
Marginally Specified Hierarchical Models for Relational Data

Andrew C. Thomas
Carnegie-Mellon University
Pittsburgh, PA
act@acthomas.ca

Joseph K. Blitzstein
Harvard University
Cambridge, MA

Abstract

We present a unified approach to modelling dyadic relational data, namely that seen in social, biological and technological networks, without restriction to the binary format. The approach involves three principles: considering the marginal specification of any edge as the fundamental unit, embedding as much dependence as possible in latent structural forms, and using distributional forms that favour high-throughput computational methods for their solution. We show that this approach allows for an extremely flexible and generalizable way of describing the structural properties of relational systems; namely, we offer alternate explanations for two approaches popular in the networks literature, the “small-world” and “scale-free” mechanisms, and demonstrate the ability of marginal hierarchical modelling to expand beyond them.

1 Introduction

Many stochastic models for networks assume conditional dyadic independence: given a set of underlying characteristics, the variability of each undirected edge, or of each complementary pair of directed arcs, is unaffected by the effect of other remaining ties. The presentation of conditional dyadic independence cuts off more complicated dependence patterns between dyads at the overt level, but with the exchange that these trends can be more cleanly explained at a level below that of observation.

As motivated, this paper contains a unifying framework for many of these previous approaches that allows for considerable extension. In the binary case, ties are represented as an observed outcome of an underlying continuous process, based primarily on the Gaussian framework but adaptable to other contexts, and the investigator can bring to bear tools developed in computational statistics, dynamic programming, and other connecting literatures in order to efficiently and correctly model these sorts of relational data. This is not the first time the approach has been proposed – similar models have been well-implemented in the work of Peter Hoff and his colleagues [8; 9; 6; 12] – but, to the author’s knowledge, it is the first large-scale attempt to unify the modelling framework for various dyadic relational data types with a wider class of models, largely focused on the GLM framework, and the generalization of computational methods for their analysis.

As the outcomes \mathbf{Y} can be considered entries in an N -by- N matrix, it is standard to group these terms in four entries: grand mean value, row effects, column effects and row-column interactions. For this reason, we begin with a redefinition of the p_1 model from the marginal point of view, capturing the four components: mean density, sender properties, receiver properties and reciprocity

between the arcs. Three properties are then brought into the current framework: latent spaces, latent characteristics and the behaviour of assortative mixing on degree.

2 Marginal Specification and Extension of p_1

The original p_1 model [10] was specified on a series of $\binom{n}{2}$ dyads with quadrinomial probability specifications for each. In the marginal case, there are $2\binom{n}{2}$ arcs to be specified, namely of the form $Y_{ij} \sim Be(p_{ij})$, or in general probit notation, $Y_{ij} \sim Be(\Phi(\mu_{ij}))$. The first simplifying step in p_1 is to simplify this probability into terms representing the grand mean, sender and receiver. This becomes

$$Y_{ij} \sim Be(\Phi(\mu + \alpha_i + \beta_j))$$

so that the terms $\mu + \alpha_i + \beta_j$ represent the same types of quantities as before – the increased likelihood of ties in general, ties from sender i and ties to receiver j – even though their numerical interpretations are slightly different, in terms of their effect on the differing likelihoods.

This formula can be represented in terms of a latent normal variable Z_{ij} , so that the previous expression is equivalent to

$$Y_{ij} \sim \mathbb{I}(Z_{ij} > 0); \quad Z_{ij} \sim N(\mu + \alpha_i + \beta_j, 1).$$

Once this step is made, the conversion from two independent normals to one bivariate normal is immediate, and the dyad (Y_{ij}, Y_{ji}) is now expressed as the realization of a latent bivariate normal:

$$\begin{bmatrix} Y_{ij} \\ Y_{ji} \end{bmatrix} \mid \begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix} = \begin{bmatrix} \mathbb{I}(Z_{ij} > 0) \\ \mathbb{I}(Z_{ji} > 0) \end{bmatrix}; \quad (1)$$

$$\begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix} \mid \alpha, \beta, \rho \sim N_2 \left(\begin{bmatrix} \mu + \alpha_i + \beta_j \\ \mu + \alpha_j + \beta_i \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \quad (2)$$

To compare to the canonical p_1 , the sender and receiver effects can be restricted to have zero sum, $\sum_i \alpha_i = \sum_i \beta_i = 0$. Each node's sender and receiver effects may also come from a common family, as expressed in [20],

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta \\ \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{bmatrix} \right) \quad (3)$$

with appropriate prior distributions on these variances and the correlation term $\rho_{\alpha\beta}$.

