# Modeling Reality Without Sacrificing Data: Inferentially Tractable Models for Complex Social Systems

**Carter T. Butts**

Department of Sociology and

Institute for Mathematical Behavioral Sciences

University of California, Irvine

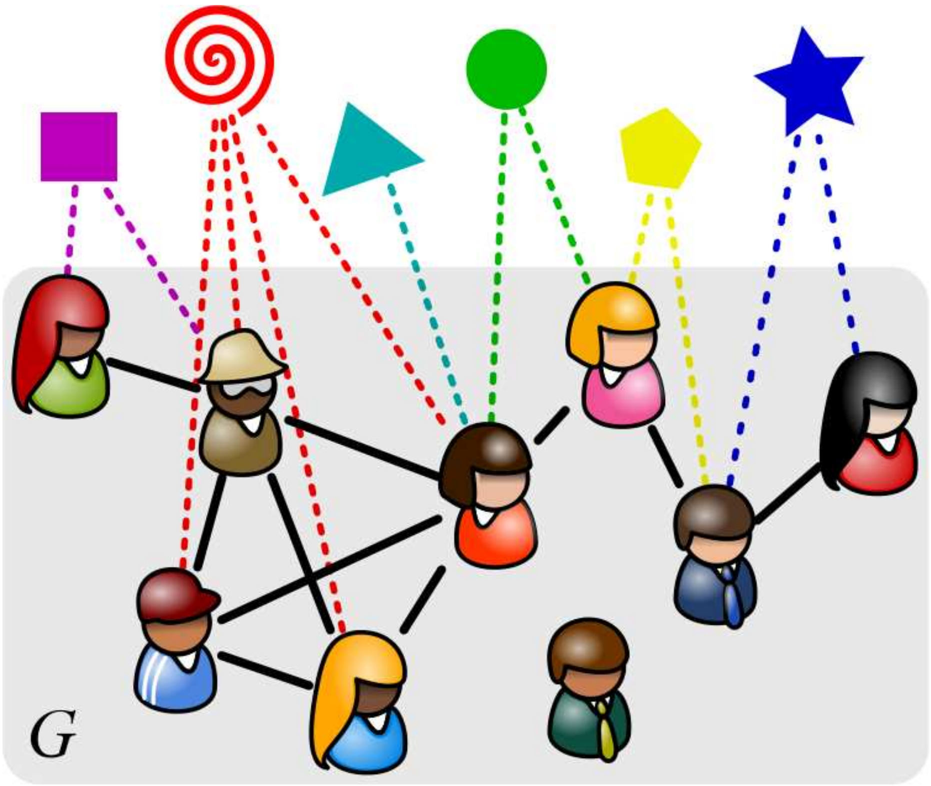
buttsc@uci.edu

Prepared for the 2008 *International Workshop on Challenges and Visions in the Social Sciences*, ETH Zürich.

This work was supported in part by NIH award 5 R01 DA012831-05 and NSF awards IIS-0031707 and CMS-0624257.

# The Problem of Complex Dependence

- Many human systems exhibit complex patterns of dependence
  - Nontrivial coupling among system elements
  - Particularly true within relational systems (i.e., social networks)
- A methodological and theoretical challenge
  - How to capture dependence without losing inferential tractability?
  - Not a new problem: also faced, e.g., by researchers in statistical physics

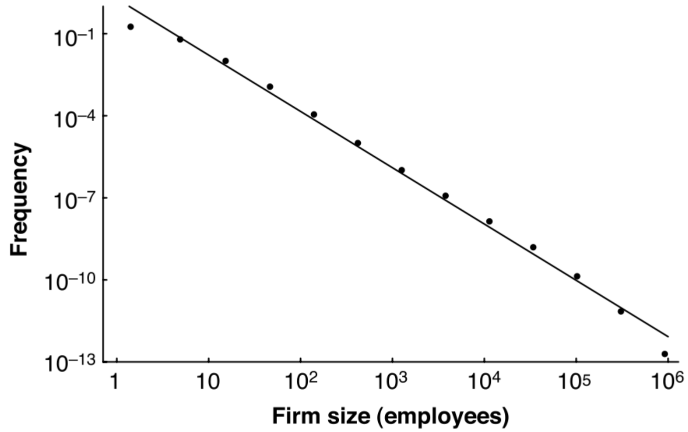


# Challenge: Modeling Reality without Sacrificing Data

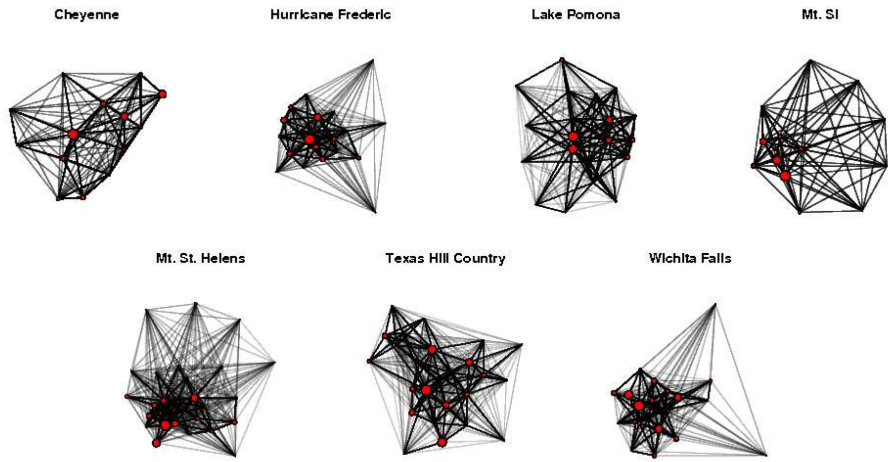
- How do we work with models which have non-trivial dependence?
- Can compare behavior of dependent-process models against stylized facts, but this has limits....
  - Not all models lead to clean/simple conditional or marginal relationships
  - Often impossible to disentangle nonlinearly interacting mechanisms on this basis
  - Very data inefficient: throws away much of the information content
  - Often need (very) large data sets to get sufficient power (which may not exist)
    - Collection of massive data sets often prohibitively costly
    - Many systems of interest *are* size-limited; studying only large systems leads to sampling bias
- Ideally, would like a framework which allows principled inference/model comparison without sacrificing (much) data

# Aside: Why "Small" Cases Can't Be Ignored

- Tempting to ignore "small" systems, but this is unwise
- Most groups, organizations relatively small (e.g., Axtell (2001); Mayhew et al. (1995))
  - Mean US firm size in 1997 apx 20 (5 w/self-employment)
  - Median individual exposure $<$150 (shades of Dunbar?)
- Many important phenomena involve few actors
  - E.g., Drabek et al. (1981) SAR EMONs: mean size 20 (similar for Lind et al.'s Katrina EMONs)
- Bottom line: methods should ideally work on "small" systems as well as large




(Figure 1, Axtell (2001))




(Drabek et al. SAR EMONs)

# Today's Case: 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 (many at this meeting(!), plus "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 $E \subseteq \left\{\{v, v'\} : v, v' \in V\right\}$, $G$ is said to be *undirected*; $G$ is *directed* iff $E \subseteq \left\{(v, v') : v, v' \in V\right\}$
  - $\{v, v\}$ 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 $G$

# Classical Examples

- Two familiar classical examples (now attributed to Erdös and Rényi):
  - Let $M_m = \binom{N}{2}$ (undirected case) or $M_m = 2\binom{N}{2}$ (directed case) be the maximum number of edges in simple random graph $G = (V, E)$
  - The $N, M$ model (size/density conditioned graph): $\Pr(G = g|N, M) = \binom{M_m}{M}^{-1}$
  - The $N, p$ model (aka homogeneous Bernoulli family):

