Statistical Sciences
UNIVERSITY OF TORONTO

Pseudo-likelihood methods for community detection in large sparse networks

Arash Amini, University of Michigan

We consider the problem of community detection in a network, that is, partitioning the nodes into groups that, in some sense, reveal the structure of the network. Many algorithms have been proposed for fitting network models with communities, but most of them do not scale well to large networks, and often fail on sparse networks. We present a fast pseudo-likelihood method for fitting the stochastic block model, a well-known model for networks with communities, as well as a variant that allows for an arbitrary degree distribution by conditioning on degrees.

We provide empirical results showing that the algorithms perform well under a range of settings, including on very sparse networks, and illustrate on the example of a network of political blogs. We also present spectral clustering with perturbations, a method of independent interest, which works well on sparse networks where regular spectral clustering fails, and use it to provide an initial value for pseudo-likelihood. We discuss theoretical results showing that pseudo-likelihood provides consistent estimates of the communities under mild conditions on the starting value, for the case of a block model with two communities. Time permitting, we give some insights as to why perturbations help with spectral clustering on sparse networks.

Tuesday February 11, 2014 at 3:30pm

Sidney Smith Hall, Room 2118

Refreshments will be served at 3:15pm