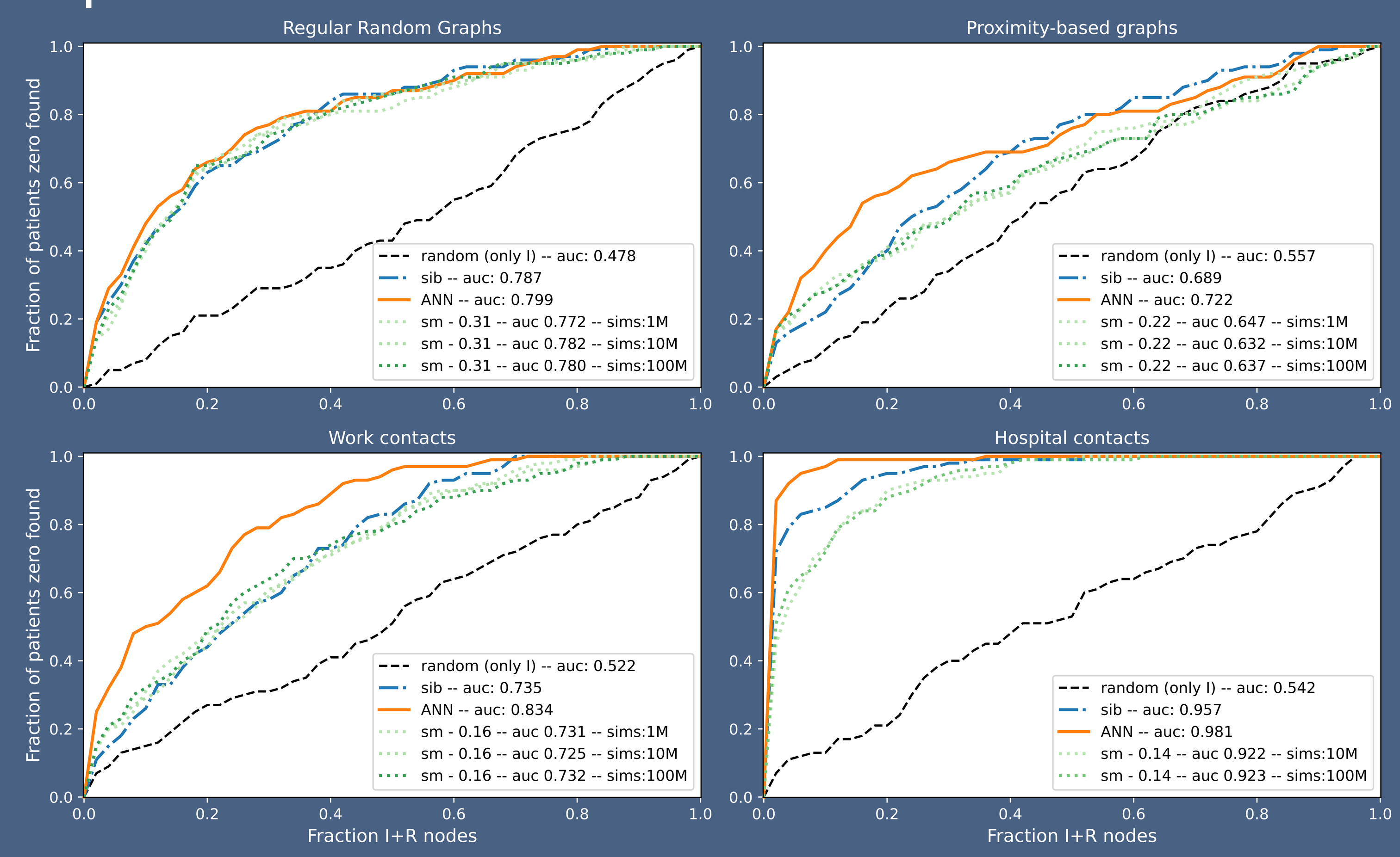
We use Autoregressive Neural Networks (ANNs), which are able to efficiently sample from a probability distribution, and have already been used for text or image generation tasks.
We train our ANNs by generating samples of epidemic cascades and then evaluating the (reversed) Kullback-Leibler divergence, from which we have derived a loss function.

When this procedure ends, we are able to generate samples from $P(\mathbf{x}|\mathcal{O})$, easily computing marginals.

