36-780: Social Network Modeling

ERGMs as models Brian Junker 132E Baker Hall brian@stat.cmu.edu

1/16/2014

Outline

- Special cases of ERGMs
 - Erdos-Renyi-Gilbert model
 - Beta model
 - \square p₁ model
- Exponential Random Graph Models (ERGMs)
 - Computational explosion
 - Independent-dyads approximation
 - □ MCMC-MLE
- ERGMs: The Good, The Bad, The R Example
- HW02, for Tuesday, is on line

The Erdos-Renyi-Gilbert Model

 G(N,p) specifies a model for a directed graph G on N nodes, with iid probability p of an edge y_{ij} between any two nodes i and j.

Then $\#E(G) = K \sim Binomial\left(\binom{N}{2}, p\right)$

• G(N,K) is the equivalent "hypergeometric-like" model, that conditions on the number of edges K, with p = $K/\binom{N}{2}$.

1/16/2014

The beta model for social networks

The E-R-G model assumes Y_{ii} iid with

$$\log \frac{P(Y_{ij})}{1 - P(Y_{ij})} \equiv \theta$$

The beta model assumes Y_{ii} independent with

$$\log \frac{P(Y_{ij})}{1 - P(Y_{ij})} = \beta_i + \beta_j$$

This allows for very simple variation in degree across nodes

1/16/2014

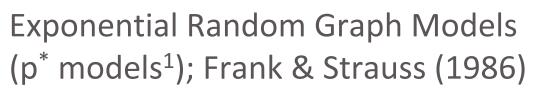
The p₁ model (Holland & Leinhardt 1981)

 Easiest to state in terms of the joint likelihood for the whole adjacency matrix Y:

 $\log P(Y=y) \propto \theta y_{++} + \sum_{i} \alpha_{i} y_{i+} + \sum_{j} \beta_{j} y_{+j} + \rho \sum_{ij} y_{ij} y_{ji}$

- A subscript + means "sum over that index"
- □ The sufficient statistics are the out-degrees y_{i+} and indegrees y_{+j} for each node, and the number of "reciprocal dyads" $\sum_{ij} y_{ij} y_{ji}$
- $\hfill\square$ θ is an overall tie propensity
- $\hfill\square$ α_i is node i's "gregariousness" or "expansiveness"
- □ β_j is node j's "attractiveness" or "popularity"
- \Box ρ is the tendency to reciprocate ties in the graph.

1/16/2014



In the p₁ model, the term

$$ho \sum_{ij} y_{ij} y_{ji}$$

counts a certain "topological feature" (reciprocal ties).

• Could replace $\sum_{ij} y_{ij} y_{ji}$ with a summary g(y) of some other topological feature(s), getting a new model term

$\theta^*g(y)$

 This may depend on more than one edge at a time, and hence the model no longer has independent dyads. E.g., g(y) = ∑_{i,j,k} y_{ij}y_{jk}y_{ki} = # of 3-cycles.

Exponential Random Graph Models (ERGMs)²

- This leads to a model of the form $\log P(Y = y) \propto \theta y_{++} + \sum_{i} \alpha_{i} y_{i+} + \sum_{j} \beta_{j} y_{+j} + \theta^{*} g(y)$
- No need to retain the p₁ terms, and we could have more than one g(y); this leads to

$$P(Y = y) = \frac{\exp\{\vec{\theta}^{\mathsf{T}} \vec{g}(y)\}}{\kappa(\vec{\theta}, \mathcal{Y})}$$

where $\vec{\theta}$ is a vector of parameters, $\vec{g}(y)$ is a vector of summary statistics, $\kappa(\vec{\theta}, \mathcal{Y})$ is a normalizing constant, summed over the set \mathcal{Y} of "all" networks on E(Y).

1/16/2014

²Following, more or less, Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3). [http://www.jstatsoft.org/v24/i03/]

ERGM computational explosion...