    $\Pr(G = g|N, p) = p^M (1 - p)^{M_m - M}$
- Both models now synonymous with term “random graph,” much theory available (e.g., Bollobás (2001))
- Still frequently used as baseline models, but have many limitations
  - No heterogeneity in marginal edge probabilities (all ties are equiprobable)
  - No edgewise dependence in $N, p$ (almost none in $N, M$)

# Simple Relaxations

- Two ways to produce more interesting models: relax independence, or relax homogeneity
- Intra-dyadic dependence
  - Let $(M, A, N)$ be number of mutual, asymmetric, and null dyads in $g$
  - $U|MAN$ (dyad census conditioned graph): $\Pr(G = g|M, A, N) = \frac{M!A!N!}{(M+A+N)!}$, for $g$ with fixed dyad census $(M, A, N)$
  - $u|man$ (dyadic multinomial family): $\Pr(G = g|m, a, n) = m^M a^A n^N$ for $g$ with dyad census $(M, A, N) : M + A + N = \binom{|V|}{2}$
- Inhomogeneous families
  - Inhomogeneous Bernoulli graphs: $\Pr(G = g|\mathbf{\Phi}) = \prod_{\{v_i,v_j\}} B(Y_{ij} = y_{ij}|\Phi_{ij})$ (undirected) or $\Pr(G = g|\mathbf{\Phi}) = \prod_{(v_i,v_j)} B(Y_{ij} = y_{ij}|\Phi_{ij})$ (directed)
  - Inhomogeneous independent dyad graphs:

$$\Pr(G = g|\mathbf{\Phi}, \mathbf{\Psi}) = \prod_{\{v_i,v_j\}} \left[ \begin{array}{c} \Phi_{ij} y_{ij} y_{ji} + \Psi_{ij} \left(y_{ij} \left(1 - y_{ji}\right) + \left(1 - y_{ij}\right) y_{ji}\right) \\ + \left(1 - \Phi_{ij} - \Psi ij\right) \left(1 - y_{ij}\right) \left(1 - y_{ji}\right) \end{array} \right]$$

# A Less Trivial Model

- An important independent dyad model put forward by (Holland and Leinhardt, 1981), the $p_1$ family
  - $p_1$ pmf: $\Pr(G = g|\alpha, \beta, \gamma, \delta) \propto \exp\left[\alpha \sum_i \sum_j y_{ij} + \sum_i \beta_i \sum_j y_{ij} + \sum_i \gamma_i \sum_j y_{ji} + \delta \sum_i \sum_j y_{ij} y_{ji}\right]$
    - Density parameter $\alpha \in \mathbb{R}$
    - Expansiveness (outdegree) and popularity (indegree) parameters $\beta, \gamma \in \mathbb{R}^N$
    - Reciprocity (symmetry) parameter $\delta \in \mathbb{R}$
- Forged a conceptual revolution in network modeling
  - Nontrivial, yet parameters are estimable
  - Combines heterogeneity and dependence in a manageable way
  - Easily generalized, e.g. to latent blockmodels
  - First model to be posited in exponential family form

# Exponential Families for Random Graphs

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

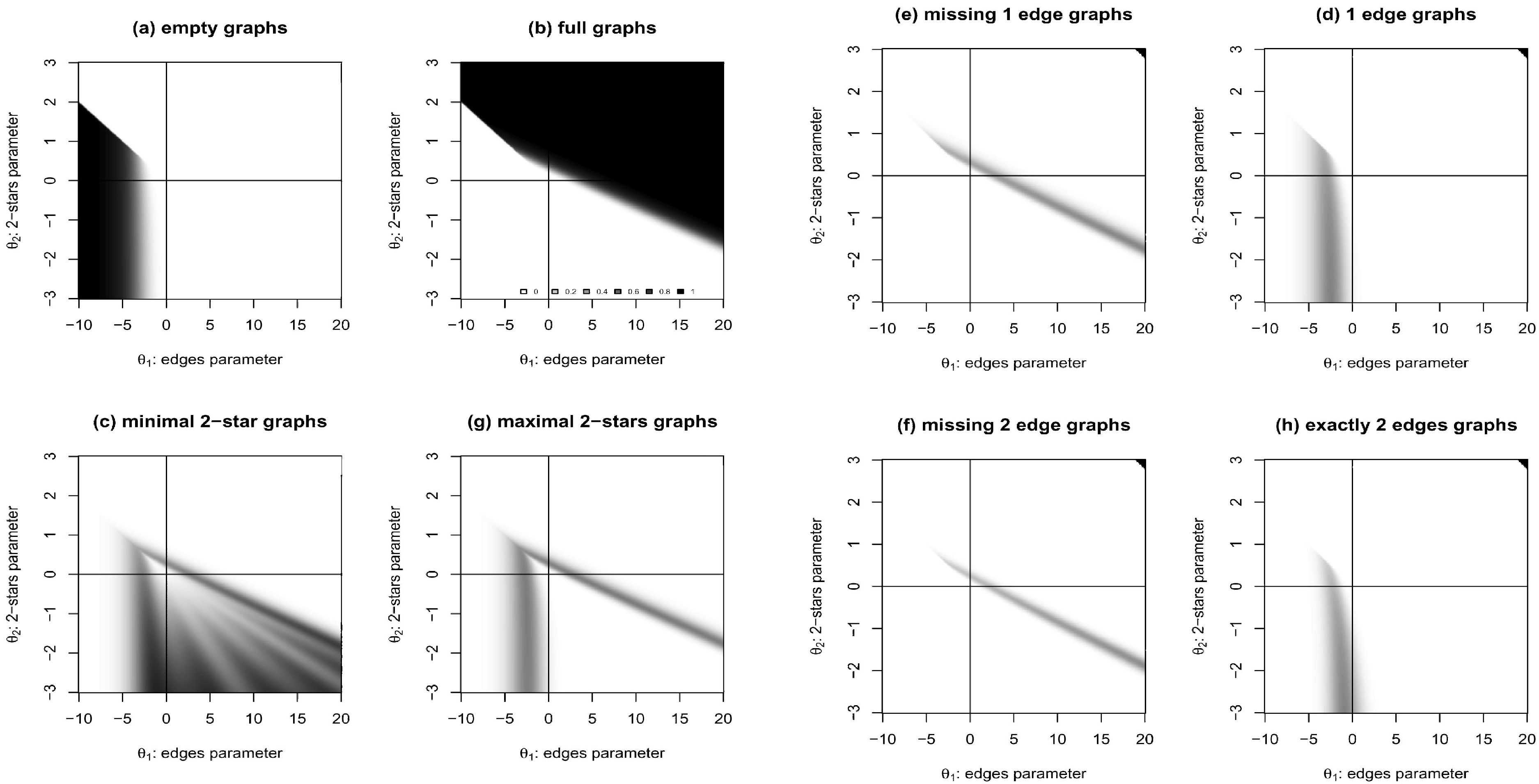
$$\Pr(G = g|\theta) = \frac{\exp\left(\theta^T \mathbf{t}(g)\right)}{\sum_{g' \in \mathcal{G}} \exp\left(\theta^T \mathbf{t}(g')\right)} I_{\mathcal{G}}(g) \quad (1)$$

- $\theta^T \mathbf{t}$: 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 $\theta$
  - Model is complete for pmfs on $\mathcal{G}$, few constraints on $\mathbf{t}$