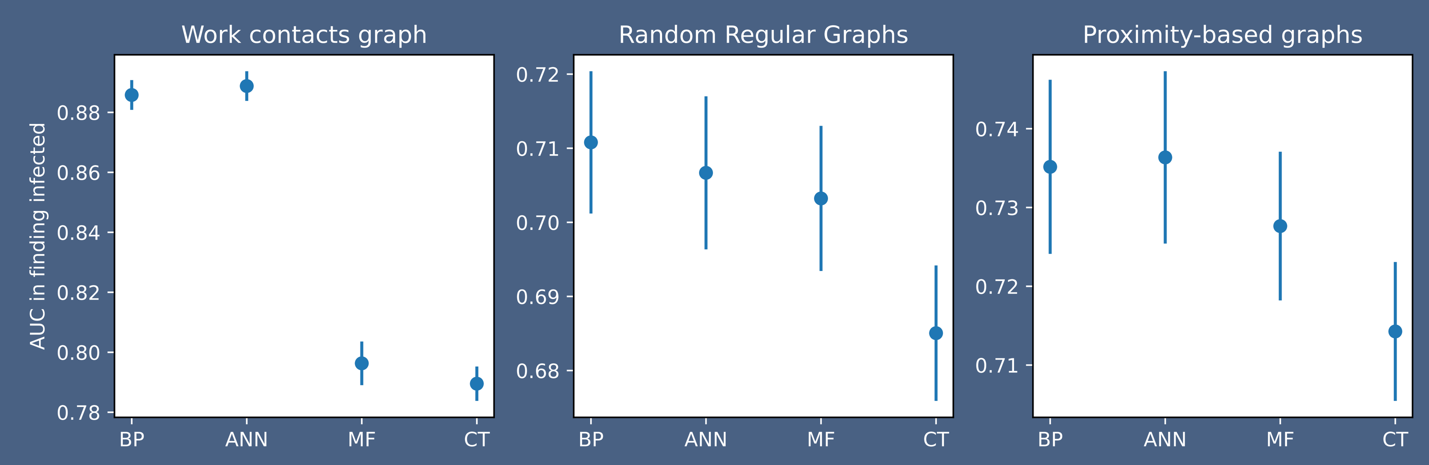
RESULTS

We apply the ANN method to two inference problems:

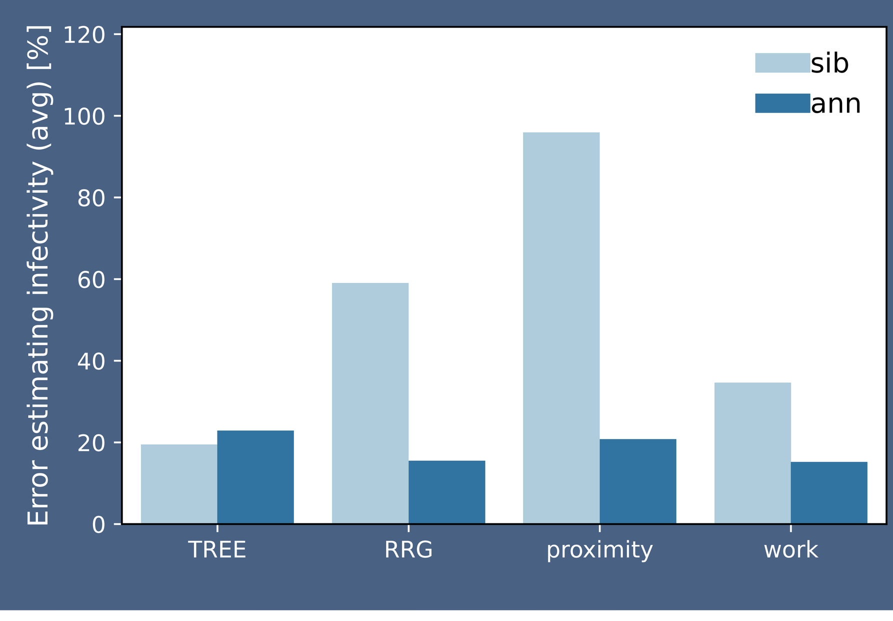
- the patient zero, where we have observed the final state and want to find the source of the epidemic



- the risk inference problem, where we want to find who is at most risk of being infected, given some sparse observations on the ongoing epidemic



We are also able to use the Bayesian framework to infer the parameters of the epidemic



CONCLUSIONS AND FUTURE WORK

We show that ANNs are able to learn to generate epidemic cascades according to the epidemic model, which are compatible with the observations or evidence; if the model's parameters are unknown, they are also capable of inferring them during the overall learning process.
The approach is flexible enough to be easily used on various inference problems and with different propagation models.