Network Formation Models for Static Network Data

Chih-Sheng Hsieh

Department of Economics
National Taiwan University

March 18, 2024

Literature on Statistical Network Formation Models

In the statistical literature, there are several network formation models used to analyze static network data.

- The Erdös-Rényi-Gilbert Random Graph Model (Erdös and Rényi, 1959; Gilbert, 1959)
- The Blockmodels and Stochastic Blockmodels (Lorrain and White, 1977; Holland et al., 1983; Airoldi et al., 2008; Nowicki and Snijders, 2001)
- Latent Position Model (Hoff et al., 2002; Handcock et al., 2007; Krivitsky et al., 2009)
- Exponential Random Graph Model (Frank and Strauss, 1986; Wasserman and Pattison, 1996; Snijders et al., 2006)

Goldenberg et al. (2010) provides a comprehensive survey on these models.

Random graph model

- Application of mathematical random graph theory (Erdös and Rényi, 1959; Gilbert, 1959).
- A network is formed by having each bilateral link form *independently* with a probability p.
- The probability for such a random graph (represented by an $n \times n$ adjacency matrix W) is modeled by

$$\Pr(W) = \prod_{i<j} p^{W_{ij}}(1-p)^{1-W_{ij}}.$$

or the probability that there are M edges in a network W of n nodes is given by the Bernoulli distribution

$$\Pr(W) = p^M(1-p)^{N-M},$$

where $N = \binom{n}{2}$ is the number of potential links.

Random graph model

- The probability of a node that has k links, i.e., degree k, is captured by the binomial distribution

$$P(k) = P(a \text{ } vertex \text{ } has \text{ } degree \text{ } k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

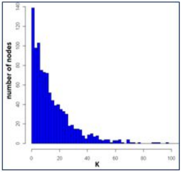
- When n is large and p is small, the binomial distribution can be approximated by the Poisson distribution.

$$P(k) = \frac{e^{-np}(np)^k}{k!}.$$

Random graph model

Degree distribution

- P(k): probability that a node has a degree of exactly k



- Common distributions:

Poisson:

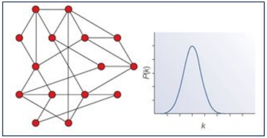
$$P(k) = \frac{e^{-d} d^k}{k!}$$

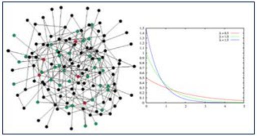
Exponential:

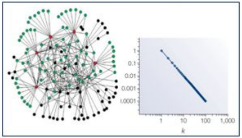
$$P(k) \propto e^{-k/d}$$

Power-law:

$$P(k) \propto k^{-c}, k \neq 0, c > 1$$





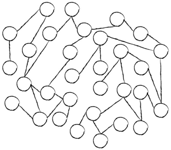


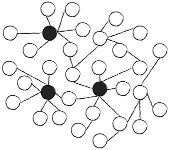
Random graph model

- The ER random graph model does not produce the "scale-free" property which is commonly observed in social networks.
- "scale-free network" describes the class of networks that exhibit a power-law degree distribution – very uneven distribution of connections. Some nodes have very high degrees of connectivity (hubs), while most have small degrees.

$$P(k) = \gamma k^{-c} \Rightarrow \; P(\alpha k) = \gamma(\alpha k)^{-c} = \gamma \alpha^{-c} k^{-c} = \gamma' k^{-c}$$

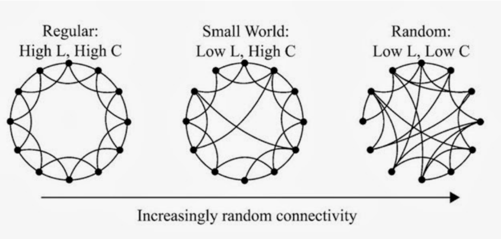
- In the typical power-law distribution, c is between 2 and 3, and the second moment (scale parameter) does not exist.


Random network


Scale-free network

Random graph model

- "small-world" network refers (1) the typical distance L between two randomly chosen nodes (the number of steps required) is small – grows proportionally to the logarithm of the number of nodes in the network and (2) the clustering coefficient C is not small (Travers and Milgram, 1967; Watts and Strogatz, 1998).
- The random graph model posses the first "small-world" property – a small average shortest path length (varying typically as the logarithm of the number of nodes) but not the second.



Random graph model

The other important property of the ER random graph is the phase transition for the size of largest (connected) component. Denote the mean degree by λ, which is computed by np, so we have $p = \lambda/n$.

Theorem

For $n = 1, 2, \cdots$, let $W_n \sim \mathrm{ER}(n, \lambda/n)$, then

1. *If $\lambda < 1$, then W_n will have no connected component of size larger than $O(\log n)$, with high probability (w.h.p) as $n \to \infty$.*
2. *If $\lambda = 1$, then W_n will have a largest component whose size is $O(n^{2/3})$, w.h.p as $n \to \infty$.*
3. *If $\lambda > 1$, then W_n will have a unique "giant" component containing a positive fraction of the nodes, w.h.p as as $n \to \infty$. No other component will contain more than $O(\log n)$ nodes, w.h.p as $n \to \infty$.*

Blockmodels and Stochastic Blockmodels

- The goal of blockmodeling (Lorrain and White, 1977) is to reduce a large, potentially incoherent network to a smaller comprehensible structure that can be interpreted more readily.
- a blockmodel is a model of network data that relies on the intuitive notion of "structural equivalence" – two nodes are defined to be structurally equivalent if their connectivity with similar nodes is similar.
- Following up this idea, we can imagine collapsing structurally equivalent nodes together to form a block in the language of blockmodels.

Blockmodels and Stochastic Blockmodels

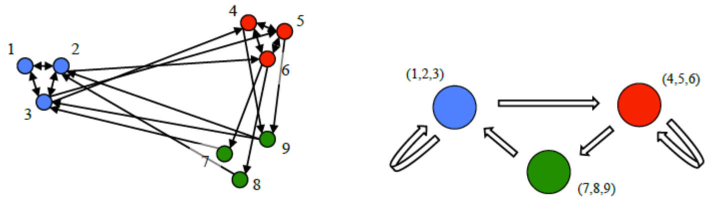


Figure 3.2: Left: An example graph. Right: The corresponding blockmodel, where red nodes have been collapsed into the red block and similarly for the other colors. Note that this problem is not a typical clustering problem, as the green nodes do not share any direct connections; each green node, however, has connections directed to blue nodes, and connections directed from red nodes. In other words, given the partition into monochromatic blocks, the nodes in the green block share patterns of connectivity to nodes in other blocks.

Blockmodels and Stochastic Blockmodels

- Given N nodes and K blocks, let W be the adjacency matrix of the graph G, then two nodes a and b are structurally equivalent, and thus belong to the same block h, if their connectivity patterns $\mathcal{C}_a$ and $\mathcal{C}_b$ with nodes in the other blocks are similar.
- The statement can be formally written as

$$\mathcal{C}_a \equiv \{W(a, i \in h_k) : \forall h_k \neq h\} \approx \mathcal{C}_b,$$

where the index i runs over the nodes other than a and b, the index k runs over the blocks other than h, h_k is the set of nodes in block k.