# Deduction with ERGs

- Given a model family (specified by $t, \mathcal{G}$), should be able to explore properties deductively
- Primary tool – simulation by Markov chain Monte Carlo
- Analytical results hard to come by, except for independent-dyad families
  - Some results known for edge clustering, triangle models (see, e.g., Strauss (1986))
- **Research opportunity alert:** much to be learned about behavior of more general model families (e.g., curved parameterizations of Snijders et al. (2006))

# Example: Edge Clustering (aka 2-Star) Model



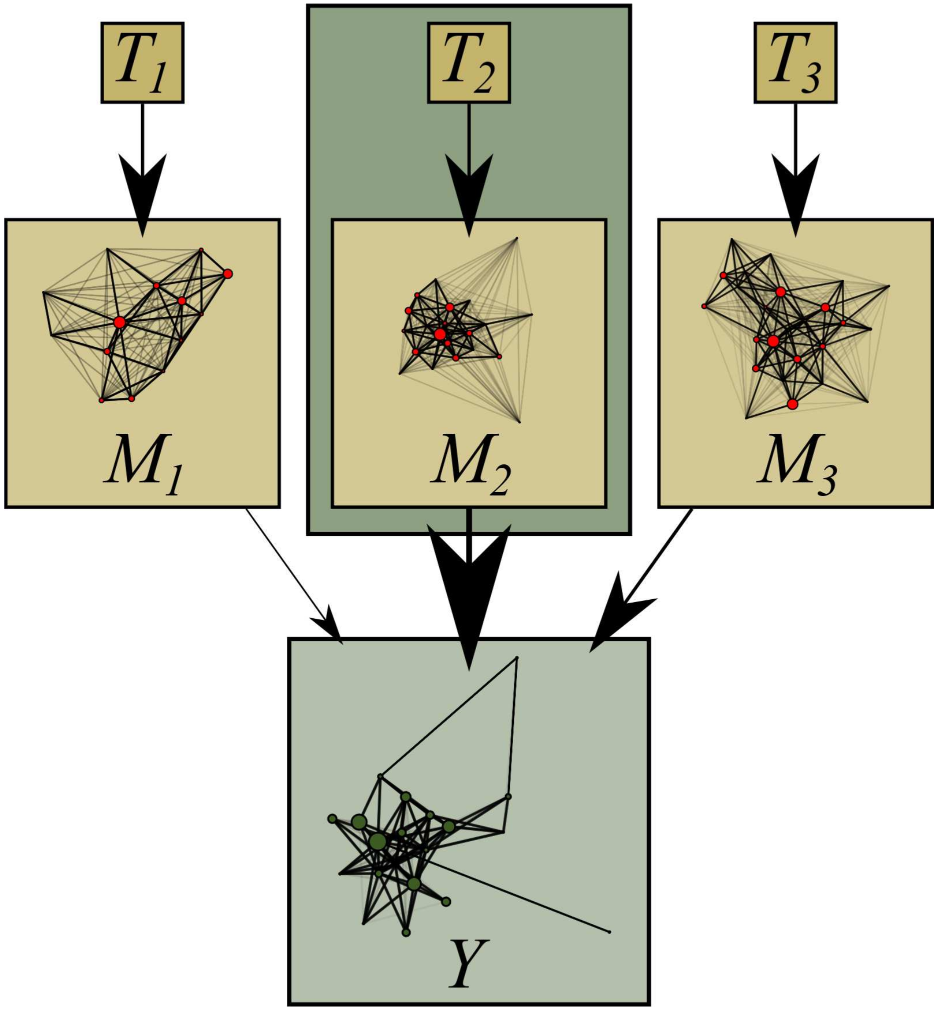

Figures 2 and 3, Handcock (2003)

# Simulating ERGs

- Calculation of ERG normalizing factor involves sum over support, and is hence infeasible
  - Simple Monte Carlo quadrature also fails, due to roughness of function – importance sampling is used (as we shall see)
- Luckily, can easily simulate using the Metropolis algorithm (Gilks et al., 1996))
  - Observe that $\frac{\Pr(G=g'|\theta)}{\Pr(G=g|\theta)} = \exp\left[\theta^T\left(t(g') - t(g)\right)\right]$ (normalizing factor divides out)
  - Algorithm sketch: Initially, draw $g^{(0)}$ from $\mathcal{G}$. For step $i > 0$, draw candidate $g'$ and let $g^{(i)} = g'$ with probability $\min\left(1, \frac{\Pr(G=g'|\theta)}{\Pr(G=g^{(i-1)}|\theta)}\right)$; otherwise, set $g^{(i)} = g^{(i-1)}$. Increment $i$, and repeat until $g^{(0)}, \ldots, g^{(n)}$ converges to $\Pr(G = g'|\theta)$.
  - Broadly speaking, convergence occurs in the infinite limit so long as the chain is irreducible and aperiodic, and so long as the proposal generation process satisfies detailed balance (i.e., the probability of proposing $g'|g^{(i-1)}$ must equal the probability of proposing $g^{(i-1)}|g'$)
  - If proposal mechanism is cleverly chosen, $t(g') - t(g)$ can be computed very rapidly for many statistics (i.e. $\mathcal{O}(1)$) – this is key to effective implementation
  - Can also use proposal mechanism to easily simulate support constraints (e.g., fixed density, degree distribution, etc.)

# From Deduction to Inference

- Modeling is important, but we also need inference
  - Need to discriminate among competing theories
  - May need to assess quantitative variation in effect strengths, etc.
- Basic logic
  - Derive ERG parameterization from prior theory
  - Assess fit to observed data
  - Select model/interpret parameters
  - Update theory and/or seek low-order approximating models
  - Repeat as necessary



# Inference for ERGs

- Inability to calculate normalizing factor is again a problem
- First approach: maximum pseudo-likelihood estimation (MPLE)
  - Given observed adjacency matrix $\mathbf{y}$, let $\mathbf{y}^+_{ij}, \mathbf{y}^-_{ij}$ be the matrices formed by setting $y_{ij}$ to 1 or 0 (respectively)
  - Note that, by definition, $\ln \frac{\Pr(Y_{ij}=1|\mathbf{y}^c_{ij},\theta)}{\Pr(Y_{ij}=0|\mathbf{y}^c_{ij},\theta)} = \theta^T \left(t\left(\mathbf{y}^+_{ij}\right) - t\left(\mathbf{y}^-_{ij}\right)\right)$ and $\Pr(Y_{ij}=1|\mathbf{y}^c_{ij},\theta) = \left(1+\exp\left[\theta^T\left(t\left(\mathbf{y}^-_{ij}\right) - t\left(\mathbf{y}^+_{ij}\right)\right)\right]\right)^{-1}$
  - Define the *pseudo-likelihood* by $\tilde{L}(G=g|\theta) = \prod_{(v_i,v_j)} \Pr(Y_{ij}=y_{ij}|\mathbf{y}^c_{ij},\theta)$
    - Pseudo-likelihood is equal to $\Pr(G=g|\theta)$ iff elements of $\mathbf{Y}$ are independent; more generally, we may hope that it is a reasonable approximation where this is not the case
  - The *maximum pseudo-likelihood estimator* of $\theta$ is defined as $\tilde{\theta} = \arg\max_\theta \tilde{L}(G=g|\theta)$
    - Can easily calculate from above, using *autologistic regression*
    - Alas, frequentist properties of $\tilde{\theta}$ unknown in the general case (likely poor; see Besag (2000); Lubbers and Snijders (2007); van Duijn et al. (2007))

