MAR16-2015-020073

Nonparametric inference of network structure and dynamics

TIAGO P. PEIXOTO, University of Bremen

The network structure of complex systems determine their function and serve as evidence for the evolutionary mechanisms that lie behind them. Despite considerable effort in recent years, it remains an open challenge to formulate general descriptions of the large-scale structure of network systems, and how to reliably extract such information from data. Although many approaches have been proposed, few methods attempt to gauge the statistical significance of the uncovered structures, and hence the majority cannot reliably separate actual structure from stochastic fluctuations. Due to the sheer size and high-dimensionality of many networks, this represents a major limitation that prevents meaningful interpretations of the results obtained with such nonstatistical methods. In this talk, I will show how these issues can be tackled in a principled and efficient fashion by formulating appropriate generative models of network structure that can have their parameters inferred from data. By employing a Bayesian description of such models, the inference can be performed in a nonparametric fashion, that does not require any *a priori* knowledge or *ad hoc* assumptions about the data. I will show how this approach can be used to perform model comparison, and how hierarchical models yield the most appropriate trade-off between model complexity and quality of fit based on the statistical evidence present in the data. I will also show how this general approach can be elegantly extended to networks with edge attributes, that are embedded in latent spaces, and that change in time. The latter is obtained via a fully dynamic generative network model, based on arbitrary-order Markov chains, that can also be inferred in a nonparametric fashion. Throughout the talk I will illustrate the application of the methods with many empirical networks such as the internet at the autonomous systems level, the global airport network, the network of actors and films, social networks, citations among websites, voting correlations among politicians, co-occurrence of disease-causing genes and others.