Blockmodels and Stochastic Blockmodels

- The concept of deterministic equivalence behind the block models is rather rigid.
- The stochastic blockmodel (SBM) (Holland et al., 1983) weakens this concept to the stochastic one – given the assignment of nodes to blocks, the probability of an edge between two nodes just depends on which blocks they are in and is independent across edges.
- The SBM of n nodes is generally defined by the following components:
  - k : a scalar value denoting the number of groups in the network
  - z : a $n \times 1$ vector where z_ℓ gives the block index of vertex ℓ.
  - P : a $k \times k$ stochastic block matrix, where P_{ij} gives the probability that a vertex of block i is connected to a vertex in block j.
  - SBM is the simplest "generative" model for data with the cluster feature.

Blockmodels and Stochastic Blockmodels

The development of the stochastic blockmodel literature:

- Snijders and Nowicki (1997) study the case of undirected graphs with only two clusters.
- Nowicki and Snijders (2001) extend to the case where relations can be directed and the number of blocks is arbitrary.
- Airoldi et al. (2008) develop a mixed memberships model which associates each unit of observation with multiple blocks rather than a single block.
- This literature is closely related to physics and computer science literature on community structure detection in social networks (Girvan and Newman, 2002; Newman, 2006; Blondel et al., 2008) – these approaches are generally mechanical and involve less economic intuition.

Latent Position model

- The latent position model is proposed by Hoff et al. (2002).
- Assume each node has a unique position in the unobserved latent space.
- $N \times N$ adjacency matrix W, with entries $w_{ij} = 1$ or 0.
- The link probability is modeled in a Logit form:

$$p(w_{ij} = 1 | z_i, z_j, x_{ij}, \theta) = \frac{\exp(\psi_{ij})}{1 + \exp(\psi_{ij})}, \quad \psi_{ij} = \theta_0 + \theta_1' x_{ij} + \theta_2 h(z_i, z_j).$$

- $h(z_i, z_j)$ could be $\| z_i - z_j \|$ (distance model) or $z_i z_j$ (projection model).
- $X = \{x_{ij}\}$ are observed pair-specific characteristics and θ and $Z = \{z_i, i = 1, \cdots, N\}$, are parameters and positions to be estimated.
- The joint probability of W is modelled as

$$P(W|Z, X, \theta) = \prod_i \prod_{j \neq i} P(w_{ij}|z_i, z_j, x_{ij}, \theta),$$

Estimating Latent Position model

- Hoff et al. (2002) combine maximum likelihood (ML) estimation and Bayesian Markov Chain Monte Carlo (MCMC) estimation. First identify MLE $\hat{Z}$ of Z by direct maximization of the likelihood. Then use the M-H algorithm to update parameters θ and Z with $\hat{Z}$ as the initial value of the MCMC chain.
- Handcock et al. (2007) use a full Bayesian approach. Treat Z as individual random effects and update them together with other parameters in MCMC.
- The R package "latentnet" provides simulation and estimation of latent position model.

Latent Position model

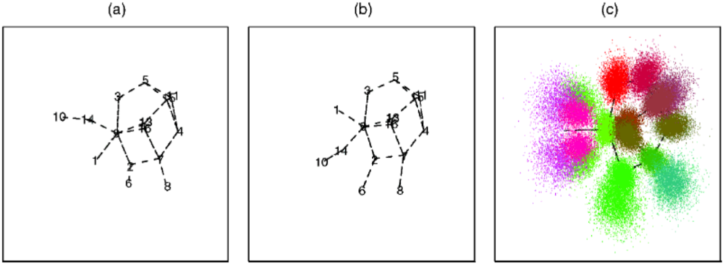


Figure 3. *(a) and (b) Alternate d_2 Representations of the Florentine Family Data. (c) Marginal posterior distributions of family positions.*

The extension of Latent Position model

- Handcock et al. (2007) extend the latent space model of Hoff et al. (2002) to capture the clustering property of social networks, assume z_i to be drawn from a finite mixture of G multivariate normal distribution

$$z_i \sim \sum_{g=1}^{G} \lambda_g \mathrm{MN}_d(\mu_g, \sigma_g^2 I_d), \qquad \lambda_g \geq 0, \sum_{g=1}^{G} \lambda_g = 1$$

- We can estimate the probability of individual i belonging to cluster g, i.e., $P(k_i = g)$, $i = 1, \cdots, N$, $g = 1, \cdots, G$ in the procedure.
- Krivitsky et al. (2009) represent heterogeneity in the propensity for nodes to form ties, add node-specific random effects into the model

$$\psi_{ij} = \theta_0 + \theta_1' x_{ij} + \theta_2 h(z_i, z_j) + \delta_i + \delta_j$$

The extension of Latent Position model

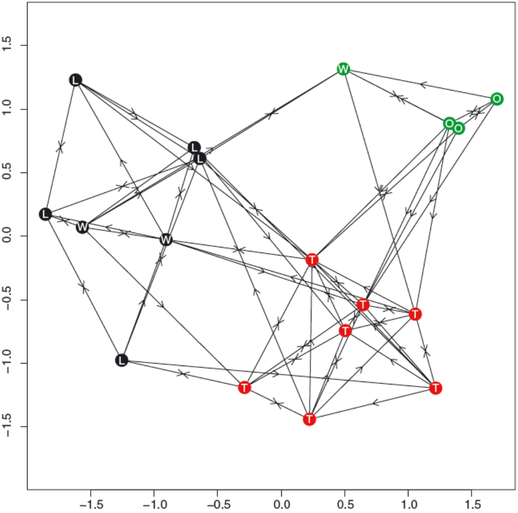


Fig. 1. Relationships between monks within a monastery: groups from two-stage maximum likelihood estimation of the latent position cluster model with three groups are shown by the colours of the nodes; a grouping given by Sampson (1969) is shown by the letters T (Turks), L (Loyal Opposition), O (Outcasts) and W (Waverers) (→, ties (i.e. the data))

Combining SAR model with latent position model (Hsieh and Lee, 2016)

- We consider a modeling approach to correct the estimation biases in the SAR model resulted from an endogenous spatial weight matrix.
- We treat entries of the spatial weight matrix as endogenous choice variables and model them by a network formation process.
- In order to reflect the idea of having unobserved factors (e.g., ability, preference, taste, etc.) which affect both friendship decisions and economic outcomes, we specify a number of unobserved (latent) variables in both the network model and the SAR model.

Combining SAR model with latent position model (Hsieh and Lee, 2016)

- The SAR model:

$$Y_g = \lambda W_g Y_g + X_g \beta_1 + W_g X_g \beta_2 + l_g \alpha_g + \epsilon_g, \quad g = 1, \cdots, G,$$

- Our network model is modified from the latent position model (Hoff et al., 2002):

$$P(w_{ij,g} | c_{i,g}, c_{j,g}, c_{ij,g}, z_{i,g}, z_{j,g}, \gamma, \delta) = \frac{\exp(w_{ij,g}\psi_{ij,g})}{1 + \exp(\psi_{ij,g})},$$

$$\psi_{ij,g} = c_{i,g}\gamma_1 + c_{j,g}\gamma_2 + c_{ij,g}\gamma_3 + \delta_1 |z_{i1,g} - z_{j1,g}| + \cdots + \delta_{\bar{d}} |z_{i\bar{d},g} - z_{j\bar{d},g}|.$$

- $c_{i,g}$ and $c_{ij,g}$: individual-specific and dyad-specific characteristics.
- $z_{i,g} = (z_{i1,g}, \cdots, z_{id,g}, \cdots, z_{i\bar{d},g})$: a $1 \times \bar{d}$ vector of the unobserved characteristics, such as ability, preference, taste, etc.