# Inference for ERGs, Cont.

- Another approach: invoke the likelihood equation to find the maximum likelihood estimator (MLE)
  - MLE exists and is unique for ERGs under fairly broad conditions (e.g., affine independence of $t$, $t(g)$ in the relative interior of the convex hull of $t$ on $\mathcal{G}$)
  - For exponential families, $\mathbf{E}_{\hat{\theta}} t(G) = t(g)$ (where $\hat{\theta}$ is the MLE of $\theta$)
  - Thus, can estimate $\theta$ by simulating draws from $G|\theta'$, and adjusting until $\mathbf{E}_{\theta'} t(G) = t(g)$
  - Sounds easy (and is often used), but can be difficult in practice
    - Requires MCMC runs at every iteration
    - $(\mathbf{E}_{\theta'} t(G)) - t(g)$ can be very flat when $\theta'$ is far from $\hat{\theta}$, particularly if the model is near-degenerate (i.e., puts substantial mass on the boundary of the convex hull of $t$ on $\mathcal{G}$)
  - Also, does not compute maximized likelihood, nor does it provide estimates of uncertainty

# Inference for ERGs, Cont.

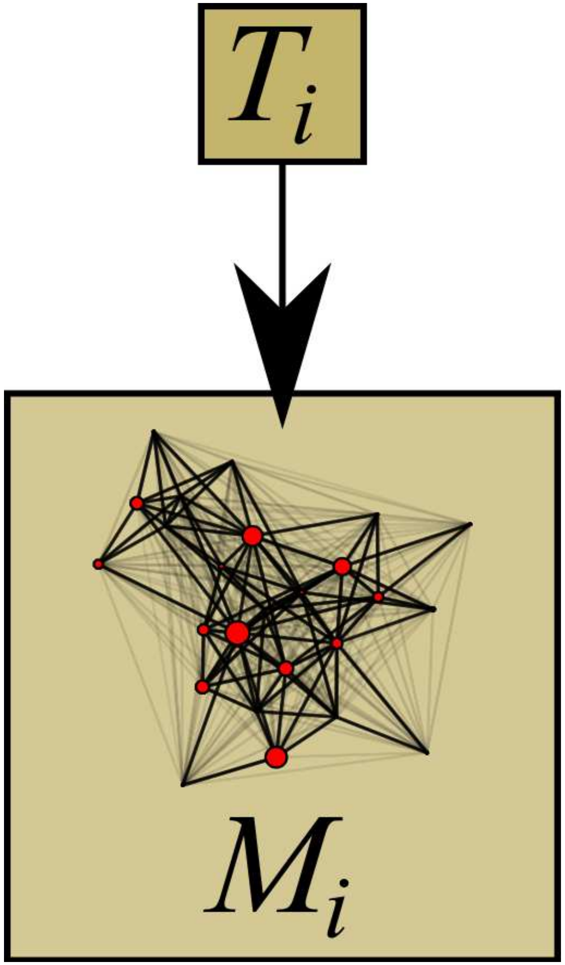
- Yet another alternative: approximate the MLE using importance sampling (Geyer and Thompson, 1992)
  - Note that if $\hat{\theta}$ is the MLE, then $\frac{\Pr(G=g|\hat{\theta})}{\Pr(G=g|\theta')} \geq \frac{\Pr(G=g|\theta'')}{\Pr(G=g|\theta')}$ for any $\theta', \theta''$; hence, maximizing the likelihood ratio versus any fixed $\theta'$ will yield $\hat{\theta}$
  - Let $\theta'$ be an initial "seed" estimate (e.g., the MPLE)
  - Draw $g^{(1)}, \ldots, g^{(n)}$ based on seed model, using Metropolis (or the like)
  - Now, choose $\hat{\theta}$ to maximize the estimated likelihood ratio

$$\frac{\Pr(G=g|\hat{\theta})}{\Pr(G=g|\theta')} \approx \exp\left[\left(\hat{\theta}-\theta\right)^T t(g)\right] \frac{1}{n}\left[\sum_{i=1}^{n} \exp\left[\left(\hat{\theta}-\theta\right)^T t\left(g^{(i)}\right)\right]\right]^{-1}$$

  - In general, converges to the true MLE as $n \to \infty$, provided that $\theta'$ is "close" to the target; can iterate the procedure if seed appears to have been too far away
  - Similar tricks can be used to get derivatives of the log-likelihood surface, etc.

# Parameterizing ERGs

- The ERG form is a way of representing distributions on $\mathcal{G}$, ***not*** a model in and of itself!
- Critical task: derive model statistics from prior theory
- Several approaches – we will show two here....



# 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,

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

  - 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 $\theta_{ij} = \operatorname{logit} \Phi_{ij}$
    - 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 $K_2$s, 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

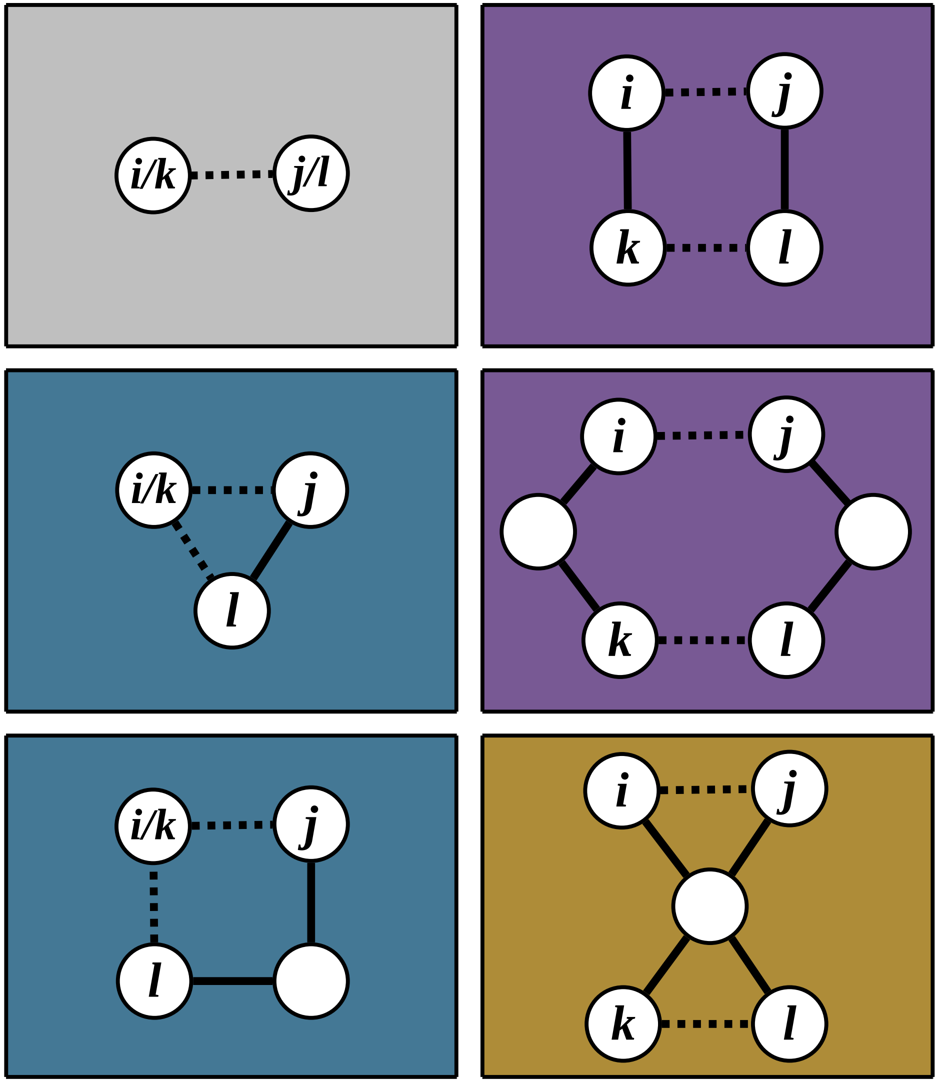
- An important advance by (Frank and Strauss, 1986): the Markov graphs
- The basic definition: $Y_{ij} \not\perp Y_{kl} | \mathbf{Y}^c_{ij,kl}$ iff $|\{i,j\} \cap \{k,l\}| > 0$
  - Intuitively, edge variables are conditionally dependent iff they share at least one endpoint
  - $D$ now has a large number 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

