Parameterizing Exponential Family Models for Random Graphs: Current Methods and New Directions

Carter T. Butts

Department of Sociology and

Institute for Mathematical Behavioral Sciences

University of California, Irvine
buttsc@uci.edu

Prepared for the 2008 SIAM Conference, San Diego, CA, 7/10/08.

This work was supported in part by NIH award 5 R01 DA012831-05.

Stochastic Models for Social (and Other) Networks

- General problem: need to model graphs with varying properties
- Many *ad hoc* approaches:
  - Conditional uniform graphs (Erdös and Rényi, 1960)
  - Bernoulli/independent dyad models (Holland and Leinhardt, 1981)
  - Biased nets (Rapoport, 1949a;b; 1950)
  - Preferential attachment models (Simon, 1955; Barabási and Albert, 1999)
  - Geometric random graphs (Hoff et al., 2002)
  - Agent-based/behavioral models (including "classics" like Heider (1958); Harary (1953))
- A more general scheme: discrete exponential family models (ERGs)
  - General, powerful, leverages existing statistical theory (e.g., Barndorff-Nielsen (1978); Brown (1986); Strauss (1986))
  - (Fairly) well-developed simulation, inferential methods (e.g., Snijders (2002); Hunter and Handcock (2006))

Basic Notation

- Assume $G = (V, E)$ to be the graph formed by edge set E on vertex set V
  - Here, we take $|V| = N$ to be fixed, and assume elements of V to be uniquely identified
  - If $E \subseteq \big\{\{v, v'\} : v, v' \in V\big\}$, G is said to be *undirected*; G is *directed* iff $E \subseteq \big\{(v, v') : v, v' \in V\big\}$
  - $\{v, v\}$ or (v, v) edges are known as *loops*; if G is defined per the above and contains no loops, G is said to be *simple*
    - Note that multiple edges are already banned, unless E is allowed to be a multiset
- Other useful bits
  - E may be random, in which case $G = (V, E)$ is a *random graph*
  - Adjacency matrix $\mathbf{Y} \in \{0, 1\}^{N \times N}$ (may also be random); for G random, will usually use notation $\mathbf{y}$ for adjacency matrix of realization g of G

Exponential Families for Random Graphs

- For random graph G w/countable support $\mathcal{G}$, pmf is given in ERG form by

$$\Pr(G = g|\theta) = \frac{\exp\left(\theta^T \mathbf{t}(g)\right)}{\sum_{g' \in \mathcal{G}} \exp\left(\theta^T \mathbf{t}(g')\right)} I_{\mathcal{G}}(g) \tag{1}$$

- $\theta^T \mathbf{t}$: linear predictor
  - $\mathbf{t} : \mathcal{G} \to \mathbb{R}^m$: vector of sufficient statistics
  - $\theta \in \mathbb{R}^m$: vector of parameters
  - $\sum_{g' \in \mathcal{G}} \exp\left(\theta^T \mathbf{t}(g')\right)$: normalizing factor (aka partition function, Z)
- Intuition: ERG places more/less weight on structures with certain features, as determined by $\mathbf{t}$ and θ
  - Model is complete for pmfs on $\mathcal{G}$, few constraints on $\mathbf{t}$

Dependence Graphs and ERGs

- Let $\mathbf{Y}$ be the adjacency matrix of G
  - $Y_{ij} = 1$ if $(i, j) \in E$ and $Y_{ij} = 0$ otherwise
  - $\mathbf{Y}^c_{ab,cd,\ldots}$ denotes cells of $\mathbf{Y}$ not corresponding to pairs $(a, b), (c, d), \ldots$
- $D = (\mathcal{E}, E')$ is the conditional dependence graph of G
  - $\mathcal{E} = \{(i, j) : i \neq j, i, j \in V\}$: collection of edge variables
  - $\{(i, j), (k, l)\} \in E'$ iff $Y_{ij} \not\perp Y_{kl} | \mathbf{Y}^c_{ij,kl}$
- From D to G: the Hammersley-Clifford Theorem (Besag, 1974)
  - Let K_D be the clique set of D. Then in the ERG case,

$$\Pr(G = g|\theta) = \frac{1}{Z(\theta, \mathcal{G})} \exp\left(\sum_{S \in K_D} \theta_S \prod_{(i,j) \in S} y_{ij} \right) \tag{2}$$

  - If homogeneity constraints imposed, then sufficient statistics are counts of subgraphs of G isomorphic to subgraphs forming cliques in D