Combining SAR model with latent position model (Hsieh and Lee, 2016)

- We assume the disturbances of the SAR outcome equation, ϵ_g, and unobserved variables, Z_g, to follow a multivariate normal distribution, i.e.,

$$(\epsilon_{i,g}, z_{i,g}) \sim i.i.d.\ \mathcal{N}_{\bar{d}+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon z} \\ \sigma_{z\epsilon} & \Sigma_z \end{pmatrix}\right),$$

  where $\sigma_{\epsilon z} = (\sigma_{\epsilon z_1}, \cdots, \sigma_{\epsilon z_{\bar{d}}})$ and Σ_z is a $\bar{d} \times \bar{d}$ covariance matrix.

- The SAR equation should be modified to

$$Y_g = \lambda W_g Y_g + X_g \beta_1 + W_g X_g \beta_2 + l_g \alpha_g + Z_g \Sigma_z^{-1} \sigma_{z\epsilon} + u_g,$$

  where $u_g \sim \mathcal{N}_{m_g}(0, \sigma_u^2 I_{m_g})$ with $\sigma_u^2 = \left(\sigma_\epsilon^2 - \sigma_{\epsilon z} \Sigma_z^{-1} \sigma_{z\epsilon}\right)$.

Combining SAR model with latent position model (Hsieh and Lee, 2016)

- The SCSAR model should be joined with the network formation model as a system to study network interaction effects on economic outcomes.
- The joint probability of Y_g and W_g, given observed exogenous variables, is

$$P(Y_g, W_g | X_g, C_g, \theta, \alpha_g)$$
$$= \int_{Z_g} P(Y_g | W_g, X_g, Z_g, \theta, \alpha_g) \cdot P(W_g | C_g, Z_g, \theta) \cdot f(Z_g) dZ_g,$$

  where $\theta = (\gamma', \delta', \lambda, \beta_1', \beta_2', \sigma_\epsilon^2, \sigma_{\epsilon z}')$.
- If $\sigma_{\epsilon z}$ were all zero, Z_g and ϵ_g would be uncorrelated. The influences from the endogeneity of W_g to outcomes would be eliminated and the SAR model can be estimated by treating W_g as exogenously given.

Combining SAR model with latent position model (Hsieh and Lee, 2016)

The identification constraints:

- Σ_z should be normalized to $I_{\bar{d}}$.
- The signs of $\sigma_{\epsilon z}$ are arbitrary. So we normalize them to positive.
- We are not able to distinguish $|z_{id,g} - z_{jd,g}|$ from $|z_{il,g} - z_{jl,g}|$, $l \neq d$ in the network model. Hence, we require $|\delta_1| \geq \cdots \geq |\delta_{\bar{d}}|$.
- For Z_g in multiple dimensions, we need to assume the distribution of Z_g and the distribution of pure disturbance to be known.

Combining SAR model with latent position model (Hsieh and Lee, 2016)

Table: Simulation Study

Parameter	True	SAR		SCSAR ($\bar{d} = 2$)		SCSAR ($\bar{d} = 1$)	
		Average	S.D.	Average	S.D.	Average	S.D.
λ	0.050	0.068	0.010	0.047	0.006	0.065	0.007
β_1	0.500	0.499	0.021	0.502	0.018	0.499	0.021
β_2	0.500	0.491	0.015	0.502	0.013	0.493	0.016
σ_u^2	0.500	0.498	0.022	0.508	0.024	0.501	0.022
$\sigma_{\epsilon z_1}$	0.000	-	-	0.071	0.091	0.149	0.092
$\sigma_{\epsilon z_2}$	0.500	-	-	0.482	0.091	-	-
γ_{31}	-0.500	-	-	-0.502	0.067	-1.342	0.057
γ_{32}	1.000	-	-	1.003	0.050	0.967	0.051
δ_1	-2.000	-	-	-2.000	0.132	-2.151	0.110
δ_2	-1.000	-	-	-1.013	0.124	-	-

Combining SAR model with latent position model (Hsieh and Lee, 2016)

Table: Estimations result of the SAR and SCSAR models

	SAR				SCSAR ($\bar{d} = 1$)			
Endogenous	0.029	[0.012, 0.046]			0.029	[0.011, 0.045]		
	Own		Contextual		Own		Contextual	
Age	-0.193	[-0.240, -0.145]	-0.005	[-0.009, -0.001]	-0.200	[-0.255, -0.149]	-0.005	[-0.009, -0.001]
Male	-0.111	[-0.180, -0.043]	-0.029	[-0.070, 0.012]	-0.110	[-0.179, -0.042]	-0.029	[-0.070, 0.013]
Black	-0.096	[-0.231, 0.035]	-0.029	[-0.065, 0.007]	-0.100	[-0.234, 0.033]	-0.030	[-0.066, 0.006]
Asian	-0.012	[-0.269, 0.249]	0.040	[-0.129, 0.207]	-0.012	[-0.274, 0.245]	0.042	[-0.124, 0.211]
Hispanic	0.032	[-0.125, 0.186]	-0.044	[-0.126, 0.040]	0.030	[-0.127, 0.185]	-0.045	[-0.130, 0.038]
Other races	-0.117	[-0.252, 0.020]	-0.076	[-0.151, -0.003]	-0.118	[-0.253, 0.019]	-0.075	[-0.148, 0.000]
Both parents	0.070	[0.009, 0.146]	0.014	[-0.035, 0.062]	0.069	[0.006, 0.144]	0.015	[-0.033, 0.063]
Less HS	-0.145	[-0.259, -0.029]	-0.040	[-0.116, 0.033]	-0.144	[-0.261, -0.030]	-0.041	[-0.116, 0.033]
More HS	0.145	[0.067, 0.226]	0.005	[-0.044, 0.051]	0.145	[0.066, 0.223]	0.005	[-0.042, 0.053]
Edu missing	-0.081	[-0.204, 0.044]	0.038	[-0.047, 0.118]	-0.079	[-0.204, 0.044]	0.039	[-0.045, 0.121]
Welfare	0.125	[-0.242, 0.499]	0.028	[-0.221, 0.270]	0.125	[-0.252, 0.492]	0.027	[-0.223, 0.271]
Job missing	-0.042	[-0.168, 0.082]	-0.007	[-0.089, 0.074]	-0.043	[-0.168, 0.081]	-0.006	[-0.090, 0.075]
Professional	-0.004	[-0.099, 0.094]	0.018	[-0.040, 0.076]	-0.004	[-0.102, 0.091]	0.018	[-0.039, 0.076]
Other jobs	0.029	[-0.056, 0.111]	0.046	[-0.003, 0.094]	0.028	[-0.055, 0.113]	0.046	[-0.002, 0.095]
σ_ϵ^2	0.459	[0.430, 0.490]			0.461	[0.430, 0.490]		
$\sigma_{\epsilon z_1}$	-				0.012	[0.000, 0.031]		