A Gibbs sampling scheme, as inspired by [1], is relatively easy to put together. Of special note is the algebra needed to demonstrate the direct draws for the sender, receiver and grand mean effects. Consider the draw for one sender component α_i ; the log-likelihood for a single bivariate normal containing the term, as divided into conditional and marginal pieces, is given as:

$$\begin{aligned} & \log(p(Z_{ij} \mid \alpha, \beta, \mu, \rho)p(Z_{ji} \mid Z_{ij}, \alpha, \beta, \mu, \rho)) \\ &= C - \frac{1}{2}(Z_{ij} - \alpha_i - \beta_j - \mu)^2 - \frac{1}{2(1-\rho^2)}(Z_{ji} - \alpha_j - \beta_i - \mu - \rho(Z_{ij} - \alpha_i - \beta_j - \mu))^2 \\ &= \frac{1}{2}((Z_{ij} - \beta_j - \mu) - \alpha_i)^2 + \frac{\rho}{2(1-\rho^2)} \left(\alpha_i - (Z_{ij} - \beta_j - \frac{1-\rho}{\rho}\mu - \frac{Z_{ji} - \alpha_j - \beta_i}{\rho}) \right)^2, \end{aligned}$$

which is in quadratic form for α_i , conditional on the remaining terms. The addition of either a prior distribution common to all α , or a hierarchical pooling model such as Equation 3, make the conditional draw for the parameter as natural as from a standard distribution.

3 Covariate Inclusion for Senders, Receivers, Edges

As introduced, node effects are modelled as indicators for the presence of a particular individual; for example, the sender effect α_i may also be considered as $\sum_k \alpha_k \delta_{ki}$, to signify the presence of an effective covariate: the indicator that the node being considered corresponds to sender i . From here, it is a simple addition to generalize to other covariates, whether or not they are uniform for all senders, receivers or edges.

The inclusion for covariates on senders, receivers and edges is straightforward:

$$\begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix} \mid \begin{bmatrix} \mu, \alpha, \beta, \rho, \mathbf{X}, \\ \mathbf{W}, \mathbf{U}, \gamma, \nu, \theta \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \mu + (\alpha_i + X_i \gamma_i) + (\beta_j + W_j \nu_j) + U_{ij} \theta \\ \mu + (\alpha_j + X_j \gamma_j) + (\beta_i + W_i \nu_i) + U_{ji} \theta \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right);$$

the steps added to the Gibbs sampler are identical in form to the node effects due to their quadratic form representations.

4 Differential Reciprocity Adjustments

[5] propose an extension of the p_1 model around the notion of differential reciprocity; that is, the tendency for one arc in a dyad to mirror the connection of the other may vary based on the information on the participating nodes. Under the original specification, the reciprocity term was considered as an odds ratio; in the GLM framework, it is considered to be a correlation.

For full specification, consider the Fisher transform $q = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$, so that the transformed quantity q is without bound. Then, the transformed correlation may take the form $q_{ij} = \mu_q + \tau_i + \tau_j + V_{ij} \psi$, so that μ_q represents the baseline reciprocity, and τ_i and τ_j represent the deviations due to each of the two nodes in the dyad, subject to a zero-sum or pooling constraint,

$$\sum_i \tau_i = 0 \quad or \quad \tau_i \sim N(0, \sigma_\tau^2).$$

Covariates V_{ij} can be included for the edge, multiplied by the coefficient vector ψ to produce the observed effect on reciprocity.

By using the inverse transform, $\rho_{ij} = \rho_{ji} = \frac{e^{2q_{ij}} - 1}{e^{2q_{ij}} + 1}$, the parameters are restored to the original $(-1, 1)$ range to act as correlations between each edge in the dyad.

