Advanced ERG Parameterization: Curved Models and Constraints

SOC 280: Analysis of Social Network Data

Carter T. Butts
Department of Sociology and
Institute for Mathematical Behavioral Sciences
University of California, Irvine

The Final Chapter

- **In our previous episode, we discussed evaluation of network models within the ERG framework**
  - Diagnostics for burn-in, convergence, and degeneracy
  - Model adequacy checking via simulation
  - Model/algorithm refinement to improve fit/performance
- **Our last topic: more advanced issues in ERG parameterization**
  - Using “curved” exponential families to avoid degeneracy
  - Incorporating support constraints (e.g., on degree)
- **Much more can be done, but that's another course....**

A Surfeit of Statistics

- **Observation: many natural ERG parameterizations involve many statistics**
  - E.g., degree, *k*-stars – could have up to *N*-1 of them
- **This poses many problems**
  - Easy to wind up on convex hull boundary, MLE doesn't exist (e.g., including *i*th degree term when no vertices of degree *i* are present)
  - Finding MLE difficult (high-dimensional optimization is hard)
  - Models not parsimonious (who wants dozens of terms?)
  - Often unrealistic: usually assume that terms like degree will be related to one another in a smooth way
    - E.g., the degree distribution models we saw earlier

Throwing a Curve

- **One solution to this problem: *curved* exponential families**
  - Alternate way to parameterize exponential family models
  - Starts with normal (regular) exponential family
  - Adds a "curved constraint" by positing a parametric form for some of or all of the natural parameters
    - Generally, the new form has fewer parameters than the old one (never more)
- **Familiar example from everyday usage:**
  - There are N+1 possible counts of "heads" from N coin flips (0,1,...N); could assign a parameter to every outcome (w/1 linear constraint), whereas a binomial distribution only uses a single parameter (not counting N itself)

ERGs in Curved Exponential Family Form

- **For adjacency matrix $\mathbf{Y}$ w/countable support $\mathcal{Y}$ and covariate set X, pmf can be written in *curved* ERG form:**

$$\Pr\left(\mathbf{Y}=\mathbf{y}\middle|\mathbf{t},\theta,\eta,\mathcal{Y},X\right)=\frac{\exp\left(\eta\left(\theta\right)^{T}\mathbf{t}\left(\mathbf{y},X\right)\right)}{\sum_{\mathbf{y}'\in\mathcal{Y}}\exp\left(\eta\left(\theta\right)^{T}\mathbf{t}\left(\mathbf{y}',X\right)\right)}I_{\mathcal{Y}}\left(\mathbf{y}\right)$$

  - $\eta(\theta)$: *canonical* parameters
    - $\theta\in\mathbb{R}^n$: vector of *curved* parameters
    - $\eta:\mathbb{R}^n\rightarrow\mathbb{R}^m$: mapping from curved to canonical parameters
- **Intuition: curved parameterization forces canonical parameters to lie on a lower-dimensional surface**
  - Fewer curved than canonical parameters implies many combinations of canonical parameters impossible
  - Allows one to impose parametric form on sets of statistics

Example: Curved Families for Degree and Shared Partners

- **Degree/star parameters are a natural candidate for curved models**
  - Too many of 'em
  - We expect them to follow a smooth form
- **Two similar examples**
  - Alternating k-stars
    - ith degree parameter becomes proportional to $\exp(-\alpha i)$
  - Geometrically weighted degree
    - ith degree parameter effectively proportional to $-\lambda^2(1-(1-1/\lambda)^i) = \gamma^i - \beta$
- **Another example: shared partner statistics**
  - Two kinds: edgewise shared partners, dyadwise shared partners
  - Alternative to triangles for clustering
  - Can avoid runaway clique formation
- **Both can be used with geometric weighting**
  - Again, have ith parameter declining as some γ^i
    - Intuitively, declining marginal tendency to form shared partners

Incorporating Constraints

- **Recall that the definition of the ERG explicitly incorporated a “support term”**
  - Set of all graphs that are realizable
- **Usually suppress in our notation – take as assumed**
  - Standard assumption: all loopless graphs/digraphs on N vertices are possible
- **In some cases, can't take support for granted**
  - Models with special classes of graphs (e.g., entailments, trees, hierarchies, etc.)
  - Constraints arising from data collection or other factors

Common Constraint Cases

- **Common practical issue: network data collection imposes constraints on what we can observe**
- **Examples:**
  - Outdegree constraints: nominate exactly *n* alters; nominate up to *n* alters
  - Total degree constraints: isolates unobserved (e.g., edge sampling)
  - Mixing constraints: no ties between groups *i* and *j* (e.g., heterosexual contact networks); nominate up to *n* alters in each group (position generators, AddHealth)
  - Observed edge constraints: certain edges present by design (e.g., examining only alters with known ties to ego, interorganizational edges mandated by law)

Constraints in `ergm`

- **`ergm` supports several types of constraints**
- **Syntax: argument `constraints=~...`**
  - Options (i.e., "`...`"):
    - `.`: no constraints (default)
    - `edges`: preserve the number of edges
    - `degrees`: preserve the observed degrees
    - `degreedist`: preserve the observed degree distribution
    - `indegreedist`/`outdegreedist`: preserve indegree/outdegree distribution
    - `bd(args)`: place bounds on degree (see help for arguments)
    - `observed`: preserve observed edges (but not non-edges)

Final Exam

- **Will be posted electronically on Monday, 6/8**
- **Will be due midnight on Sunday, 6/15**
  - Turn in via EEE Drop Box
- **Content will be similar to homework**
  - Perform data analysis, interpret results
- **Ground rules**
  - You may use your notes, books, articles, software, etc. for reference
  - You may not discuss the exam with others (except me)
    - If questions arise, please don't hesitate to email me

Some Closing Comments

- **We've come a long way in this class – from basic visualization to ERG models!**
- **Much more that can be (and is) done in this fast-growing field**
  - To learn more, consult journals like *Social Networks, JMS, CMOT, Sociological Methodology,* and the *Journal of Social Structure*
- **Many ways to get involved**
  - Keep an eye out for more classes in Sociology and elsewhere
  - Consider joining INSNA, and attending the annual Sunbelt social network conference
  - Come to the UCI Social Network Research Group meetings next fall!
- **Good luck with your research – and remember, never underestimate the power of relationships!**