Model Construction Using Dependence Graphs

- Hammersley-Clifford allows us to specify random graph models which satisfy particular edge dependence conditions
- Simple examples (directed case):
  - Independent edges: $Y_{ij} \not\perp Y_{kl}|\mathbf{Y}^c_{ij,kl}$ iff $(i,j)=(k,l)$
    - D is the null graph on $\mathcal{E}$; thus, the only cliques are the nodes of D themselves (which are the edge variables of G)
    - From this, H-C gives us $\Pr(G=g|\theta) \propto \exp\left(\sum_{(v_i,v_j)} \theta_{ij} y_{ij}\right)$, which is the inhomogeneous Bernoulli graph with $\theta_{ij} = \mathrm{logit}\Phi_{ij}$
    - Assuming homogeneity, this becomes $\Pr(G=g|\theta) \propto \exp\left(\theta \sum_{(v_i,v_j)} y_{ij}\right)$, which is the N,p model – note that $|E|$ is the unique sufficient statistic!

Model Construction Using Dependence Graphs, Cont.

- Examples (cont.):
  - Independent dyads: $Y_{ij} \not\perp Y_{kl} | \mathbf{Y}^c_{ij,kl}$ iff $\{i,j\} = \{k,l\}$
    - D is a union of K_2s, each corresponding to an $\{(i,j),(j,i)\}$ pair; thus, each dyad of G contributes a clique, as does each edge (remember, nested cliques count)
    - H-C gives us $\Pr(G=g|\theta,\theta') \propto \exp\left(\sum_{\{v_i,v_j\}} \theta_{ij} y_{ij} y_{ji} + \sum_{(v_i,v_j)} \theta'_{ij} y_{ij}\right)$; this is the inhomogeneous independent dyad model with $\theta = \ln \frac{2mn}{a^2}$ and $\theta' = \ln \frac{a}{2n}$
    - As before, we can impose homogeneity to obtain $\Pr(G=g|\theta,\theta') \propto \exp\left(\theta \sum_{\{v_i,v_j\}} y_{ij} y_{ji} + \theta' \sum_{(v_i,v_j)} y_{ij}\right)$, which is the $u|man$ model with sufficient statistics M and $2M+A$

