MICHIGAN STATE UNIVERSITY
Department of Statistics and Probability

COLLOQUIUM

David Hunter
Department of Statistics
Pennsylvania State University

Model-based Clustering of Large Networks

(Joint work with Duy Q. Vu & Michael Schweinberger)

Tuesday, January 28, 2014
10:20 a.m. - 11:10 am
Refreshments 10:00 am
C405 Wells Hall

Abstract:

A network clustering framework, based on finite mixture models, is described. It can be applied to discrete-valued networks with hundreds of thousands of nodes and billions of edge variables. Relative to other recent model-based clustering work for networks, we introduce a more flexible modeling framework, improve the variational-approximation estimation algorithm, discuss and implement standard error estimation via a parametric bootstrap approach, and apply these methods to much larger datasets than those seen elsewhere in the literature. The more flexible modeling framework is achieved through introducing novel parameterizations of the model, giving varying degrees of parsimony, using exponential family models whose structure may be exploited in various theoretical and algorithmic ways. The algorithms, which we show how to adapt to the more complicated optimization requirements introduced by the constraints imposed by the novel parameterizations we propose, are based on variational generalization EM algorithms, where the E-steps are augmented by a minorization-maximization (MM) idea. The bootstrapped standard error estimates are based on an efficient Monte Carlo network simulation idea. Last, we demonstrate the usefulness of the model-based clustering framework by applying it to a discrete-valued network with more than 131,000 nodes and 17 billion edge variables.

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.