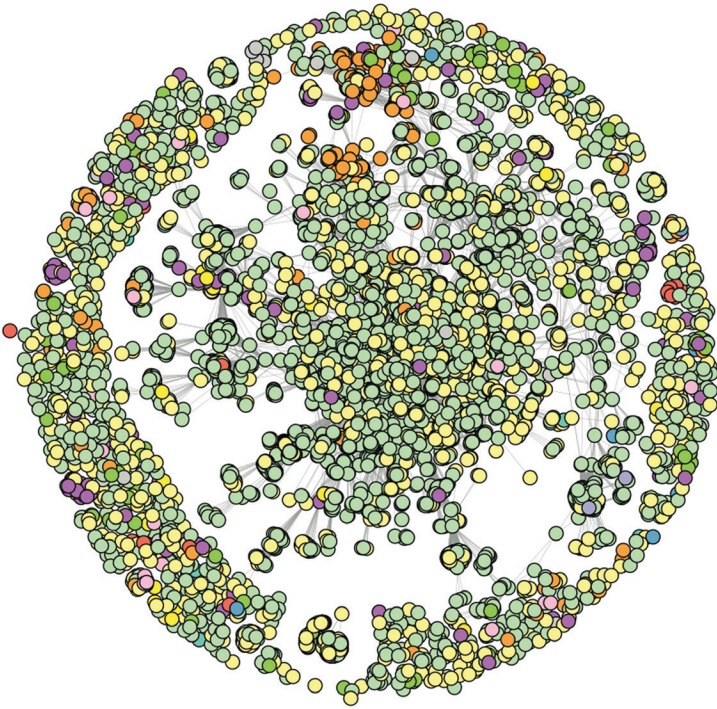


NETWORK ANALYSIS

Box-Steffensmeier team tweaking model to account for randomization

The analysis of biological and social networks has become increasingly important in recent years. Inferential and predictive statistical models that analyze networks have been put to use in such areas as epidemiology, public health, molecular biology and the social sciences.



$$\Pr(\mathbf{y} \mid \mathbf{x}, \theta, \mathcal{Y}) = \frac{\exp\{\theta^T \Gamma(\mathbf{y}, \mathbf{x})\}}{\sum_{\forall \tilde{\mathbf{y}} \in \mathcal{Y}} \exp\{\theta^T \Gamma(\tilde{\mathbf{y}}, \mathbf{x})\}},$$

$$y_{ij} = -3.25\text{edges}_{ij} + 0.75\text{gwes}_{ij} + 0.25 \times \mathbf{1}(\text{group}_i = \text{group}_j) + \gamma(\sigma_i - \sigma_j).$$

$$\Pr(\mathbf{y} \mid \cdot) = \frac{1}{\kappa} \exp\{\theta^T \Gamma(\mathbf{y}, \mathbf{x}) + w_i + w_j\},$$

Above: This graph displays the network mapping of all interest groups, color-coded by industry area, that signed amicus briefs on Court per curiam or full opinions from 2000 to 2009.

Yet, while network analysis continues to attract a great deal of attention from scholars and the broader public, the quantitative study of networks remains in the early stages of development. Only recently have the statistical theory and computational techniques been developed to rigorously analyze many types of networks, but there remain significant gaps in the statistical tools that limit their usefulness to scientists.

A research group from The Ohio State University is seeking to improve the well known and widely used Exponential Random Graph Model, or ERGM, by adding a randomization effect – called a frailty term – to account for heterogeneity when modeling network formation.

“ERGMs provide a means of incorporating both structural network characteristics and individual-level traits in models that try to explain the network as a whole,” said Janet Box-Steffensmeier, Ph.D., the Vernal Riffe Professor of Political Science at The Ohio State University. “However, while ERGMs offer important advantages over other methods of modeling network data, the approach currently lacks a method of incorporating unexplained heterogeneity, whether unmeasured, unobserved or unimagined, into these models. This is an important gap that limits the applicability of ERGMs in many areas of potential interest.”

The first goal of the project has been to refine the statistical model by introducing individual (or group) random effects in order to model that heterogeneity in the propensity for individuals in the network to form ties. The second goal of the project has been to evaluate the statistical properties of the proposed model and to apply the model to well-known health behavior and social science data sets. The researchers leveraged Ohio Supercomputer Center systems to evaluate the model with a series of Monte Carlo experiments, which is a standard workhorse approach to assessing the performance of complex models in computational statistics.

“The significance of the project is that it will help fill a gap in the ERGM literature, allowing researchers to draw correct inferences in a broader set of circumstances,” said Box-Steffensmeier. “And, for our own purposes, we are specifically interested in analyzing the diffusion of smoking and obesity in large social networks.”

Project Lead: Janet M. Box-Steffensmeier, Ph.D., The Ohio State University

Research Title: Modeling unobserved heterogeneity in network formation with the frailty exponential graph model

Funding Source: National Science Foundation

Website: <http://polisci.osu.edu/faculty/jbox/>



Ohio Supercomputer Center

An OH·TECH Consortium Member