A More Complex Example: The Markov Graphs

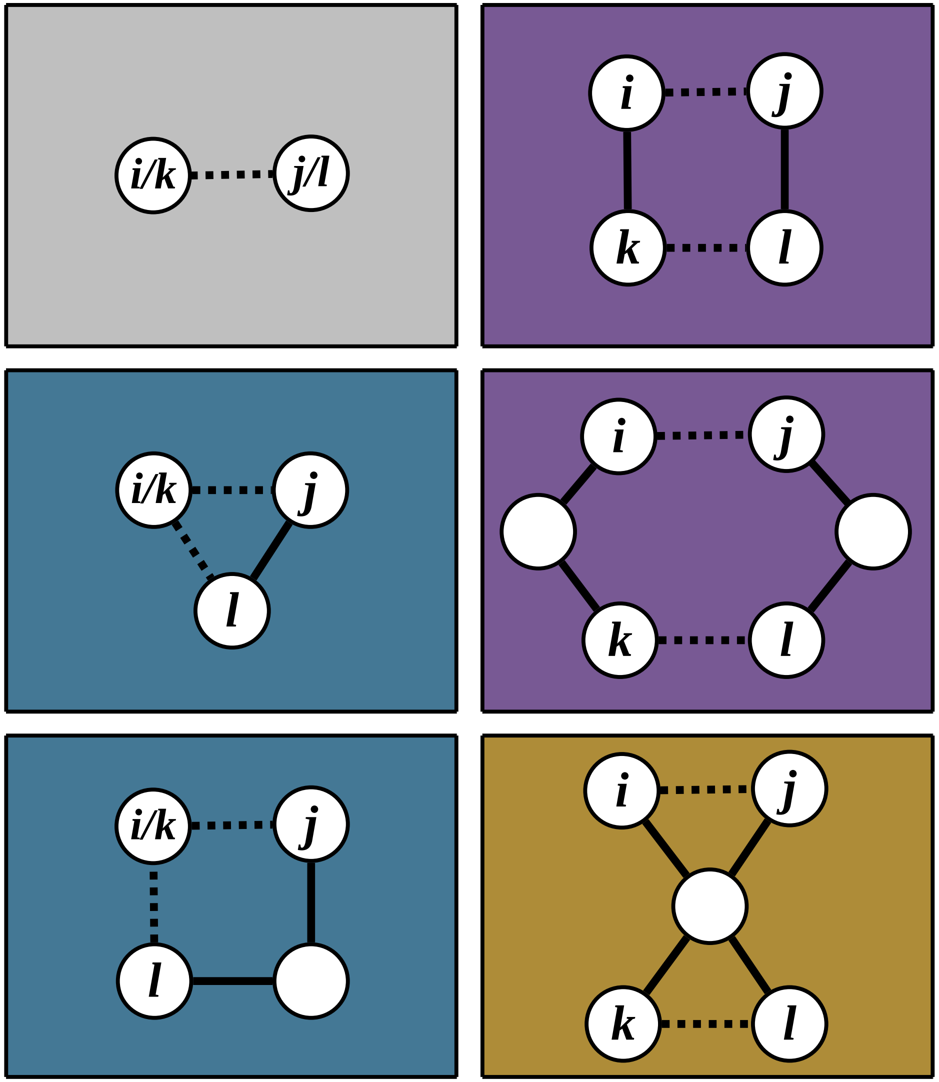
- An important advance by (Frank and Strauss, 1986): the Markov graphs
- The basic definition: $Y_{ij} \not\perp Y_{kl} | \mathbf{Y}^c_{ij,kl}$ iff $|\{i,j\} \cap \{k,l\}| > 0$
  - Intuitively, edge variables are conditionally dependent iff they share at least one endpoint
  - D now has a large number of cliques; these are the edge variables, stars, and triangles of G
    - In undirected case, sufficient statistics are the k-stars and triangles of G (or counts thereof, if homogeneity is assumed)
    - In directed case, sufficient statistics are in/out/mixed k-stars and the full triangle census of G (minus the superfluous null triad)
- Markov graphs capture many important structural phenomena
  - Trivially, includes density and (in directed case) reciprocity
  - k-stars equivalent to degree count statistics, hence includes degree distribution (and mixing, in directed case)
  - Through triads, includes local clustering as well as local cyclicity and transitivity in digraphs
- The downside: hard to work with, prone to poor behavior – but, nothing's free....

Beyond the Markov Graphs: Partial Conditional Dependence

- Bad news: Hammersley-Clifford doesn't help much for long-range dependence
  - In general, D becomes a complete graph – all subsets of edges generate potential sufficient statistics
- Alternate route: partial conditional dependence models
  - Based on Pattison and Robins (2002): $Y_{ij} \not\perp Y_{kl} | \mathbf{Y}^c_{ij,kl}$ only if some condition is satisfied (e.g., $\mathbf{y}^c_{ij}$ belongs to some set C)
  - Lead to sufficient statistics which are subset of H-C stats
- Example: *reciprocal path dependence* (Butts, 2006)
  - Assume edges independent unless endpoints joined by (appropriately directed) paths