5 Latent Spaces and Parameters

Latent spaces and parameters have been introduced mainly in undirected contexts, but there is little reason why they cannot be integrated into the current approach. Consider first the marginal distribution of a single arc. If there is assumed to be a k -dimensional latent space where increased distance represents a decreased likelihood of connection, where d_i is a k -dimensional vector in the latent space, and the general marginal expression for an arc is $Z_{ij} \sim N(\mu_{ij}, 1)$, then the mean of the latent strength can be expressed as

$$\mu_{ij} \mid \omega, \mathbf{d} = \omega |d_i - d_j|,$$

so that $\omega < 0$ guarantees that greater distance decreases connections. Sampling this model can prove to be troublesome, because there is a nonidentifiability of scale between ω and the position d_i .

This can be standardized with two steps: fixing $\omega = -1$, and fixing one dyad in the latent space: $d_1 = \vec{0}$, and $d_2 = (1, 0, \dots, 0)$. If desired, further constraints can be placed on all of the first k nodes.

The issue becomes one of multimodality. The act of compressing n nodes into a k -dimensional space will ensure that there will be an exceedingly large number of local modes in the system, since given the other nodes, each node will have at least one locally preferred location, even if the other nodes are not themselves optimally placed. One solution to this problem is to incorporate a simulated annealing ladder into the maximization routine, so that the local nodes are free to sort themselves on a rough scale in the early iterations of the procedure, increasing the likelihood of finding a preferred global configuration.

Once this is done, it is a simple matter to add these latent positions into the Gibbs sampler through a Metropolis step: propose a random step in the latent space, then accept the new position if a uniform random variable is below the ratio of the new posterior probability over the original.

The construction of latent parameters has a similar issue. For a k -dimensional parameter space, the latent strength is expressed as

$$\mu_{ij}|z, C = z_i' C z_j,$$

where z_i is a length- k vector and C is a k -by- k matrix of magnitudes. This can be interpreted as the inner product between character vectors z_i and z_j with respect to the Euclidean space transformed by C , but with one important addition: the diagonal elements of C can be negative, implying that the coordinate is heterophilic, as opposed to a positive value implying homophily on the latent characteristic.

6 Assortative Mixing on Popularity and Gregariousness, Rather than Degree

An observation that has been observed in real networks is the notion of assortative mixing: individuals with similar numbers of ties are more likely to associate with each other than would otherwise be expected by their own gregariousness or popularity, even though it is reasonable to expect individuals with a large number of ties to connect to each other with great likelihood. If this is the case, it is likely that additional forces are at work.¹

Newman [15] measures assortative mixing within a network as a descriptive statistic: a coefficient of the correlation between the joint degree distribution of two connected nodes and the degree of nodes in the marginal sense, then normalized with respect to the maximum value. Consider the measure of “remaining degree” of one node ($d_i - 1$), and the joint distribution of two connected nodes ($(d_i - 1), (d_j - 1)$). The assortativity is defined as the correlation between the joint remaining degree probability of a pair of nodes and their marginal remaining degree probabilities, with respect to each edge in the system; that is, nodes with higher degree have a higher tendency to contribute to the mixing statistic. As this is a statistical description, the inclusion of this behaviour in a generative model requires a corresponding parameter.

Consider the p_1 -type model

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

as a starting point, where α and β have mean 0 and the error term $\varepsilon_{ij} \sim N(0, 1)$. To alter the level of assortative mixing, the parameter χ is introduced and an additional term is included, directly proportional to the popularity and gregariousness of the individuals:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \chi\alpha_i\beta_j + \varepsilon_{ij}.$$

¹In particular, the fact that nodes have the appearance of organizing according to their network structure represents an endogeneity in the modelling step that static generative models may have difficulty in handling.