Combining SAR model with latent position model (Hsieh and Lee, 2016)

Table: Estimations result of the SAR and SCSAR models (continued)

	SCSAR ($\bar{d} = 2$)				SCSAR ($\bar{d} = 3$)			
Endogenous	0.020	[0.003, 0.036]			0.019	[0.000, 0.033]		
	Own		Contextual		Own		Contextual	
Age	-0.191	[-0.241, -0.140]	-0.004	[-0.008, 0.001]	-0.200	[-0.255, -0.149]	-0.005	[-0.009, -0.001]
Male	-0.110	[-0.179, -0.043]	-0.029	[-0.070, 0.012]	-0.110	[-0.179, -0.042]	-0.029	[-0.070, 0.013]
Black	-0.089	[-0.229, 0.049]	-0.026	[-0.063, 0.009]	-0.100	[-0.234, 0.033]	-0.030	[-0.066, 0.006]
Asian	-0.013	[-0.268, 0.247]	0.051	[-0.119, 0.216]	-0.012	[-0.274, 0.245]	0.042	[-0.124, 0.211]
Hispanic	0.047	[-0.110, 0.202]	-0.031	[-0.116, 0.052]	0.030	[-0.127, 0.185]	-0.045	[-0.130, 0.038]
Other races	-0.120	[-0.254, 0.019]	-0.072	[-0.145, 0.003]	-0.118	[-0.253, 0.019]	-0.075	[-0.148, 0.000]
Both parents	0.080	[0.002, 0.156]	0.022	[-0.025, 0.070]	0.069	[-0.002, 0.144]	0.015	[-0.033, 0.063]
Less HS	-0.136	[-0.249, -0.019]	-0.038	[-0.113, 0.035]	-0.144	[-0.261, -0.030]	-0.041	[-0.116, 0.033]
More HS	0.139	[0.060, 0.216]	0.001	[-0.046, 0.048]	0.145	[0.066, 0.223]	0.005	[-0.042, 0.053]
Edu missing	-0.095	[-0.219, 0.030]	0.024	[-0.059, 0.105]	-0.079	[-0.204, 0.044]	0.039	[-0.045, 0.121]
Welfare	0.133	[-0.244, 0.503]	0.017	[-0.222, 0.263]	0.125	[-0.252, 0.492]	0.027	[-0.223, 0.271]
Job missing	-0.038	[-0.164, 0.085]	-0.006	[-0.076, 0.088]	-0.043	[-0.168, 0.081]	-0.006	[-0.090, 0.075]
Professional	-0.002	[-0.100, 0.092]	0.024	[-0.033, 0.082]	-0.004	[-0.102, 0.091]	0.018	[-0.039, 0.076]
Other jobs	0.026	[-0.058, 0.110]	0.048	[0.000, 0.097]	0.028	[-0.055, 0.113]	0.046	[-0.002, 0.095]
σ_ϵ^2	0.469	[0.438, 0.502]			0.469	[0.439, 0.503]		
$\sigma_{\epsilon z_1}$	0.148	[0.080, 0.223]			0.104	[0.002, 0.191]		
$\sigma_{\epsilon z_2}$	0.098	[0.015, 0.168]			0.069	[0.000, 0.146]		
$\sigma_{\epsilon z_3}$	-				0.149	[0.049, 0.232]		

Exponential Random Graph Model

- The Exponential Random Graph (ERG) model is proposed by Frank and Strauss (1986).
- ERGM takes a general exponential form given by

$$\Pr(W|\theta) = \left(\frac{1}{\kappa(\theta)}\right) \exp\left(\theta^T z(W)\right),$$

  where $z(W)$ is a vector of sufficient statistics for different network configurations and θ is the corresponding vector of model parameters.
- When any element of θ is zero, edge variables in the corresponding network configuration are conditionally independent of each other and therefore that configuration is irrelevant to the network model.
- A normalizing constant $\kappa(\theta) = \sum_{W' \in \Omega_W} \exp\left(\theta^T z(W')\right)$ is presented to ensure it is a valid probability measure.

Exponential Random Graph Model

- The theoretical foundation of ERGM is Markov random field (MRF).
- Markov random field (or Markov network): Given an undirected graph $G = (V, E)$ and $\mathbf{X} = (X_1, \cdots)'$ be a collection of discrete random variables defined on V. $\mathbf{X}$ is a MRF on G if
  i) $P(\mathbf{X} = x) > 0$ for all x;
  ii) $P(X_i = x_i | \mathbf{X}_{-i} = x_{-i}) = P(X_i = x_i | \mathbf{X}_{N_i} = x_{N_i})$, where N_i denotes the set of neighborhoods of i.
- The key feature of MRFs is that they satisfy the Markov property, which states that the probability of a "node" being in a particular state is only dependent on the states of its neighboring nodes.
- ERGMs are an extension of MRF that model the probability distribution over a set of "edges" (binary variables) in a network.

Exponential Random Graph Model

- Because dependence is defined in local neighborhoods (cliques), so the joint probability can be factorized over the cliques of G: $P(\mathbf{X}=x)=\prod_{C\in \mathrm{cl}(G)}\phi_C(x_C)$.
- Gibbs random field: a random vector $\mathbf{X}$ has the Gibbs distribution $P(\mathbf{X}=x)=\frac{1}{\kappa}\exp(U(x))$, where $U(x)=\sum_{c\in\Omega}U_c(x)$, with Ω denotes the set of all cliques of all sizes in G, and $\kappa=\sum_{x\in\mathbf{X}}\exp(U(x))$.
- Connection between Markov Random field and Gibbs random field (under appropriate condition) is established by Hammersley-Clifford theorem. (Besag, 1974).
- The network can be viewed as a dependence graph and we can pick different types of configurations (a small subset of possible network links) as relevant to the graph.

Exponential Random Graph Model

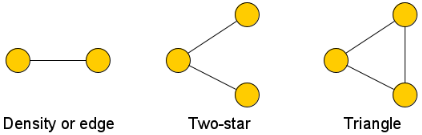
- Specification of ERGM hinges on the choice of network statistics $z(W)$, which reflects the dependence assumption imposed in the model.
- The simplest dependence assumption is dyadic independence, which means that all distinct links are independent of one another. Under dyadic independence, the corresponding model (*Bernoulli* random graph) in the form of ERGM is given by

$$\Pr(W|\theta) = \left(\frac{1}{\kappa(\theta)}\right) \exp\left(\theta_\nu \nu(W)\right),$$

  where $\nu(W) = \sum_{i,j} W_{ij}$ is the number of edges in the network W.
- A homogeneity assumption is imposed here on the effect of each link so that the model only contains one parameter θ_ν, which is called the *edge* or *density* parameter.

Exponential Random Graph Model

- Dyadic independence can be easily violated in social networks.
- To capture the possible transitivity of relationships across social links, a more appropriate dependence assumption is the Markov dependence introduced by Frank and Strauss (1986).
- Under Markov dependence, any two possible network links are dependent when they connect to a common vertex.
- The following Figure shows three network configurations which could be included in a Markov graph.