Reciprocal Path Conditions

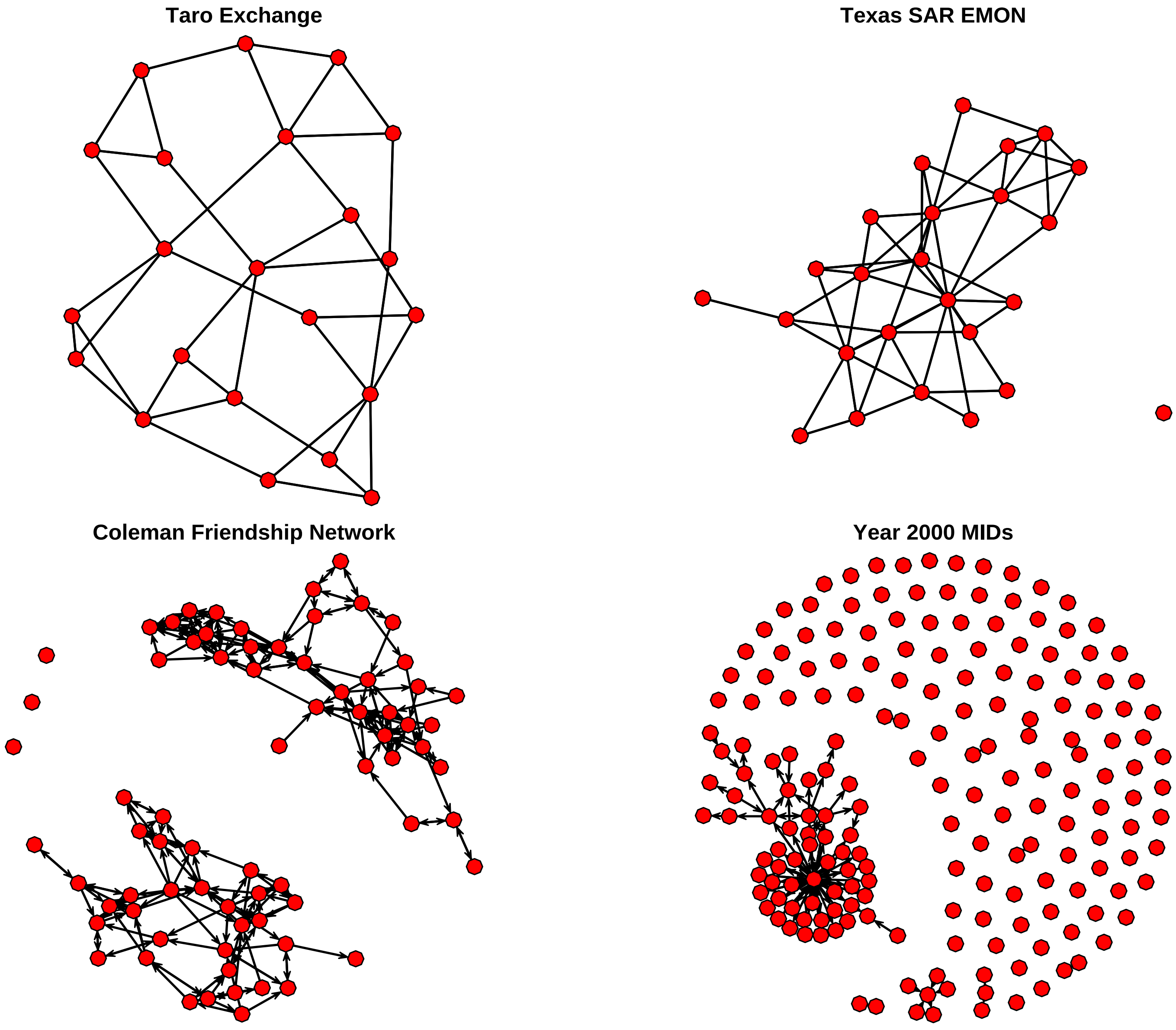
- Basic idea: head of each edge can reach the tail of the other
  - Weak case: (directed) paths each way are sufficient
  - Strong case: paths cannot share internal vertices
- Intuition: *extended reciprocity*
  - Possibility of feedback through network
  - In strong case, channels of reciprocation share no intermediaries



Reciprocal Path Dependence Models

- Define $aRb \equiv$"a and b satisfy the reciprocal path condition"
  - Negation written as $a\overline{R}b$
  - $aRb \Leftrightarrow bRa$, $a\overline{R}b \Leftrightarrow b\overline{R}a$
- Theorem: Let $\mathbf{Y}$ be a random adjacency matrix whose pmf is a discrete exponential family satisfying a reciprocal path dependence assumption under condition R. Then the sufficient statistics for $\mathbf{Y}$ are functions of edge sets S such that $(i,j)R(k,l)\ \forall\ \{(i,j),(k,l)\} \subseteq S$.
- Sufficient statistics under reciprocal path dependence, homogeneity:
  - Strong, directed: cycles
  - Weak, directed: cycles, certain unions of cycles
  - Strong, undirected: subgraphs w/spanning cycles
  - Weak, directed: subgraphs w/spanning cycles, some unions thereof

Application to Sample Networks



Cycle Census ERG Fits

	Taro Exchange			Texas EMON		
	$\hat{\theta}$	s.e.	$\Pr(> \lvert Z \rvert)$	$\hat{\theta}$	s.e.	$\Pr(> \lvert Z \rvert)$
Edges	2.0526	1.4914	0.1687	**−2.5933**	**0.4064**	**0.0000**
Cycle3	1.1489	1.0175	0.2588	**2.6117**	**0.9033**	**0.0038**
Cycle4	**−2.1619**	**0.8713**	**0.0131**	−0.7302	0.5911	0.2167
Cycle5	−0.0789	0.6297	0.9003	0.1765	0.2081	0.3964
Cycle6	−0.4999	0.2772	0.0714	−0.0300	0.0316	0.3423
	ND 320.234; RD 56.112 on 226 df			ND 415.89; RD 97.14 on 295 df		

