Network Reconstruction from Noisy and Incomplete Spreading Dynamics

Mateusz Wilinski[1] and Andrey Y Lokhov[1]

[1] Theoretical Division, Los Alamos National Laboratory, Los Alamos, US

Long distance connections in modern interconnected world play an important role in many areas of life, such as information spreading, epidemics, financial contagion or opinion dynamics. This drives the need for proper understanding of diffusion processes on networks. Another unprecedented feature of current era is the data availability, which, together with rapid development of machine learning tools, allow to learn and predict models from observed processes. In reality, however, these large amounts of data are often incomplete, noisy or biased. We address this problem in the case of spreading processes on networks and propose a general framework, which allow to learn spreading models from data, when the latter is incomplete or subject to uncertainty.

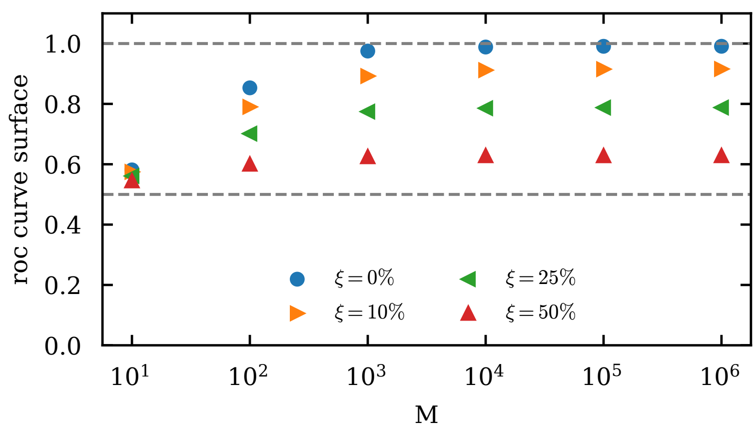

Figure 1: ROC curve surface for reconstruction of network edges, as a function of available sample size (number of observed spreads). Percentage of unobserved nodes is denoted by ξ. Experiment was conducted for an Erdos-Renyi graph with average degree $\langle k \rangle = 3$, number of nodes $N = 100$ and maximum spreading time $T = 5$. The results were averaged over five different network instances.

Starting from the algorithm described in [1] we propose an extended methodology, which allows to learn spreading parameters, together with the network structure, when not only part of the nodes is unobserved, but there is an additional uncertainty regarding the observed part. Results of network reconstruction in the case of Erdos-Renyi graphs are shown in FIG. 1. Since the algorithm is based on a *dynamics message passing* inference procedure, it is particularly useful in the case of locally tree-like graphs. Additionally, we present an effective implementation of the algorithm, which assures linear complexity even for heterogeneous networks and show how the procedure can be more effective when an additional information about the process is known.

References

[1] Wilinski, M., and Lokhov, A. "Prediction-Centric Learning of Independent Cascade Dynamics from Partial Observations." *International Conference on Machine Learning*. PMLR, 2021.