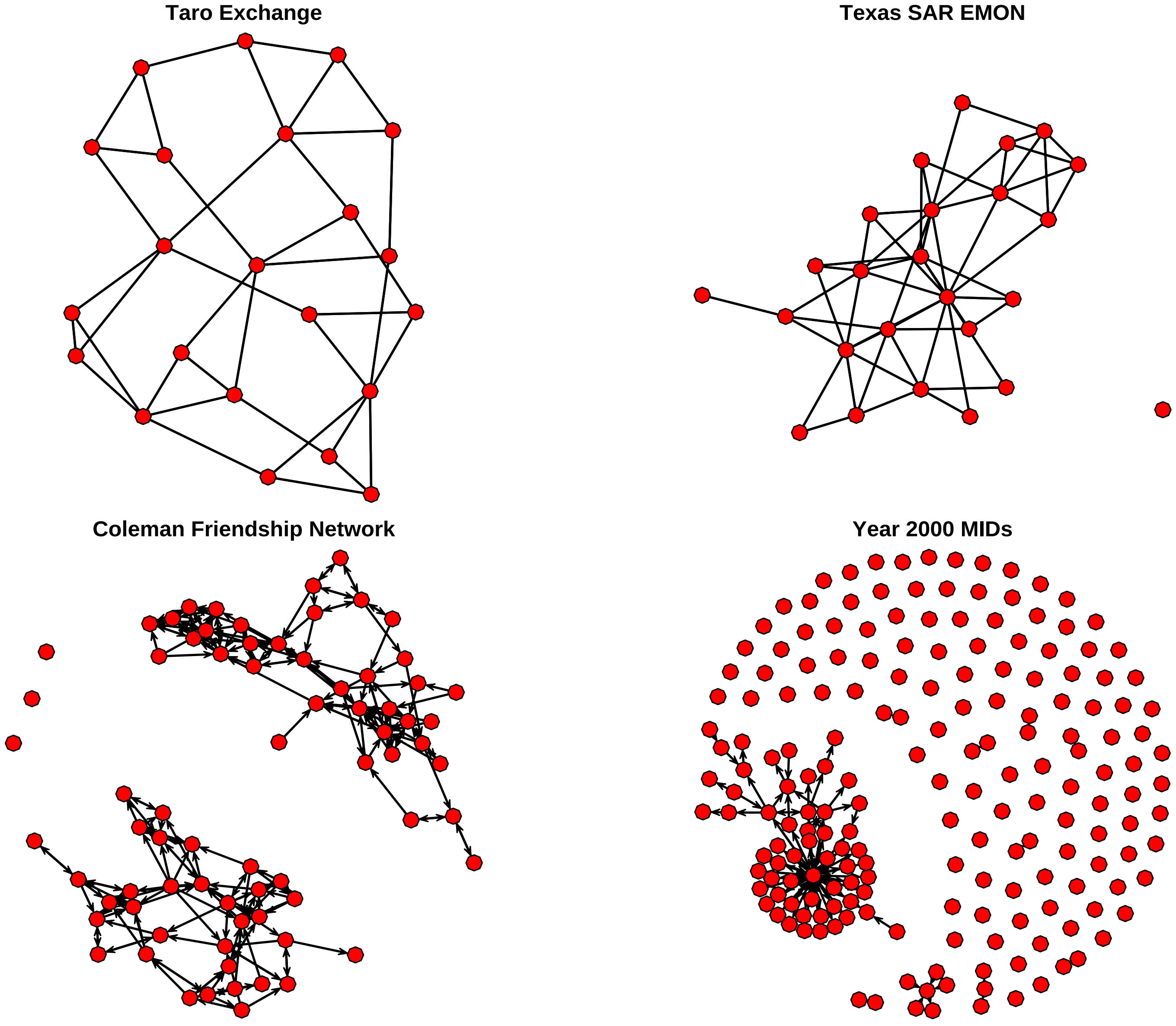
- Basic idea: head of each edge can reach the tail of the other
  - Weak case: (directed) paths each way are sufficient
  - Strong case: paths cannot share internal vertices
- Intuition: *extended reciprocity*
  - Possibility of feedback through network
  - In strong case, channels of reciprocation share 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

# Illustrative Application to Sample Networks



# Cycle Census ERG Fits

| | Taro Exchange | | | Texas EMON | | |
|---|---|---|---|---|---|---|
| | $\hat{\theta}$ | s.e. | $\Pr(>\|Z\|)$ | $\hat{\theta}$ | s.e. | $\Pr(>\|Z\|)$ |
| 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. | $\Pr(>\|Z\|)$ | $\hat{\theta}$ | s.e. | $\Pr(>\|Z\|)$ |
| 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

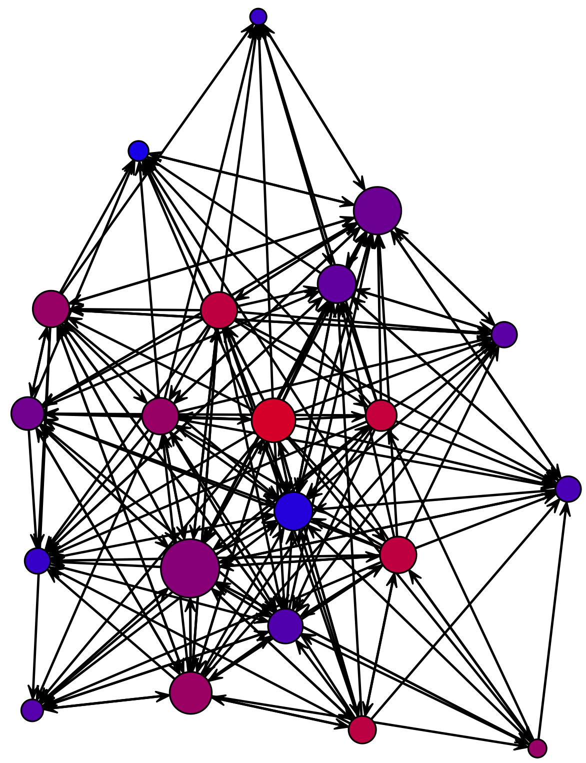
- Potential games (Monderer and Shapley, 1996)
  - Let $X$ by a strategy set, $u$ a vector utility functions, and $V$ a set of players. Then $(V, X, u)$ is said to be a *potential game* if $\exists\, \rho : X \mapsto \mathbb{R}$ such that $u_i\left(x_i', x_{-i}\right) - u_i\left(x_i, x_{-i}\right) = \rho\left(x_i', x_{-i}\right) - \rho\left(x_i, x_{-i}\right)\ \forall i \in V, x, x' \in X$.
- 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 $\rho$; (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.