	Friendship			MIDs		
	$\hat{\theta}$	s.e.	$\Pr(> \lvert Z \rvert)$	$\hat{\theta}$	s.e.	$\Pr(> \lvert Z \rvert)$
Edges	**−4.1778**	**0.0957**	**0.0000**	**−6.9336**	**0.3406**	**0.0000**
Cycle2	**1.5615**	**0.2082**	**0.0000**	**7.8360**	**2.4368**	**0.0013**
Cycle3	**0.7222**	**0.2092**	**0.0006**	**−3.0203**	**0.7638**	**0.0001**
Cycle4	**0.6866**	**0.1819**	**0.0002**	**43.3479**	**0.0188**	**0.0000**
Cycle5	0.1663	0.1062	0.1173	**−1.9328**	**0.0029**	**0.0000**
Cycle6	−0.0063	0.0334	0.8508			
	ND 7286.4; RD 1384.4 on 5256 df			ND 50308.62; RD 988.48 on 36285 df		

A New Direction: Potential Games

- So far, our focus has been on *dependence hypotheses*
  - Define the conditions under which one relationship could affect another, and hope that this is sufficiently reductive
  - Complete agnosticism regarding underlying mechanisms – could be social dynamics, unobserved heterogeneity, or secret closet monsters
- A choice-theoretic alternative?
  - In some cases, reasonable to posit actors with some control over edges (e.g., out-ties)
  - Existing theory often suggests general form for utility
  - Reasonable behavioral models available (e.g., multinomial choice)
- The link between choice models and ERGs: *potential games*
  - Increasingly wide use in economics, engineering
  - Equilibrium behavior provides an alternative way to parameterize ERGs

Potential Games and Network Formation Games

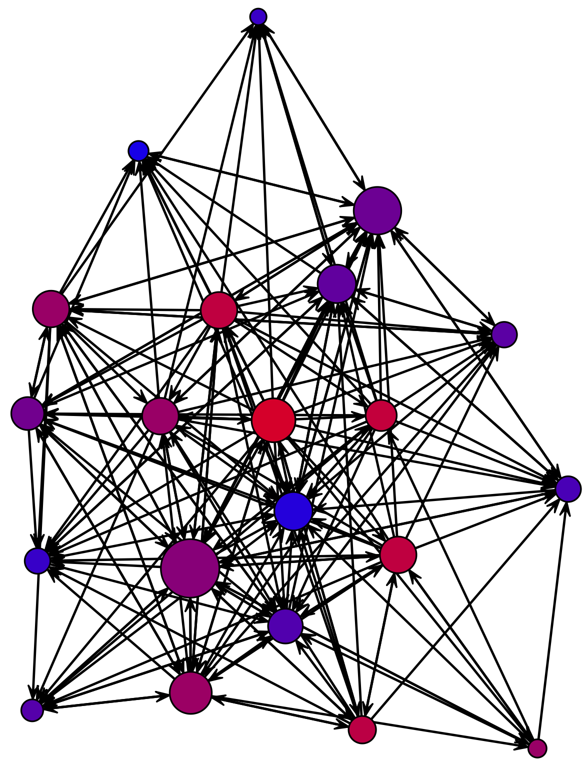
- Potential games (Monderer and Shapley, 1996)
  - Let X by a strategy set, u a vector utility functions, and V a set of players. Then (V, X, u) is said to be a *potential game* if $\exists\, \rho : X \mapsto \mathbb{R}$ such that $u_i\left(x'_i, x_{-i}\right) - u_i\left(x_i, x_{-i}\right) = \rho\left(x'_i, x_{-i}\right) - \rho\left(x_i, x_{-i}\right)\ \forall i \in V, x, x' \in X$.
- Consider a simple family of *network formation games* (Jackson, 2006) on $\mathcal{Y}$:
  - Each i, j element of $\mathbf{Y}$ is controlled by a single player $k \in V$ with finite utility u_k; can choose $y_{ij} = 1$ or $y_{ij} = 0$ when given an "updating opportunity"
    - We will here assume that i controls $\mathbf{Y}_{i\cdot}$, but this is not necessary
  - Theorem: Let (i) $(V, \mathcal{Y}, u)$ in the above form a game with potential ρ; (ii) players choose actions via a logistic choice rule; and (iii) updating opportunities arise sequentially such that every (i, j) is selected with positive probability, and (i, j) is selected independently of the current state of $\mathbf{Y}$. Then $\mathbf{Y}$ forms a Markov chain with equilibrium distribution $\Pr(\mathbf{Y} = \mathbf{y}) \propto \exp(\rho(\mathbf{y}))$, in the limit of updating opportunities.
- One can thus obtain an ERG as the long-run behavior of a strategic process, and parameterize in terms of the hypothetical underlying utility functions