Exponential Random Graph Model

- The resulting model in a form of ERGM is

$$\Pr(W|\theta) = \left(\frac{1}{\kappa(\theta)}\right) \exp\left(\theta_\nu \nu(W) + \theta_\rho \rho(W) + \theta_\tau \tau(W)\right),$$

where $\rho(W) = 3\sum_i \sum_{k>i} \sum_{j \neq i,k} W_{ij} W_{jk} W_{ki}$ denotes the number of triangles and $\tau(W) = \sum_i \sum_{k>i} \sum_{j \neq i,k} W_{ij} W_{jk}$ denotes the number of two-stars.

Exponential Random Graph Model

Estimating an ERGM:

- The difficulty comes from an intractable normalizing constant in the likelihood function.
- Strauss and Ikeda (1990): the pseudo likelihood maximization approach which maximizes

$$L(\theta) = \sum_{i \neq j} \log\left(\Pr\{W_{ij} = w_{ij} | W_{-ij}\}\right),$$

  where $\Pr\{W_{ij} = 1 | W_{-ij}\} = \frac{\Pr\{W_{ij}=1, W_{-ij}\}}{\Pr\{W_{ij}=0, W_{-ij}\} + \Pr\{W_{ij}=1, W_{-ij}\}}$.
- This estimation method is not reliable because it ignores the dependence between links. But it provides good initial values for other estimation methods.

ERGM

Estimating an ERGM:

- Geyer and Thompson (1992) propose Monte Carlo MLE: the likelihood function is approximated by using m samples $\{W^i\}_{i=1}^m$ generated through MCMC procedure. An iteration process to reach the final estimate. Need good initial values to prevent unconvergence.
- Snijders (2002) uses method of moments – Robbins-Monro algorithm with MCMC simulation. it prevents the convergence problem in MCMLE if the initial value θ_0 is not chosen well.
- Liang (2010) and Caimo and Friel (2011): Bayesian estimation with double Metropolis-Hastings algorithm.

Exponential Random Graph Model

- Example: Caimo and Friel (2011) study Sampson's Monk network.
- They propose the following three-dimensional model

$$P(w) = \frac{\exp\left(\theta_1 s_1(w) + \theta_2 s_2(w) + \theta_3 s_3(w)\right)}{z(\theta)},$$

where $s_1(w) = \sum_i w_{ij}$ (number of edges),
$s_2(w) = \sum_{i \neq j} w_{ij} w_{ji}$ (number of mutual edges),
$s_3(w) = \sum_{i \neq j} w_{ij} w_{jk} w_{ki}$ (number of cyclic triads).

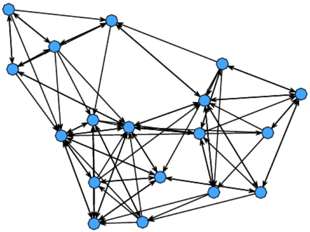


Figure 16: Sampson's Monk network.

Exponential Random Graph Model

	Parameter	Post. Mean	Post. Sd.
Chain 1	θ_1 (edges)	-1.72	0.31
	θ_2 (mutual)	2.34	0.43
	θ_3 (ctriple)	-0.05	0.16
Chain 2	θ_1 (edges)	-1.73	0.30
	θ_2 (mutual)	2.38	0.42
	θ_3 (ctriple)	-0.02	0.16
Chain 3	θ_1 (edges)	-1.74	0.29
	θ_2 (mutual)	2.37	0.43
	θ_3 (ctriple)	-0.04	0.15
Chain 4	θ_1 (edges)	-1.72	0.29
	θ_2 (mutual)	2.33	0.44
	θ_3 (ctriple)	-0.06	0.16
Chain 5	θ_1 (edges)	-1.73	0.30
	θ_2 (mutual)	2.30	0.43
	θ_3 (ctriple)	-0.06	0.16
Chain 6	θ_1 (edges)	-1.72	0.30
	θ_2 (mutual)	2.27	0.44
	θ_3 (ctriple)	-0.06	0.16
Overall	θ_1 (edges)	-1.72	0.30
	θ_2 (mutual)	2.33	0.43
	θ_3 (ctriple)	-0.04	0.16

Table 7: Monks dataset: Summary of posterior parameter density of the model (14).

Exponential Random Graph Model

Example of using R package "statnet" to estimate ERGM

- install.packages('statnet')
- library(statnet)
- data(package='ergm')
- data(florentine)
- flomarriage
- plot(flomarriage, main="Florentine Marriage", cex.main=0.8)
- summary(flomarriage $\sim$ edges+triangle)
- flomodel $\leftarrow$ ergm(flomarriage edges+triangle)
- summary(flomodel)

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- We propose economic (and econometric) modeling approaches to unify network formation and network interactions.
- Our economic model is a two-stage network formation game in which the network is formed in the first stage and then individuals choose the intensity of their economic activities in the second stage.
- Our econometric model is formulated based on the equilibrium of the network formation game which allows us to estimate structural parameters in the utility function.
- Two advantages of our approach:
  - correct endogenous friendship selection biases on the estimate of network interaction effect.
  - allow researchers to evaluate which outcomes (activities) provide important incentives for individuals to form friendship.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- Assume individuals make their decisions on friendship links and economic activities sequentially in a **two-stage** game.
- We focus on a "sub-game perfect equilibrium": individuals anticipate the second stage of the game when choosing the network.
- Formation of network, i.e., the first stage, follows the literature regarding the stability and efficiency of social networks (e.g., Jackson and Wolinsky, 1996; Dutta and Jackson, 2000; Jackson, 2005; Caulier et al., 2015) – accordingly, we adopt a **transferable utility** framework that allows individuals to make side payments.
- In the second stage, we follow the literature dealing with games on networks to characterize individual's choice of activity intensity by the Nash equilibrium.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- The preference of any given individual i is represented by the utility function:

$$U_i(W, Y_1, ..., Y_{\bar{d}}) = v_i(W) + \sum_{d=1}^{\bar{d}} \delta_d u_{i,d}(y_{i,d}, Y_{-i,d}, W),$$

- $v_i(W)$ represents an explicit preference over the network structure W.
- $u_{i,d}(y_{i,d}, Y_{-i,d}, W)$ is the utility derived from activity d.
- Coefficient $\delta_d \geq 0$ captures the relative importance (or weight) of the utility of activity d with respect to the utility of the network $v_i(W)$. We call this **incentive effect** of activity d on network formation.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- Individual i's (explicit) preference for the network structure

$$v_i(W) = \underbrace{\sum_{j=1}^{m} w_{ij}\psi_{ij}}_{\text{local network effects}} + \underbrace{\varpi_i(w_i, W_{-i})\eta}_{\text{global network effects}} + \tau_{i,W},$$

where $\tau_{i,W}$ is an idiosyncratic shock on the value of the network W for individual i.

