Social Network Analysis: Node and Graph Level Statistics Part 2

EPIC - SNA, Columbia University

Zack W Almquist

June 13th, 2018

University of Minnesota

Introduction to Statisticsl Inference

Cohesion and Subgroups

Introduction to Statisticsl Inference

A First Look at Random Graphs

- So far, we have been measuring properties of graphs (and modeling their effects)
- Next step: modeling graphs themselves
  - Long-running and ongoing research area
  - Will crop up repeatedly in the coming weeks
- Today – a quick introduction to some basic families
  - We'll see some uses of these model families in the next lecture...

The Notion of Random Graphs

- Let $G = (V, E)$ be a graph. If E (and perhaps V) is a random set, then G is a random graph
  - Can consider G to be a random variable on some set G of possible graphs (“multinomial" representation)
  - Write probability mass function (pmf) as $\Pr(G = g)$
- Let Y be the adjacency matrix of random graph G. Then Y is a random matrix
  - Write graph pmf as $\Pr(Y = y)$
  - Y_{ij} is a binary random variable which indicates the state of the (random) i, j edge
  - $\Pr(Y_{ij} = y_{ij})$ is the (marginal) proability of the Y_{ij} edge state

Carter Butts. Social Network Methods. University of California, Irvine.

“Classical" Random Graphs

- Two families from the early (mathematical) literature:
  - The “N,M" family (Erdös-Rényi, size/density CUG)
  - Let M_m be the maximum number of edges in G. Then:

$$\Pr(G = g \mid N, M) = \binom{M_m}{M}^{-1}$$

  - The “N,p" family (homogeneous Bernoulli graphs)

$$\Pr(G = g \mid N, p) = p^M (1-p)^{M_m - M}$$

  - Both used as baseline models, but very limited
    - No heterogeneity, (almost) no dependence
    - Starting point for more complex models

Relaxation 1: Intra-dyadic Independence

- First way to build richer models: relax independence
- Intra-dyadic dependence models
  - Allow for size, density, reciprocity effects
- Two parallel models
  - Dyad census conditioned CUG ($U|MAN$)
  - Let G have fixed dyad census M, A, N. Then

$$\Pr(G = g \mid M, A, N) = \frac{M!A!N!}{(M + A + N)!}$$

  - Homogeneous dyadic multinomial family ($u|man$)

$$\Pr(G = g \mid m, a, n) = m^M a^A n^N$$

Relaxation 2: Homogeneity

- Second way to build richer models: relax homogeneity in edge probabilities
- Development for Bernoulli, multinomial cases:
- Inhomogeneous Bernoulli graph
  - Let $\Phi \in [0,1]^{N \times N}$ be a parameter matrix, and $B(X = x \mid p)$ the Bernoulli pmf.

$$\Pr(G = g \mid \Phi) = \prod_{(i,j)} B(Y_{ij} = y_{ij} \mid \Phi_{ij})$$

  - Inhomogeneous independent dyad graph
  - Let $\Phi, \Psi \in [0,1]^{N \times N}$ be parameter matrices w/ $\Phi_{ij} + \Psi_{ij} \leq 1$. Then

$$\Pr(G = g \mid \Phi, \Psi) = \prod_{(i,j)} \left[\Phi_{ij} y_{ij} y_{ji} + \Psi_{ij}(y_{ij}(1 - y_{ij}) + (1 - y_{ij} y_{ij}) + (1 - \Phi_{ij} - \Psi_{ij})(1 - y_{ij})(1 - y_{ij})\right]$$

  - Intuitively, Φ sets the probability of mutuals, and Ψ sets the probability of asymmetrics

Simple Random Graph Models in Practice

- Models without trivial cross-dyadic dependence still have many uses
  - Baseline models for null hypothesis testing
  - Mathematical tools for exploring the space of graphs
  - Serious data models (in the inhomogeneous case)
    - Start with inhomogenous family, model parameter matrix using greression-like model (see, e.g., `sna::netlogit`) and/or with latent variables (e.g. package `latentnet`)
    - Can be extremely effective, if sufficiently strong covariates are available
  - Dyad dependent models are much more complex, but we'll see them later...

Cohesion and Subgroups

Introduction to Cohesion and Subgroups

- R Lab

References and Places for More Information i

Introduction to Statisticsl Inference

Cohesion and Subgroups