Network reconstruction from noisy and incomplete spreading dynamics

Mateusz Wilinski, Andrey Lokhov
Los Alamos National Laboratory, Los Alamos, United States

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. Starting from an algorithm based on dynamic message passing, 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. Since the algorithm is based on a message passing inference procedure, it is particularly useful in the case of locally treelike 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. The approach can also be extended towards building effective models, which can have significant predictive power, even in the case where the graph is loopy. We show the effectiveness of our algorithm both on different types of synthetic networks, including scale-free and regular graphs, as well as on real-world networks. We test different scenarios of data incompleteness and uncertainty, together with their combinations. We also explore the limits of the algorithm and its dependence on the network structure. Finally, we present a full derivation of the dynamic message passing equations from the canonical dynamic belief propagation, utilizing the knowledge about particular type of dynamics. Quality of results in the case of a Facebook subgraph with both unobserved nodes and noisy observations, is shown in the attached figure. It shows the expected error of reconstructed spreading parameters as a function of the sample size and for different percentage of unobserved nodes.