The normalizing constant

$$\kappa(\vec{\theta},\mathcal{Y}) = \sum_{y\in\mathcal{Y}} \exp\{\vec{\theta}^\mathsf{T} \vec{g}(y)\}$$

is a sum over all possible graphs in $\boldsymbol{\mathcal{Y}}$ and this can be a big number:

$$\mathsf{card}(\mathcal{Y}) = 2^{\binom{N}{2}}$$

- Usually infeasible to compute the MLE directly. Instead, we see
 - Independent-dyad approximation (Strauss & Ikeda, 1990)
 - MCMC-MLE (Snijders, 2002)

ERGM: Independent-dyads approx.

Let Y_{ij} indicate (1 or 0) an edge, let Y^c_{ij} be the rest of the graph, and let y⁺ and y⁻ be the graph with y_{ij}=1, and y_{ij}=0, respectively. Then

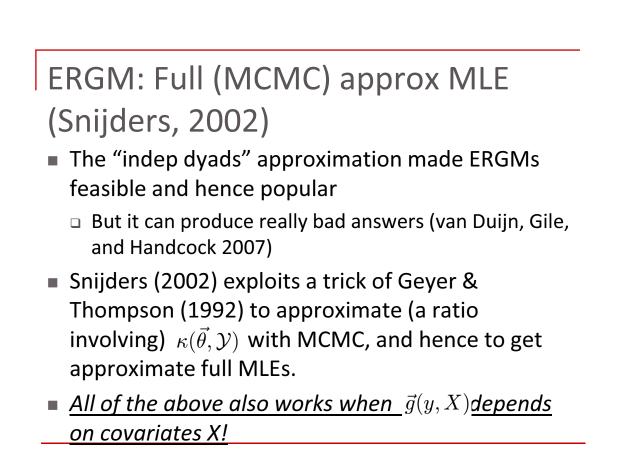
$$\log \frac{P[Y_{ij}=1|Y_{ij}^{c}=y_{ij}^{c}]}{P[Y_{ij}=0|Y_{ij}^{c}=y_{ij}^{c}]} = \vec{\theta}^{\mathsf{T}}[\vec{g}(y^{+}) - \vec{g}(y^{-})]$$

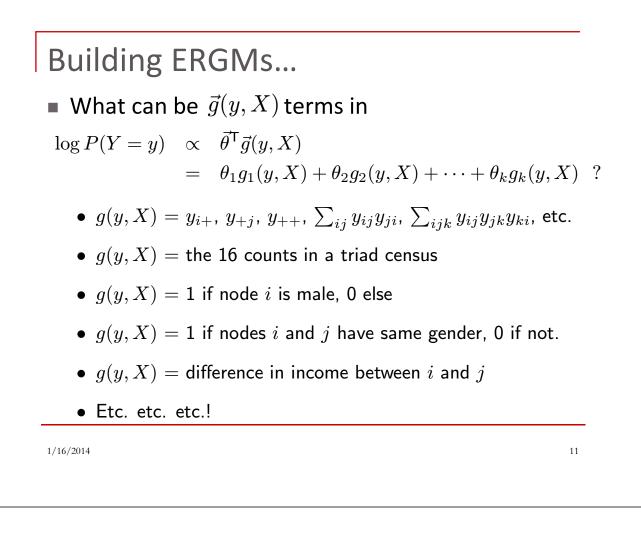
 If we now "pretend" that g
 ['](y) does not depend on Y^c_{ii}, we get an independent dyads model

$$\log \frac{P[Y_{ij}=1]}{P[Y_{ij}=0]} = \vec{\theta}^{\mathsf{T}}[\vec{g}(y_{ij}|_{y_{ij}=1}) - \vec{g}(y_{ij}|_{y_{ij}=0})]$$

that is easy to estimate (e.g., glm).

1/16/2014





ERGMs in R: The Good, The Bad, ...

- The above is the GOOD news about ERGMs.
- We will look at an example in R.
- You will also look at the "ergm" package in your hw.
- The BAD news about ERGMs will come next week (and, somewhat, in the R example...)

Outline

- Special cases of ERGMs
 - Erdos-Renyi-Gilbert model
 - Beta model
 - \square p₁ model
- Exponential Random Graph Models (ERGMs)
 - Computational explosion
 - Independent-dyads approximation
 - MCMC-MLE
- ERGMs: The Good, The Bad, The R Example
- HW02, for Tuesday, is on line

1/16/2014