Various Utility/Potential Components

- Edge payoffs (homogeneous)
  - $u_i(\mathbf{y}) = \theta \sum_j y_{ij}$
  - $\rho(\mathbf{y}) = \theta \sum_i \sum_j y_{ij}$
- Edge payoffs (inhomogeneous)
  - $u_i(\mathbf{y}) = \theta_i \sum_j y_{ij}$
  - $\rho(\mathbf{y}) = \sum_i \theta_i \sum_j y_{ij}$
- Edge covariate payoffs
  - $u_i(\mathbf{y}) = \theta \sum_j y_{ij} x_{ij}$
  - $\rho(\mathbf{y}) = \theta \sum_i \sum_j y_{ij} x_{ij}$
- Reciprocity payoffs
  - $u_i(\mathbf{y}) = \theta \sum_j y_{ij} y_{ji}$
  - $\rho(\mathbf{y}) = \theta \sum_i \sum_{j<i} y_{ij} y_{ji}$
- 3-Cycle payoffs
  - $u_i(\mathbf{y}) = \theta \sum_{j \neq i} \sum_{k \neq i,j} y_{ij} y_{jk} y_{ki}$
  - $\rho(\mathbf{y}) = \frac{\theta}{3} \sum_i \sum_{j \neq i} \sum_{k \neq i,j} y_{ij} y_{jk} y_{ki}$
- Transitive completion payoffs
  - $u_i(\mathbf{y}) =$ $\theta \sum_{j \neq i} \sum_{k \neq i,j} \left[\begin{array}{c} y_{ij} y_{ki} y_{kj} + y_{ij} y_{ik} y_{jk} \\ + y_{ij} y_{ik} y_{kj} \end{array} \right]$
  - $\rho(\mathbf{y}) = \theta \sum_i \sum_{j \neq i} \sum_{k \neq i,j} y_{ij} y_{ik} y_{kj}$
- And many more! (But caveats apply...)
  - Not all reasonable u lead to potential games – e.g., 2-path and shared partner effects cannot be separated
  - Not all heterogeneity can be modeled (e.g., individual-specific reciprocity payoffs)

Empirical Example: Advice-Seeking Among Managers

- Sample empirical application from Krackhardt (1987): self-reported advice-seeking among 21 managers in a high-tech firm
  - Additional covariates: friendship, authority (reporting)
- Demonstration: selection of potential behavioral mechanisms via ERGs
  - Models parameterized using utility components
  - Model parameters estimated using maximum likelihood (Geyer-Thompson)
  - Model selection via AIC



Advice-Seeking ERG – Model Comparison

- First cut: models with independent dyads:

	Deviance	Model df	AIC	Rank
Edges	578.43	1	580.43	7
Edges+Sender	441.12	21	483.12	4
Edges+Covar	548.15	3	554.15	5
Edges+Recip	577.79	2	581.79	8
Edges+Sender+Covar	385.88	23	431.88	2
Edges+Sender+Recip	405.38	22	449.38	3
Edges+Covar+Recip	547.82	4	555.82	6
Edges+Sender+Covar+Recip	**378.95**	**24**	**426.95**	**1**

- Elaboration: models with triadic dependence:

	Deviance	Model df	AIC	Rank
Edges+Sender+Covar+Recip	378.95	24	426.95	4
Edges+Sender+Covar+Recip+CycTriple	361.61	25	411.61	2
Edges+Sender+Covar+Recip+TransTriple	368.81	25	418.81	3
Edges+Sender+Covar+Recip+CycTriple+TransTriple	**358.73**	**26**	**410.73**	**1**

- Verdict: data supplies evidence for heterogeneous edge formation preferences (w/covariates), with additional effects for reciprocated, cycle-completing, and transitive-completing edges.