| $\mathbf{Y} \mathbf{T}$ | $= f(T^1, T^2, \dots)$: Parameter | Symbol |
|-------------------------|---|--|
| $T_{ij}^k =$ | global mean | μ |
| + | sender covariate term(i) | $\alpha_i + X_i\gamma_i$ |
| + | receiver covariate term(j) | $\beta_j + W_j\nu_j$ |
| + | sender/receiver mixing term(ij) | $\chi(\alpha_i + X_i\gamma_i)(\beta_j + W_j\nu_j)$ |
| + | arc covariate term(ij) | $U_{ij}\theta + \varepsilon_{ij}$ |
| + | latent geometric term(ij) | $- d_i - d_j $ |
| + | latent property term(ij) | $z_i' Cz_j$ |
| Definitions | $\mu, \alpha, \beta, \gamma, \nu, \chi, \theta, \omega$ | Effects (fixed, random, mixed) |
| | X_i, W_j, U_{ij} | Covariates |
| | d_i | (Latent) position |
| | z_i | Latent characteristic vector |
| | C | Latent characteristic factor matrix |
| | ε_{ij} | Noise or Overdispersion |

Table 1: The framework for all GLM network estimation, with broad definitions of each term involved. Each of the terms in the general functional framework can be composed in terms of these effect groupings. The function $f(T^1, T^2, \dots)$ may be deterministic or stochastic.

As the sender and receiver terms are naturally centered at zero, there are four regimes to consider: when each of these terms is greater or less than zero respectively. Positive values of χ raise the tie strength when α_i and β_j have the same sign, and lower for opposite signs, the key characteristic of assortative mixing; likewise, negative values for χ lower the tie strength for opposite-signed gregariousness and popularity in the individuals for this particular arc.

6.1 Additional Extensions

The present form allows for a great deal of expansion to other phenomena and other forms of data:

- Robust analyses with the multivariate t distribution, in place of the Gaussian. This implementation is a trivial addition to most Gibbs sampling algorithms through data augmentation.
- Normal-family outcomes are immediately derivable.
- Partial correlations as network ties can be implemented as inverse Fisher-transformed normal outcomes.
- Count data and ordinal data can have modified forms for the outcome.

These implementations are presented in other works.

6.2 The General Case

The cases presented have common roots: each expression required for the stochastic generation of the relational structure can be decomposed into grand mean, sender, receiver and interaction terms. These terms are summarized in Table 1.

As dyadic data, the pair (Y_{ij}, Y_{ji}) are taken together as a unit and may share many characteristics. They may be independent given their characteristics, or dependent under a chosen framework like a Generalized Estimating Equations method, the aforementioned bivariate probit, or a more general latent copula formulation [11; 16; 17; 7].

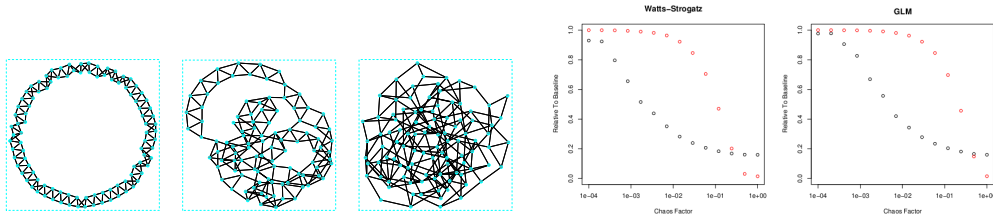


Figure 1: A series of networks formed by the Watts-Strogatz small-world algorithm. Left three: an “ordered” lattice, in which every node is connected to its two nearest neighbours; a small number of rewirings is permitted, maintaining the close connection of neighbours but decreasing the geodesic path lengths; continued rewiring of connections at random. Right two: The median path length and clustering statistics for small-world graphs generated by the Watts-Strogatz and GLM methods. Each display the region between the curves where the graphs have small-world properties.

7 Reformulations of Classic Examples, with Extensions

Several popular approaches from outside the statistical literature are built around generative schemes that propose to explain how real networks came to be in existence. However, it is all too easy to confuse the map with the territory – in this case, the mistake of accepting a proposed generative model for the network both as the best (and possibly only) story, and as predictive of future growth and of similar networks – when alternate explanations are available. It is for these reasons that we demonstrates the applicability of a workhorse GLM approach to model the same circumstances as described by small-world and preferential-attachment mechanisms.