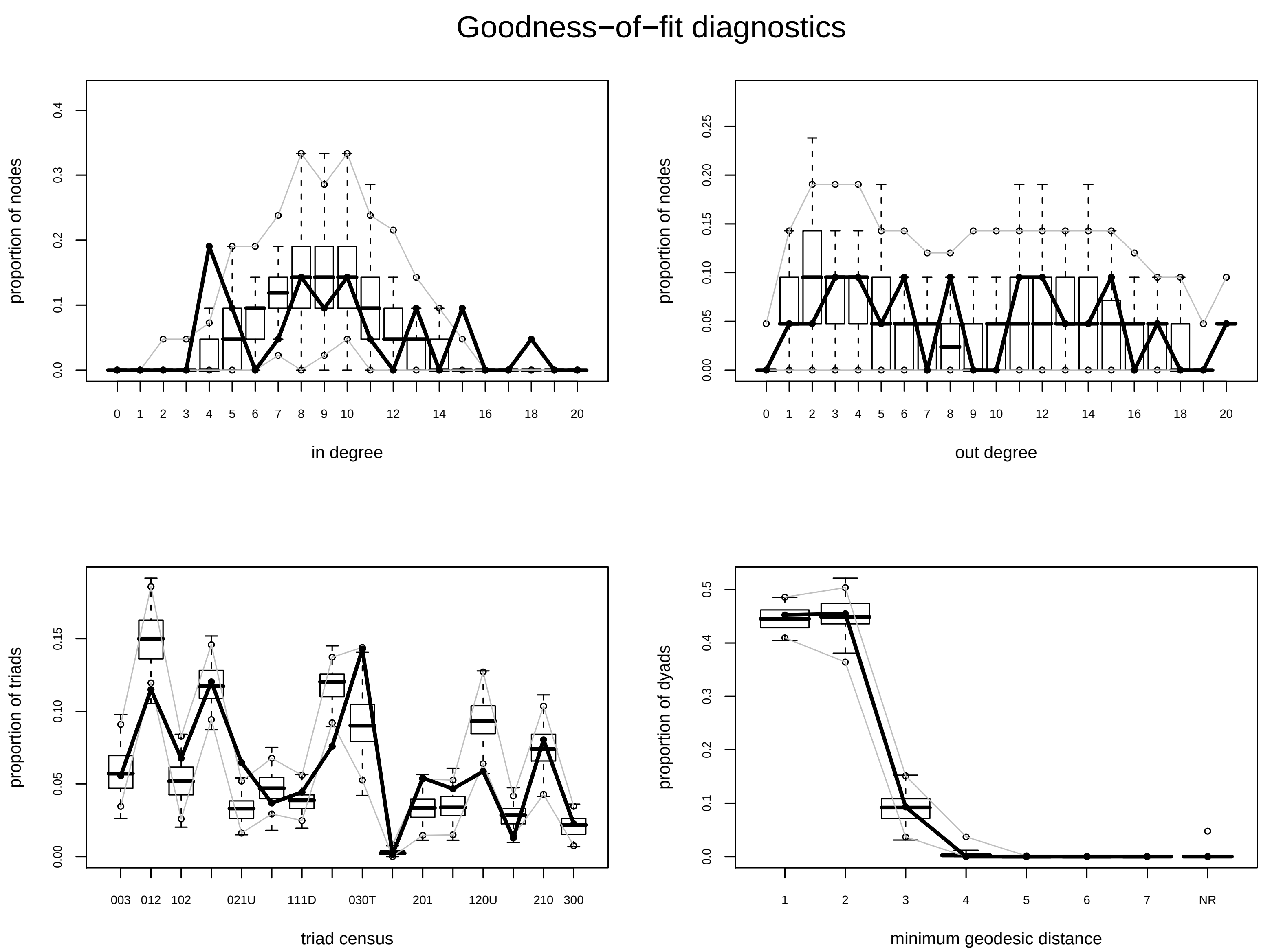
# Advice-Seeking ERG – AIC Selected Model

| 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?

# Model Adequacy Check



# Where Would One Go Next?

- Model refinement
  - Goodness-of-fit is not unreasonable, but some improvement is clearly possible
  - Could refine existing model (e.g., by adding covariates) or propose more alternatives
- Replication on new cases
  - Given a smaller set of candidates, would replicate on new organizations
  - May lead to further refinement/reformulation
- Simplification
  - Given a model family that works well, can it be simplified w/out losing too much?
  - Seek the smallest model which captures essential properties of optimal model; general behavior can then be characterized (hopefully)

# Summary

- Models for complex networks pose complex problems
  - 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

# In Closing: Where Physics Can Help

- Some areas in which input from the physics community would accelerate progress:
  - Pointers to relevant literature in statistical physics
  - Characterization of phase transitions in complex models
  - Analytical approximations (and hints as to where they are most likely to be found; can something like the virial expansion be developed here?)
  - Probably others which are not obvious to me!
- Great potential for contributions – perhaps the most "physics-ready" problem area in the network field!

# Proof of Potential Game Theorem

Assume an updating opportunity arises for $y_{ij}$, and assume that player $k$ has control of $y_{ij}$. By the logistic choice assumption,

$$\Pr\left(\mathbf{Y} = \mathbf{y}_{ij}^{+} \left| \mathbf{Y}_{ij}^{c} = \mathbf{y}_{ij}^{c}\right.\right) = \frac{\exp\left(u_k\left(\mathbf{y}_{ij}^{+}\right)\right)}{\exp\left(u_k\left(\mathbf{y}_{ij}^{+}\right)\right) + \exp\left(u_k\left(\mathbf{y}_{ij}^{-}\right)\right)} \quad (3)$$

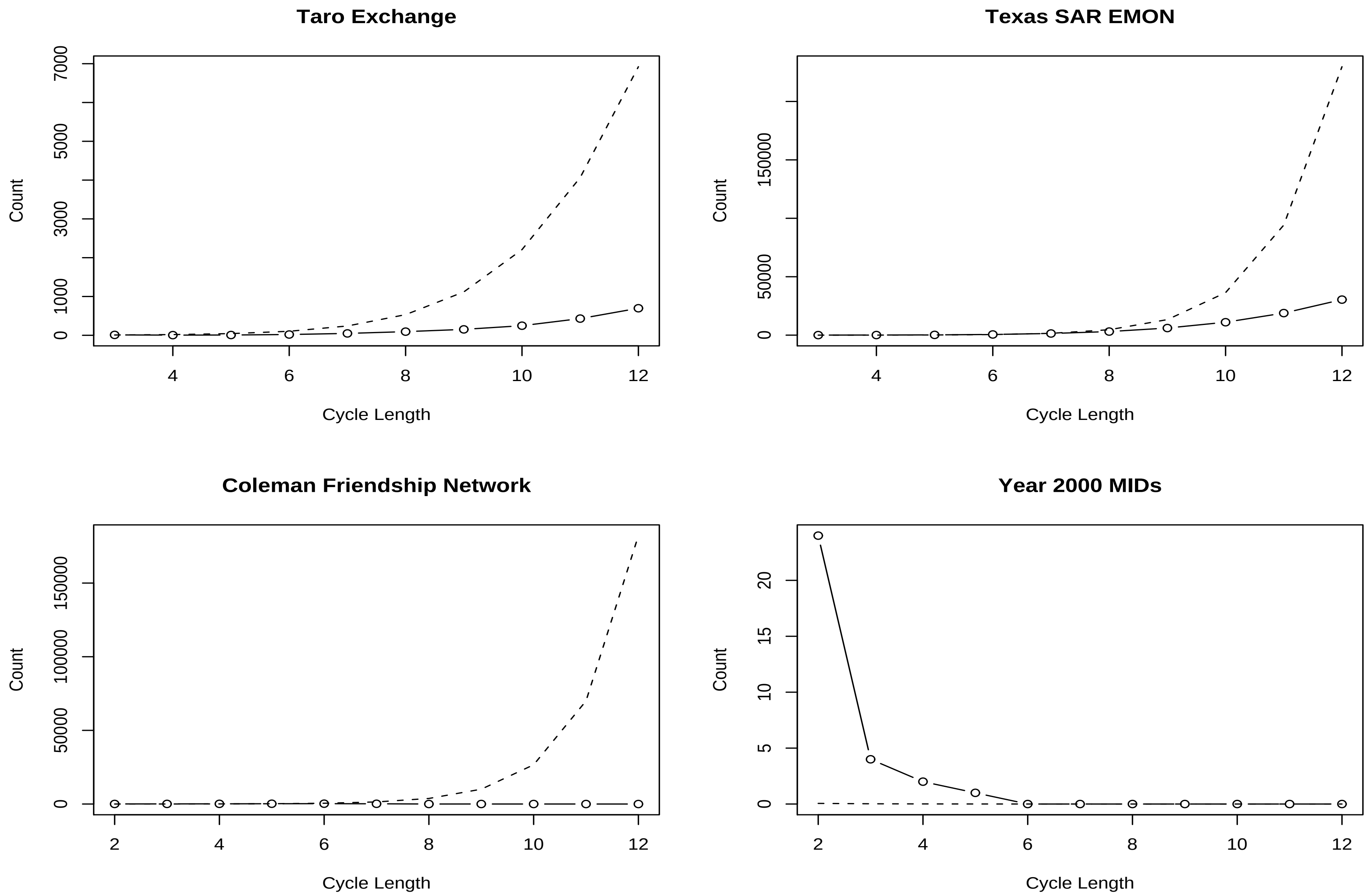
$$= \left[1 + \exp\left(u_k\left(\mathbf{y}_{ij}^{-}\right) - u_k\left(\mathbf{y}_{ij}^{+}\right)\right)\right]^{-1}. \quad (4)$$