Advice-Seeking ERG – AIC Selected Model

Effect	$\hat{\theta}$	s.e.	$\Pr(>\|Z\|)$		Effect	$\hat{\theta}$	s.e.	$\Pr(>\|Z\|)$	
Edges	**−1.022**	**0.137**	**0.0000**	$***$	**Sender14**	**−1.513**	**0.231**	**0.0000**	$***$
Sender2	**−2.039**	**0.637**	**0.0014**	$**$	**Sender15**	**16.605**	**0.336**	**0.0000**	$***$
Sender3	0.690	0.466	0.1382		**Sender16**	**−1.472**	**0.232**	**0.0000**	$***$
Sender4	−0.049	0.441	0.9112		**Sender17**	**−2.548**	**0.197**	**0.0000**	$***$
Sender5	0.355	0.495	0.4734		**Sender18**	**1.383**	**0.214**	**0.0000**	$***$
Sender6	**−4.654**	**1.540**	**0.0025**	$**$	**Sender19**	**−0.601**	**0.190**	**0.0016**	$**$
Sender7	−0.108	0.375	0.7726		Sender20	0.136	0.161	0.3986	
Sender8	−0.449	0.479	0.3486		Sender21	0.105	0.210	0.6157	
Sender9	0.393	0.496	0.4281		**Reciprocity**	**0.885**	**0.081**	**0.0000**	$***$
Sender10	0.023	0.555	0.9662		**Edgecov (Reporting)**	**5.178**	**0.947**	**0.0000**	$***$
Sender11	**−2.864**	**0.721**	**0.0001**	$***$	**Edgecov (Friendship)**	**1.642**	**0.132**	**0.0000**	$***$
Sender12	**−2.736**	**0.331**	**0.0000**	$***$	**CycTriple**	**−0.216**	**0.013**	**0.0000**	$***$
Sender13	**−0.986**	**0.194**	**0.0000**	$***$	**TransTriple**	**0.090**	**0.003**	**0.0000**	$***$

Null Dev 582.24; Res Dev 358.73 on 394 df

- Some observations...
  - Arbitrary edges are costly for most actors
  - Edges to friends and superiors are “cheaper” (or even positive payoff)
  - Reciprocating edges, edges with transitive completion are cheaper...
  - ...but edges which create (in)cycles are more expensive; a sign of hierarchy?

Conclusion

- Models for complex networks pose complex problems of parameterization
  - Many ways to describe dependence among elements
  - Once one leaves simple cases, not always clear where to begin
- Three basic approaches for ERG parameterization
  - "Straight" Hammersley-Clifford (conditional dependence)
  - Partial conditional dependence
  - Potential games
- We've come a long way, but many open problems remain
  - "Inverse" conditional/partial conditional dependence: given a graph statistic, what dependence conditions give rise to it?
  - More reductive partial conditional dependence conditions
  - Generalizations of the potential game result

1 References

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 206:509–512.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley and Sons, New York.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*. Institute of Mathematical Statistics, Hayward, CA.

Butts, C. T. (2006). Cycle census statistics for exponential random graph models. IMBS Technical Report MBS 06-05, Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA.

Erdös, P. and Rényi, A. (1960). On the evolution of random graphs. *Public Mathematical Institute of Hungary Academy of Sciences*, 5:17–61.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81:832–842.

Harary, F. (1953). On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 3:37–41.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley and Sons, New York.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76(373):33–50.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15:565–583.

Jackson, M. (2006). A survey of models of network formation: Stability and efficiency. In Demange, G. and Wooders, M., editors, *Group Formation Economics: Networks, Clubs, and Coalitions*. Cambridge University Press, Cambridge.

Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9(2):109–134.

Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, 14:124–143.

Pattison, P. and Robins, G. (2002). Neighborhood-based models for social networks. *Sociological Methodology*, 32:301–337.

Rapoport, A. (1949a). Outline of a probabilistic approach to animal sociology I. *Bulletin of Mathematical Biophysics*, 11:183–196.

Rapoport, A. (1949b). Outline of a probabilistic approach to animal sociology II. *Bulletin of Mathematical Biophysics*, 11:273–281.

Rapoport, A. (1950). Outline of a probabilistic approach to animal sociology III. *Bulletin of Mathematical Biophysics*, 12:7–17.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42:425–440.

Snijders, T. A. B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2).

Strauss, D. (1986). On a General Class of Models for Interaction. *SIAM Review*, 28(4):513–527.