7.1 Watts-Strogatz “Small-World” Networks

While studying the mechanisms of coupled harmonic oscillations in biological networks, [19] identified a structural class of networks now known as “small world” networks, as inspired by the work of social psychologist Stanley Milgram in the 1960s [14], which itself was also the source of the expression “six degrees of separation”. While the initial work of Watts and Strogatz focused on structural aspects, in later work the notion was generalized as being an interpolation between an “orderly” ring lattice and a “chaotic” Erdős-Rényi random graph, so that a ring lattice had a fraction of its ties rewired.

This can be easily described from the GLM approach. Begin with n nodes equally positioned around a circle with circumference n . Let s_{ij} be the distance between nodes i and j along the circle.² Let Y_{ij} be drawn from a Bernoulli $\{0, 1\}$ random variable with probability of success as the sum of two pieces.

First, there is the propensity to connect to an immediate neighbour. In the complete “order” case, an individual connects with probability one to the closest connections, those within a distance of $\lfloor k \rfloor = \lfloor \frac{(n-1)d}{2} \rfloor$, and with proportional probability if just outside this range. That is, if $k = 2.5$, then the nearest two nodes on each side would be connected, and those a distance three away would connect with one-half probability, and with probability zero for any nodes farther away. All together, this represents a success probability

$$o_{ij} = \delta(s_{ij} \leq \lfloor k \rfloor) + (k - \lfloor k \rfloor)\delta(0 < s_{ij} - \lfloor k \rfloor \leq 1).$$

²Other distance functions may possibly be substituted here to produce different network topologies; the ring structure is presented to maintain consistency with the original model.

Second, there is the piece more indicative of chaotic behaviour. In the standard small-world model, this takes the form of an Erdős-Rényi probability, so that the probability of connection is proportional only to the density, defined as

$$c_{ij} = d.$$

The small-world GLM is then composed by weighing the order and chaos probabilities according to the factor p , so that

$$Y_{ij} \sim Be((1-p)o_{ij} + pc_{ij}).$$

This simultaneous model with a latent space is sufficient to explain small-world characteristics as well as the original Watts-Strogatz model. However, the ability to expand this model beyond these characteristics, not the least of which is the addition of nodal properties, is an advantage that the GLM approach has over the original. These characteristics are on displayed at various levels of p in Figure ???. As the degree of “chaos” increases with rising p , there is a region where the degree of local clustering remains high, while the median distance between points decreases markedly, for both the new GLM model and the original small-world model of [19].

7.2 Preferential Attachment Models

As many networks form through a process of aggregation, there is great interest in explaining a network’s structure through a process of evolution. The mechanism proposed in [2], largely known today by the term “preferential attachment”, follows this general mechanism:

1. Begin with a small collection of k connected nodes (a “seed” network) with some configuration of ties between them. Make note of the degree of each node, $d_i = \sum_j Y_{ij}$.
2. Add a new node labelled $k + 1$ to the system, and create a link with one of the current k nodes with respective probabilities $p_j = \frac{d_j}{\sum_i d_i}$, proportional to the degree of each of the nodes at this time.
3. Repeat step 2, updating the degree distribution with each step.

These methods represent the evolution of a system in which the active age of a node is partly responsible for its propensity to have ties attached. But it is also reasonable to model this association as a function of the intrinsic popularity of a node. For example, a system of nodes whose popularities are heterogeneous can be generated as $\beta_j \sim N(\mu, \sigma^2)$, so that μ is the mean popularity and σ^2 the heterogeneity between nodes; the subsequently generated popularities are then the basis for the generation of a directed graph. In keeping with previous examples and motivations, a probit link with data augmentation is used to obtain the directed graph; a symmetrized version of this graph is produced to get the undirected equivalent, so that $Y_{ij} = \max(\mathbb{I}(Z_{ij} > 0), \mathbb{I}(Z_{ji} > 0))$.

The next section shows that generative systems with this level of heterogeneity can produce systems that have the signature characteristics of the scale-free model, in common with the growth-plus-preferential-attachment mechanism, yet also have additional interesting properties.

