

UCLA

# BIOSTATISTICS SEMINAR

FALL 2012

## Fast Inference for Model-Based Clustering of Networks Using an Approximate Case-Control Likelihood

**Adrian Raftery, PhD**

Professor of Statistics and Sociology  
University of Washington

Tuesday, December 4, 2012

3:00pm - 4:00pm, 3656 Geology Bldg

Refreshments served at 2:20 PM in room 51-254 CHS

**ABSTRACT:** The model-based clustering latent space network model represents relational data visually and takes account of several basic network properties. Due to the structure of its likelihood function, the computational cost is of order  $O(N^2)$ , where  $N$  is the number of nodes. This makes it infeasible for large networks. We propose an approximation of the log likelihood function. We adapt the case-control idea from epidemiology and construct an approximate case-control log likelihood which is an unbiased estimator of the full log likelihood. Replacing the full likelihood by the case-control likelihood in the MCMC estimation of the latent space model reduces the computational time from  $O(N^2)$  to  $O(N)$ , making it feasible for large networks. We evaluate its performance using simulated and real data. We fit the model to a large protein-protein interaction data using the case-control likelihood and use the model fitted link probabilities to identify false positive links. This is joint with with Xiaoyue Niu, Peter Hoff and Ka Yee Yeung.