- We assume a linear quadratic specification for the utility of activity conditional on network structure:

$$u_{i,d}(y_{i,d}, Y_{-i,d}, W) = \mu_{i,d} y_{i,d} - \frac{1}{2} y_{i,d}^2 + \lambda_d y_{i,d} \sum_{j=1}^{m} w_{ij} y_{j,d},$$

where $\mu_{i,d}$ captures individual exogenous heterogeneity.
$\lambda_d \geq 0$ ($\lambda_d < 0$) reflects a complementary (or competitive) effect from peers' activity intensities.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

Definition

An (sub-game perfect) equilibrium of the two-stage game is a collection $(W, Y_1, ..., Y_{\bar{d}})$ such that:

1. $(Y_1, ..., Y_{\bar{d}})$ is in a Nash equilibrium, conditional on W. We denote such an equilibrium by $(Y_1^*(W), ..., Y_{\bar{d}}^*(W))$.
2. The network value

$$T_{Y^*}(W) = \sum_i U_i(W, Y_1^*(W), ..., Y_{\bar{d}}^*(W))$$

is strongly efficient and individually stable under some allocation rules.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

Proposition

Assume that $|\lambda_d| < 1/\|W\|_\infty$*, where* $\|.\|_\infty$ *denotes the maximum row sum norm for all* $d = 1, ..., \bar{d}$ *and that* $\tau_W \equiv \sum_i \tau_{i,W}$ *is distributed according to a Type-I extreme value distribution. Then, there exists a (generically) unique equilibrium of the two-stage game. Moreover,*

(i) *for all d, such that* $y_{i,d} \in \mathbb{R}$*, we have:*

$$Y_d^*(W) = (I_m - \lambda_d W)^{-1} \mu_d,$$

where I_m *is an* $m \times m$ *identity matrix and* $\mu_d = (\mu_{1,d}, \cdots, \mu_{m,d})'$*.*

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

Proposition

(ii) *The equilibrium value is given by $T_{Y^*}(W) = V(W) + \tau_W$, where:*

$$\begin{aligned} V(W) &= \sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\psi_{ij} + \sum_{i=1}^{m} \varpi_i(w_i, W_{-i})\eta \\ &+ \sum_{d=1}^{\bar{d}} \delta_d \left[\mu_d' Y_d^*(W) - \frac{1}{2} Y_d^{*\prime}(W) Y_d^*(W) + \lambda_d Y_d^{*\prime}(W) W Y_d^*(W) \right]. \end{aligned}$$

Therefore, the probability of W at equilibrium is given by:

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\widetilde{W}\in\Omega} \exp\{V(\widetilde{W})\}},$$

where Ω is the set of all $m \times m$ network matrices.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- We specify individual exogenous heterogeneity via $\mu_{i,d} = x_i\beta_{1d} + \sum_{j=1}^{m} w_{ij}x_j\beta_{2d} + z_i\rho_{1d} + \sum_{j=1}^{m} w_{ij}z_j\rho_{2d} + \alpha_d + \epsilon_{i,d}$.
- Following Proposition, the equilibrium for uncensored activities is given by

$$Y_d^*(W) = (I_m - \lambda_d W)^{-1} \left(X\beta_{1d} + WX\beta_{2d} + Z\rho_{1d} + WZ\rho_{2d} + I_m\alpha_d + \epsilon_d \right).$$

- It matches the reduced form of the spatial autoregressive (SAR) model (Bramoulle et al., 2009; Lee et al., 2010; Lin, 2010) for studying social interactions. The coefficient λ_d represents the endogenous (peer) effect, which has been the focus of recent literature due to its policy implications (Glaeser et al., 2003).
- We also consider a censored activity case in the paper.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- Recall we have

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\widetilde{W}} \exp\{V(\widetilde{W})\}},$$

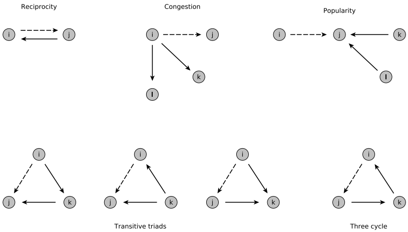
$$V(W) = \sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\psi_{ij} + \sum_{i=1}^{m} \varpi_i(w_i, W_{-i})\eta + \frac{1}{2}\sum_{d=1}^{\bar{d}} \delta_d Y_d^{*\prime}(W) Y_d^{*}(W).$$

- local network effects:

$$\psi_{ij} = \gamma_0 + c_i\gamma_1 + c_j\gamma_2 + c_{ij}\gamma_3 + \sum_{\ell=1}^{\bar{\ell}} \gamma_{4\ell}|z_{i\ell} - z_{j\ell}|.$$

- The variables $|z_{i\ell} - z_{j\ell}|$ for $\ell = 1, \cdots, \bar{\ell}$ are meant to capture the homophily on *unobserved* characteristics.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

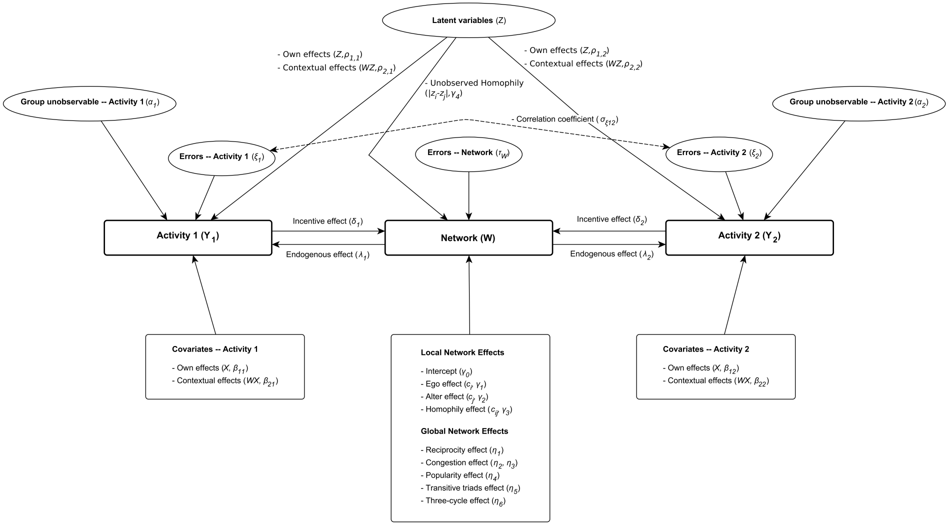


- global network effects:

$$\varpi_i(w_i, W_{-i})\eta = \sum_{j=1}^{m} w_{ij} \Big\{ \underbrace{\eta_1 w_{ji}}_{\text{reciprocity effect}} + \underbrace{\eta_2 \sum_{k\neq j}^{m} w_{ik} + \eta_3 \left(\sum_{k \neq j}^{m} w_{ik} \right)^2}_{\text{congestion effect}} + \underbrace{\eta_4 \sum_{k \neq i}^{m} w_{kj}}_{\text{popularity effect}}$$

$$+ \underbrace{\sum_{k}^{m} \eta_{51} w_{ik} w_{kj} + \eta_{52} w_{ki} w_{kj} + \eta_{53} w_{ik} w_{jk}}_{\text{transitive triads effects}} + \underbrace{\eta_6 \sum_{k}^{m} w_{jk} w_{ki}}_{\text{three-cycle effect}} \Big\}.$$

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)



Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

The joint probability function of the variable, Y_{cg}, and the network, W_g, for group $g = 1, \cdots, G$, can be written as

$$
\begin{aligned}
P(W_g, Y_g|\theta_g, \alpha_g, Z_g) &= P(Y_g|W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g, W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|\theta_g, \alpha_g, Z_g) \cdot P(W_g|\xi_g, \theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|\theta_g, \alpha_g, Z_g) \cdot \frac{\exp(\mathrm{V}(W_g, \xi_g, \theta_g, \alpha_g, Z_g))}{\sum_{\widetilde{W}_g \in \Omega_g} \exp\left(\mathrm{V}(\widetilde{W}_g, \xi_g, \theta_g, \alpha_g, Z_g)\right)},
\end{aligned}
$$

where $S_g(W_g) = I_{m_g} - \lambda W_g$,

$\xi_g = S_g(W_g)Y_g - X_g\beta_1 - W_gX_g\beta_2 - Z_g\rho_1 - W_gZ_g\rho_2 - I_{m_g}\alpha_g$,

$$
f(\xi_g|\theta_g, \alpha_g, Z_g) = (2\pi)^{-\frac{m_g}{2}} \left(\sigma^2_{\xi_g}\right)^{-\frac{m_g}{2}} \exp\left(-\frac{1}{2\sigma^2_{\xi_g}}\xi_g'\xi_g\right).
$$

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- The probability function of $y = (\{Y_{cg}\}, \{W_g\})$ given parameter θ has form $P(y|\theta) = f(y;\theta)/D(\theta)$, where $D(\theta)$ is an intractable normalizing term.
- Denote $\pi(\theta)$ as the prior probability of θ. When generating a new parameter $\tilde{\theta}$ from a proposal distribution $q(\cdot|\theta)$, the acceptance probability α of the M-H algorithm is NOT calculable,

$$
\begin{aligned}
\alpha_{MH}(\theta_{new}, \theta_{old}) &= \min\left\{1, \frac{P(\theta_{new}|y)q(\theta_{old}|\theta_{new})}{P(\theta_{old}|y)q(\theta_{new}|\theta_{old})}\right\} \\
&= \min\left\{1, \frac{\pi(\theta_{new})f(y;\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})f(y;\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{D(\theta_{old})}{D(\theta_{new})}\right\}.
\end{aligned}
$$

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- The acceptance probability α of the double M-H algorithm by Liang (2010): simulate auxiliary variable $\widetilde{y}$ from $P(\widetilde{y}|\theta_{new}) = f(\widetilde{y};\theta_{new})/D(\theta_{new})$ using m runs of M-H algorithm,

$$\begin{aligned}\alpha_{MH}(\widetilde{y},\theta_{new},\theta_{old}) &= \min\left\{1, \frac{\pi(\theta_{new})P(y|\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})P(y|\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{P(\widetilde{y}|\theta_{old})}{P(\widetilde{y}|\theta_{new})}\right\} \\ &= \min\left\{1, \frac{\pi(\theta_{new})f(y;\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})f(y;\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{f(\widetilde{y};\theta_{old})}{f(\widetilde{y};\theta_{new})}\right\},\end{aligned}$$

where $\widetilde{y}$ are simulated from the likelihood function $P(\widetilde{y}|\theta_{new}) = f(\widetilde{y};\theta_{new})/D(\theta_{new})$ with the exact sampling (Propp and Wilson, 1996).

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

Table: Results of Monte Carlo experiments for the uncensored activity variable.

		Full		No latent		No global		Latent only		Activity alone	
Parameter	True	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
λ	0.0500	0.0500	0.0073	0.0679	0.0091	0.0167	0.0052	0.0595	0.0091	0.0846	0.0073
β_1	0.3000	0.2817	0.0280	0.2986	0.0283	0.2992	0.0299	0.3038	0.0271	0.2965	0.0281
β_2	0.1000	0.0927	0.0099	0.1036	0.0196	0.1084	0.0116	0.1317	0.0185	0.1243	0.0214
ρ_1	0.3000	0.2838	0.0605	-	-	0.2359	0.0689	0.2713	0.0923	-	-
ρ_2	0.1000	0.0946	0.0248	-	-	0.0696	0.0316	0.0864	0.0685	-	-
γ_0	-0.5000	-0.5262	0.0655	-1.9885	0.1387	-1.4266	0.0617	-1.0418	0.0362	-	-
γ_3	0.3000	0.3106	0.0505	0.2733	0.0538	0.3519	0.0723	0.3620	0.0611	-	-
γ_4	-1.0000	-1.1367	0.0715	-	-	-1.1137	0.0784	-1.0580	0.0514	-	-
η_1	0.3000	0.2959	0.0405	0.4401	0.0499	-	-	-	-	-	-
η_2	0.2000	0.2172	0.0475	0.3101	0.0602	-	-	-	-	-	-
η_3	-0.1000	-0.1003	0.0084	-0.1098	0.0100	-	-	-	-	-	-
η_4	0.0400	0.0418	0.0036	0.0325	0.0042	-	-	-	-	-	-
η_5	0.3000	0.3023	0.0241	0.4015	0.0285	-	-	-	-	-	-
η_6	-0.2000	-0.1976	0.0216	-0.1777	0.0273	-	-	-	-	-	-
δ	0.3000	0.3000	0.0259	0.2094	0.0569	1.0049	0.1001	-	-	-	-
σ_ξ^2	0.5000	0.5381	0.1759	0.6852	0.1843	0.6647	0.2403	0.5259	0.1764	0.6740	0.1784

Note: This Monte Carlo study consists of 100 repetitions. The values reported in this table are the mean and the standard deviation of parameter estimates across repetitions. In each repetition, we estimate each of the corresponding models with 50,000 MCMC draws. We drop the first 10,000 draws due to burn-in and calculate the (posterior) mean of the remaining draws as parameter estimates.

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- We study high school students' friendship networks and two outcome variables, GPA and smoking frequency, with the Add Health data in U.S.
- Our empirical findings show that
  - Exogenous dyad-specific effects of age, sex, and race are significant.
  - there are significant homophily effects from individual unobservables.
  - network structure effects matter, especially reciprocality.
  - the interaction benefit resulted from GPA provides a significant incentive for forming friendships; while the interaction benefit from smoking does not.
  - both GPA and smoking frequency are subject to significant network interactions (peer effects).

Network Formation and Interactions with Exponential Distribution (Hsieh et al., 2019)

- The Add Health data:
  - A nationally representative study of adolescents' health-related behaviors which covers students in 7^{th} to 12^{th} grade from 132 schools.
  - Each respondent was asked to nominate up to five male and five female friends which provides information about students' friendship networks
- Assign respondents into groups defined by respondents' school.
- We only look at small school which consist of number of students less than 120 to ease the computational burden.
- We consider two activity variables: GPA (uncensored variable) and smoking frequency (censored variable).

