Chapter 48 Outline

Network-Analysis: Estimation and Inference

John P. Schoeneman[1] **and Bruce A. Desmarais**[1]

[1]Penn State, Political Science, University Park, Pond Lab

Overview: Most statistical models used in the social sciences rely on the assumption that the observations in a dataset (or at least clusters of observation) are drawn independently from a common data generating process. However, when analyzing networks it is often inappropriate to assume that observations are independent. Indeed, the network analyst is often interested in studying the ways in which observations depend upon each other (e.g., do friends influence each others' choices to vote, do legislators reciprocate support for legislation). In this chapter we will review three of the most popular statistical models that have been developed to analyze network data, along with their extensions. These include the exponential random graph model (ERGM), the latent space model (LSM), and the stochastic block model (SBM). We will review the application of each method to networks in which ties are binary (e.g., are two legislators on at least one committee together), count (e.g., on how many committees do two legislators both serve), and continuous (e.g., how much time do two legislators spend together in committee meetings each week). The complexity and analytical intractability of the mathematical forms of these models render exact approaches to estimation infeasible. As such, estimation in these models proceeds by sampling from a Bayesian posterior and/or approximate methods of maximum likelihood. We will review these estimation methods, and how they are applied to the ERGM, LSM, and SBM. One complication that arises from the variety of sampling-based and approximate approaches to inference with network models is that standard information criteria that would normally be used to select between different families of statistical models (e.g., AIC, BIC, DIC) cannot be calculated in a manner that is comparable across different families of network models. Recently, researchers have turned to various methods of out-of-sample prediction in order to compare and select among network models in terms of generalization error. We will conduct a simulation study to evaluate the effectiveness of comparing the performance of models in terms of generalization error. This simulation study will be, to our knowledge, the first experiment designed to evaluate approaches to comparing network models based on generalization error.

This chapter will be divided into three sections. In the first section we will give an overview of the most common models used in network analysis. In the second section we will discuss the approaches to estimation that are commonly used with these models. In the last section we will conduct a simulation study with the goal of evaluating the effectiveness of measuring generalization error to assist practitioners in selecting among these models.

Models and Variations

- MRQAP, LSM (Bilinear and Euclidean), ERGM (including TERGM and GERGM), Mixed Membership Stochastic Blockmodel (MMSBM), SAOM/Siena

Inference Frameworks
- ML, pseudo-ML, and Bayesian estimation

Model Selection

- Directed vs Undirected

- Edge Types: Binary, Ordinal ,Count, Continuous

- Parameters

- Network vs Actor Inference

- Longitudinal variations

  - Discrete time, Autoregressive

  - Continuous time, process based

- Missingness

Simulation Study

- Using a set of common parameters from publications in political science network applications, we simulate networks for common edge types and evaluate performance in out of sample predictions for the models discussed above.

- In addition to the types of edges, we can vary levels of missingness and sparsity.