Since $u, \mathcal{Y}$ form a potential game, $\exists\, \rho : \rho\left(\mathbf{y}_{ij}^{+}\right) - \rho\left(\mathbf{y}_{ij}^{-}\right) = u_k\left(\mathbf{y}_{ij}^{+}\right) - u_k\left(\mathbf{y}_{ij}^{-}\right)\ \forall\, k, (i,j), \mathbf{y}_{ij}^{c}$. Therefore, $\Pr\left(\mathbf{Y} = \mathbf{y}_{ij}^{+} \middle| \mathbf{Y}_{ij}^{c} = \mathbf{y}_{ij}^{c}\right) = \left[1 + \exp\left(\rho\left(\mathbf{y}_{ij}^{-}\right) - \rho\left(\mathbf{y}_{ij}^{+}\right)\right)\right]^{-1}$. Now assume that the updating opportunities for $\mathbf{Y}$ occur sequentially such that $(i,j)$ is selected independently of $\mathbf{Y}$, with positive probability for all $(i,j)$. Given arbitrary starting point $\mathbf{Y}^{(0)}$, denote the updated sequence of matrices by $\mathbf{Y}^{(0)}, \mathbf{Y}^{(1)}, \ldots$. This sequence clearly forms an irreducible and periodic Markov chain on $\mathcal{Y}$ (so long as $\rho$ is finite); it is known that this chain is a Gibbs sampler on $\mathcal{Y}$ with equilibrium distribution $\Pr(\mathbf{Y} = \mathbf{y}) = \frac{\exp(\rho(\mathbf{y}))}{\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\rho(\mathbf{y}'))}$, which is an ERG with potential $\rho$. By the ergodic theorem, then $\mathbf{Y}^{(i)} \rightarrow_{i \rightarrow \infty} ERG(\rho)$. QED.

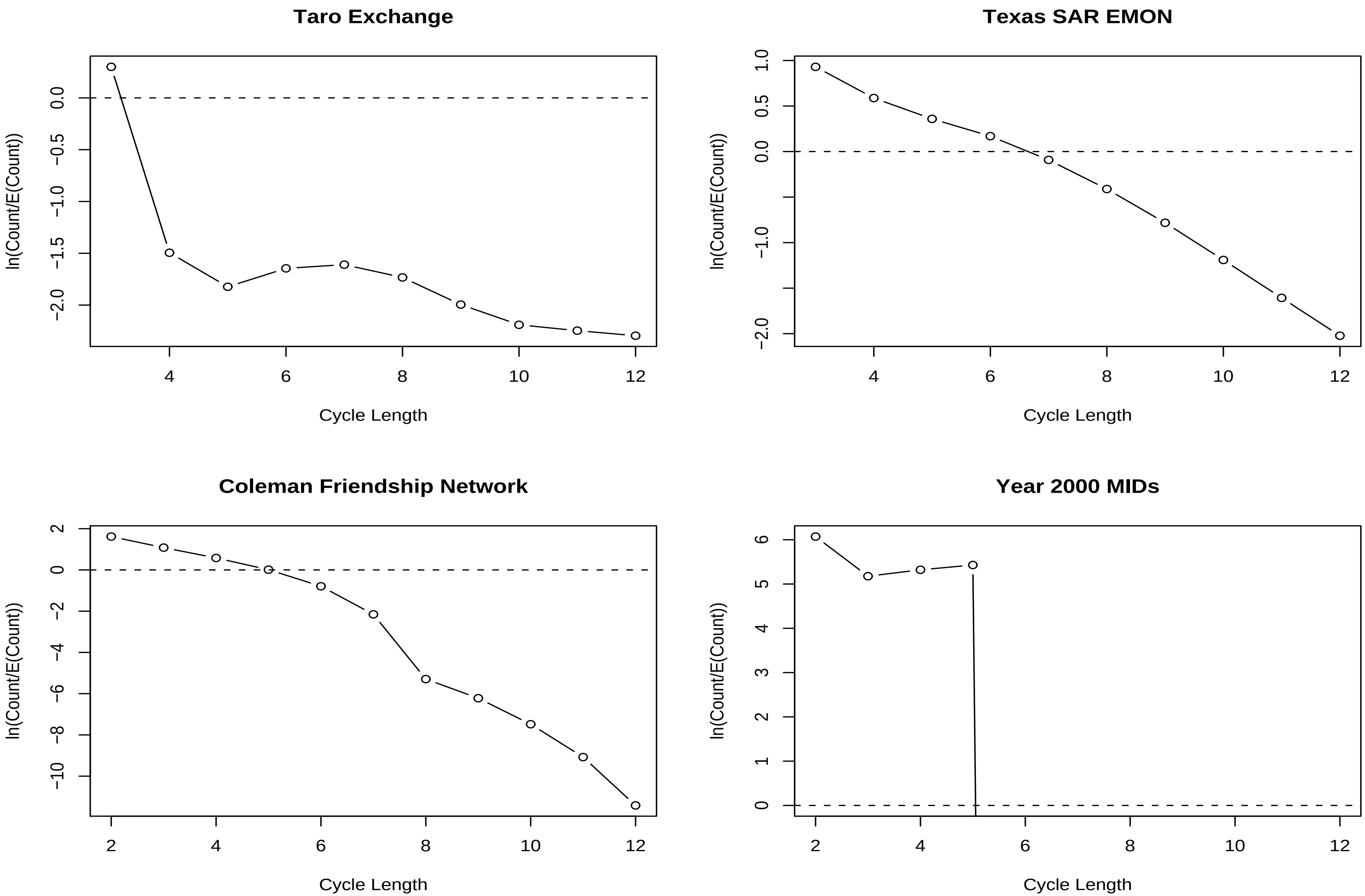
# Cycle Counts for the Sample Networks

| | Taro Exchange | Texas EMON | Friendship | MIDs |
|---|---|---|---|---|
| 2-Cycles | 0 | 0 | 62 | 24 |
| 3-Cycles | 10 | 40 | 88 | 4 |
| 4-Cycles | 4 | 89 | 136 | 2 |
| 5-Cycles | 7 | 226 | 202 | 1 |
| 6-Cycles | 20 | 592 | 240 | 0 |
| 7-Cycles | 48 | 1411 | 164 | 0 |
| 8-Cycles | 94 | 3068 | 19 | 0 |
| 9-Cycles | 152 | 6078 | 20 | 0 |
| 10-Cycles | 247 | 11059 | 15 | 0 |
| 11-Cycles | 430 | 18889 | 8 | 0 |
| 12-Cycles | 697 | 30403 | 2 | 0 |