Table: Summary statistics

Variable	Min	Max	Mean	SD
GPA	1	4	3.059	0.742
Smoking	0 (73.26%)	7	0.543	1.715
Age	11	18	13.788	1.641
Male	0	1	0.457	0.498
Female	0	1	0.543	0.498
White	0	1	0.741	0.438
Black	0	1	0.128	0.335
Asian	0	1	0.017	0.131
Hispanic	0	1	0.050	0.218
Other race	0	1	0.063	0.243
Both parents	0	1	0.821	0.383
Less HS	0	1	0.069	0.254
HS	0	1	0.326	0.469
More HS	0	1	0.460	0.499
Edu missing	0	1	0.094	0.291
Professional	0	1	0.273	0.446
Staying home	0	1	0.230	0.421
Other jobs	0	1	0.375	0.484
Job missing	0	1	0.069	0.254
Welfare	0	1	0.002	0.044
Num. of other students at home	0	6	0.591	0.850
Network size	29	101	69.29	23.96
Network density	0.016	0.136	0.076	0.062
Out-degree	0.000	10.000	3.752	2.688
Clustering coefficient	0.025	0.186	0.095	0.048
Sample size		1,036		
Num. of networks		15		

	Full	No latent	No global	Latent only	Activity alone
Network Interaction					
Endogenous (GPA)	0.0177***	0.0189***	0.0162**	0.0239**	0.0330***
	(0.0063)	(0.0070)	(0.0071)	(0.0091)	(0.0105)
Endogenous (smoking)	0.1052***	0.1103***	0.1068***	0.1056***	0.1125***
	(0.0196)	(0.0225)	(0.0193)	(0.0197)	(0.0191)
Network Formation					
Constant (γ_0)	-4.8531***	-5.7071***	-3.9329***	1.1101***	-
	(0.0366)	(0.1194)	(0.1134)	(0.0958)	-
Experience in school (sender) (γ_1)	-0.0405***	-0.0159	0.0892***	0.0528***	-
	(0.0114)	(0.0212)	(0.0211)	(0.0148)	-
Experience in school (receiver) (γ_2)	0.0374**	0.0515***	0.1586***	0.1301***	-
	(0.0143)	(0.0198)	(0.0247)	(0.0147)	-
Same age (γ_{31})	0.3675***	0.5340***	1.0300***	0.4450***	-
	(0.0431)	(0.0496)	(0.0773)	(0.0586)	-
Same sex (γ_{32})	0.3471***	0.4889***	0.3454***	0.5477***	-
	(0.0441)	(0.0608)	(0.0686)	(0.0517)	-
Same race (γ_{33})	0.3116***	0.2152***	0.3716***	0.3961***	-
	(0.0430)	(0.0789)	(0.0839)	(0.0595)	-
Latent distance (γ_{41})	-0.1469***	-	-0.3945***	-3.9387***	-
	(0.0173)	-	(0.0671)	(0.1214)	-
Latent distance (γ_{42})	-0.0966**	-	-0.2231***	-2.5592***	-
	(0.0414)	-	(0.0662)	(0.1000)	-
Latent distance (γ_{43})	-0.0125	-	-0.0868	-2.3793***	-
	(0.0528)	-	(0.0594)	(0.0887)	-
Reciprocity (η_1)	1.4309***	1.4080***	-	-	-
	(0.0476)	(0.0552)	-	-	-
Congestion (η_2)	0.2699***	0.3521***	-	-	-
	(0.0151)	(0.0281)	-	-	-
Congestion (η_3)	-0.0247***	-0.0304***	-	-	-
	(0.0017)	(0.0022)	-	-	-
Popularity (η_4)	0.0049	0.0015	-	-	-
	(0.0068)	(0.0061)	-	-	-
Trans. triads (η_5)	0.4715***	0.4767***	-	-	-
	(0.0220)	(0.0189)	-	-	-
Three cycles (η_6)	-0.2071***	-0.2083***	-	-	-
	(0.0171)	(0.0166)	-	-	-
Incentive from GPA (δ_1)	0.2145**	0.1825***	0.2921***	-	-
	(0.0956)	(0.0344)	(0.1122)	-	-
Incentive from smoking (δ_2)	0.0197	0.0083	0.0214	-	-
	(0.0134)	(0.0053)	(0.0131)	-	-

Note: The MCMC runs for 150,000 iterations and the first 50,000 runs are dropped for the burn-in. Values in parentheses are

Reference

Airoldi, Edoardo M, David M Blei, Stephen E Fienberg, and Eric P Xing (2008) "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, Vol. 9, pp. 1981–2014.

Besag, Julian (1974) "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (2008) "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, p. P10008.

Caimo, Alberto and Nial Friel (2011) "Bayesian inference for exponential random graph models," *Social Networks*, Vol. 33, pp. 41–55.

Erdös, P. and A. Rényi (1959) "On Random Graphs, I.," *Publicationes Mathematicae*, Vol. 6, pp. 290–297.

Frank, Ove and David Strauss (1986) "Markov graphs," *Journal of the American Statistical Association*, Vol. 81, pp. 832–842.

Geyer, Charles J and Elizabeth A Thompson (1992) "Constrained Monte Carlo maximum likelihood for dependent data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 657–699.

Gilbert, Edgar N (1959) "Random graphs," *The Annals of Mathematical Statistics*, Vol. 30, pp. 1141–1144.

Girvan, Michelle and Mark EJ Newman (2002) "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, Vol. 99, pp. 7821–7826.

Goldenberg, Anna, Alice X Zheng, Stephen E Fienberg, Edoardo M Airoldi et al. (2010) "A survey of statistical network models," *Foundations and Trends® in Machine Learning*, Vol. 2, pp. 129–233.

Handcock, Mark S, Adrian E Raftery, and Jeremy M Tantrum (2007) "Model-based clustering for social networks," *Journal of the Royal*

Statistical Society: Series A (Statistics in Society), Vol. 170, pp. 301–354.

Hoff, Peter D, Adrian E Raftery, and Mark S Handcock (2002) "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, Vol. 97, pp. 1090–1098.

Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt (1983) "Stochastic blockmodels: First steps," *Social Networks*, Vol. 5, pp. 109–137.

Hsieh, Chih-Sheng and Lung Fei Lee (2016) "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, Vol. 31, pp. 301–319.

Hsieh, Chih-Sheng, Lung-Fei Lee, and Vincent Boucher (2019) "Specification and estimation of network formation and network interaction models with the exponential probability distribution."

Krivitsky, Pavel N, Mark S Handcock, Adrian E Raftery, and Peter D Hoff (2009) "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models," *Social Networks*, Vol. 31, pp. 204–213.

Liang, Faming (2010) "A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants," *Journal of Statistical Computation and Simulation*, Vol. 80, pp. 1007–1022.

Lorrain, Francois and Harrison C White (1977) "Structural equivalence of individuals in social networks," in *Social Networks*: Elsevier, pp. 67–98.

Newman, Mark EJ (2006) "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, Vol. 103, pp. 8577–8582.

Nowicki, Krzysztof and Tom A B Snijders (2001) "Estimation and prediction for stochastic blockstructures," *Journal of the American Statistical Association*, Vol. 96, pp. 1077–1087.

Snijders, Tom AB (2002) "Markov chain Monte Carlo estimation of exponential random graph models," *Journal of Social Structure*, Vol. 3, pp. 1–40.

Snijders, Tom AB and Krzysztof Nowicki (1997) "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," *Journal of Classification*, Vol. 14, pp. 75–100.

Snijders, Tom AB, Philippa E Pattison, Garry L Robins, and Mark S Handcock (2006) "New specifications for exponential random graph models," *Sociological Methodology*, Vol. 36, pp. 99–153.

Strauss, David and Michael Ikeda (1990) "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, Vol. 85, pp. 204–212.

Travers, Jeffrey and Stanley Milgram (1967) "The small world problem," *Phychology Today*, Vol. 1, pp. 61–67.

Wasserman, Stanley and Philippa Pattison (1996) "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp," *Psychometrika*, Vol. 61, pp. 401–425.

Watts, Duncan J and Steven H Strogatz (1998) "Collective dynamics of "small-world" networks," *Nature*, Vol. 393, p. 440.