

A Bayesian generative model for Pervasive Overlap

Sune Lehmann, Morten Mørup, and Mikkel N. Schmidt
Technical University of Denmark

February 29, 2012

When a complex network exhibits pervasive overlap, nodes can belong to multiple densely connected groups. Pervasive overlap describes structures in social networks where each person is a member of multiple social circles; in biological networks where many proteins and metabolites play multiple roles, or in information networks, such as the word association network, where each word is used in multiple contexts. We propose a non-parametric Bayesian generative model for pervasive overlap based on the assumption that modules act as independent causes of links. The developed techniques are principled, interpretable, and computationally efficient. Our generative framework connects the model to well studied techniques for statistical inference, as well as analysis by standard statistical tests (including link-prediction), and principled handling of missing data. Computationally, the method scales linearly in the number of edges in the graph allowing for use in large scale network analysis. The model is able to adapt to the complexity of the data by learning the number of groups through the use of Bayesian non-parametrics.