# Cycle Census for the Sample Networks



# Cycle Census, Versus Expectation



# Cycle Statistics Defined

- Given graph $G = (V, E)$, define the $i$th *cycle statistic*, $c_i$, of $G$ by $c_i(G) \equiv |\{g \subseteq G : g \cong C_i\}|$, where $C_i$ is the cycle on $i$ vertices
- $\mathbf{c}(G)$ is the *cycle census* of $G$
- Familiar special cases
  - $c_2$ is the number of mutuals
  - $c_3$ is the number of cyclic triads (directed case) or triangles (undirected case)
  - $c_{|V|}$ is the indicator for a spanning (aka Hamiltonian) cycle
- Computation is NP-complete problem, but there are tricks....

# Expected Cycle Counts

- For $G$ Bernoulli, expectation of $\mathbf{c}$ can be obtained analytically:

$$\mathbf{E}_\theta\big(c_i\left(G\right)\big) = \begin{cases} \frac{\prod_{j=0}^{i-1}(N-j)}{i}\theta^i & \text{if } G \text{ directed} \\ \frac{\prod_{j=0}^{i-1}(N-j)}{2i}\theta^i & \text{if } G \text{ undirected} \end{cases} \quad (5)$$

- Variances are nontrivial, but can approximate by binomial process:

$$\mathbf{V}_\theta\big(c_i\left(G\right)\big) \approx \begin{cases} \frac{\prod_{j=0}^{i-1}(N-j)}{i}\theta^i(1-\theta^i) & \text{if } G \text{ directed} \\ \frac{\prod_{j=0}^{i-1}(N-j)}{2i}\theta^i(1-\theta^i) & \text{if } G \text{ undirected} \end{cases} \quad (6)$$

- Note that $\mathbf{c}$ *very* right-skewed for moderate/long cycles, so $z$-scores are not so useful....ERGs provide a better modeling approach

# Dependency Graphs and ERGs

- Let $\mathbf{X}$ be the adjacency matrix of $G$
  - $X_{ij} = 1$ if $(i,j) \in E$ and $X_{ij} = 0$ otherwise
  - $\mathbf{X}^c_{ab,cd,\ldots}$ denotes cells of $\mathbf{X}$ 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 $X_{ij} \not\perp X_{kl} | \mathbf{X}^c_{ij,kl}$
- From $D$ to $G$: the Hammersley-Clifford Theorem (Besag, 1974)
  - Let $K_D$ be the clique set of $D$. Then in the ERG case,

$$\Pr(G = g|\theta) = \frac{1}{Z(\theta,\mathcal{G})} \exp\left(\sum_{S \in K_D} \theta_S \prod_{(i,j) \in S} x_{ij}\right) \quad (7)$$

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

# Cycle Stats and the Dependence Graph

- Bad news: Hammersley-Clifford doesn't help much for cycle stats
  - 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): $X_{ij} \not\perp X_{kl} | \mathbf{X}^c_{ij,kl}$ only if some condition is satisfied
  - Lead to sufficient statistics which are subset of H-C stats
- One approach: *reciprocal path dependence*
  - Assume edges independent unless endpoints joined by (appropriately directed) paths

# Generalized Location Systems

- General framework for assignment of entities to positions
- Elements
  - Object set: $O = (o_1, \ldots, o_n)$
  - Location set: $L = (l_1, \ldots, l_m)$
  - Configuration vector: $\ell \in \mathcal{C}$
  - Accessible configuration set: $\mathcal{C} \subseteq \{1, \ldots, m\}^n$
  - Object/location feature sets: $F_L, F_O$
- Model family: posit discrete exponential family on $\mathcal{C}$, using functions of feature sets
- Some connections
  - Multigraph permutation model (Butts, 2007)
  - Zhang's (2004) segregation model
  - Schelling and Sakoda relocation processes

# A Form for the "Social Potential"

- Social potential $P(\ell)$ relates configural properties to probability
  - Vector features $\mathbf{Q}, \mathbf{X}, \mathbf{Y}, \mathbf{R}$
  - Relational features $\mathbf{B}, \mathbf{A}, \mathbf{W}, \mathbf{D}$
- Four classes of effects
  - Attraction/repulsion
  - Object heterogeneity
  - Location heterogeneity
  - Alignment
- $\Pr(\ell) \propto \exp(P(\ell))$
  - $P(\ell) = \alpha^T t^\alpha(\ell) + \beta^T t^\beta(\ell) + \gamma^T t^\gamma(\ell) + \delta^T t^\delta(\ell)$
  - $t_i^\alpha = \sum_{j=1}^n \mathbf{Q}_{\ell_j i} \mathbf{X}_{ji}$
  - $t_i^\beta = \sum_{j=1} \sum_{k=1}^n \mathbf{B}_{i \ell_j \ell_k} \left| \mathbf{Y}_{ji} - \mathbf{Y}_{ki} \right|$
  - $t_i^\gamma = \sum_{j=1} \sum_{k=1}^n \mathbf{A}_{ijk} \left| \mathbf{R}_{\ell_j i} - \mathbf{R}_{\ell_k i} \right|$
  - $t_i^\delta = \sum_{j=1} \sum_{k=1}^n \mathbf{A}_{ijk} \left| \mathbf{R}_{\ell_j i} - \mathbf{R}_{\ell_k i} \right|$

# 1 References


Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, 293:1818–1820.

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.

Besag, J. (2000). Markov chain Monte Carlo for statistical inference. Technical Report CSSS Working Paper 9, University of Washington.

Bollobás, B. (2001). *Random Graphs*. Cambridge University Press, Cambridge, second edition.

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.

Butts, C. T. (2007). Permutation models for relational data. *Sociological Methodology*, 37:257–281.

Drabek, T. E., Tamminga, H. L., Kilijanek, T. S., and Adams, C. R. (1981). *Managing Multiorganizational Emergency Responses: Emergent Search and Rescue Networks in Natural Disaster and Remote Area Settings*. Number Monograph 33 in Program on Technology, Environment, and Man. Institute of Behavioral Sciences, University of Colorado, Boulder, CO.

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.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood calculations (with discussion). *Journal of the Royal Statistical Society, Series C*, 54:657–699.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Handcock, M. S. (2003). Assessing degeneracy in statistical models of social networks. CSSS working paper no 39, University of Washington.

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.

Lubbers, M. J. and Snijders, T. A. B. (2007). A comparison of various approaches to the exponential random graph model: a reanalysis of 102 student networks in school classes. *Social Networks*, 29(4):489–507.

Mayhew, B. H., McPherson, J. M., Rotolo, M., and Smith-Lovin, L. (1995). Sex and race homogeneity in naturally occurring groups. *Social Forces*, 74(1):15–52.

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).

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36.

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

van Duijn, M. A. J., Gile, K., and Handcock, M. S. (2007). Comparison of pseudo-maximum likelihood and maximum likelihood estimation of exponential family random graph models. Technical Report CSS Working Paper 74, University of Washington.

Zhang, J. (2004). A dynamic model of residential segregation. *Journal of Mathematical Sociology*, 28:147–170.