7.2.1 Example: A Simultaneous Heterogeneous Degree Network

To produce a network with similar “obvious” characteristics to a preferential attachment modelled network with n total nodes, a large number N of networks are simulated under the preferential attachment mechanism and recording the degree distribution for each, and taking the mean at each position; that is, let

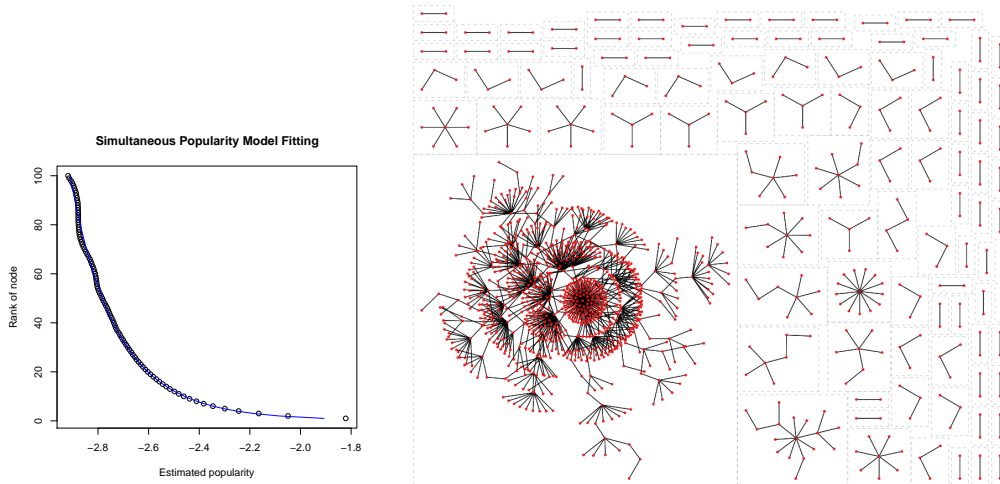


Figure 2: Left: Fitting the expected degree of each node under a GLM model to the observed degrees of networks constructed by preferential attachment. Right: A representative simulation of the GLM model for scale-free-type networks. 1221 nodes with connections are shown, 906 of which are in the largest component.

$$D_m = \frac{1}{N} \sum_k d_k(m)$$

be the mean degree of the m^{th} most popular node across all simulations.

Correspondingly, as node m should have D_m incoming connections, and the probability of any one incoming connection is $\Phi(\beta_m)$, there is an estimate for each individual $\beta_m = \Phi^{-1}(D_m/(n-1))$. A curve is fit to the cumulative distribution curve (see Figure 2 Left), corresponding to the expected fraction of edges per node for various ranges of beta using a simple least-squares criterion.

Three quantities are obtained in the fit: the mean shift, the standard deviation (scale), and the total number of points that would make the completed curve – including the addition of nodes with degree zero that would be unaccounted for with an algorithm that guarantees a fully connected graph. Notably, to get a system with 1000 nodes in the largest component, a base population of 5000 is required; if this is the generative mechanism for the system, then the preferential attachment model will miss them. A simulation from the GLM model is plotted in Figure 2 to show the similarities between it and the original model. The largest component in this simulation has 906 nodes; there are an additional 315 nodes that are connected in some way to others, leaving 3879 isolated and undisplayed nodes.

The dynamic properties of this class of system are considerably different from preferential attachment, not the least of which is the likelihood that new nodes are not guaranteed to join the giant component. This generative scheme can capture many of the same features of an observed network for which preferential attachment is a plausible mechanism, while at the same time noting that the use of a preferential attachment model would not indicate the presence of so many disconnected nodes that are no less involved in the system under study.

References

- [1] J.H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.
- [2] AL Barabasi and R Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [3] Derek de Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27:292–306, 1976.
- [4] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Advances in Physics*, 51(4):1079–1187, 2002.
- [5] S. Fienberg and S. Wasserman. Categorical data analysis of single sociometric relations. *Sociological methodology*, 11:156–192, 1981.
- [6] Peter D. Hoff. Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 295(469):286–295, 2005.
- [7] Peter D. Hoff. Extending the rank likelihood for semiparametric copula estimation. *Annals of Applied Statistics*, 1(1):265–283, February 2007.
- [8] Peter D. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. *Neural Information Processing Systems*, 20, 2007.
- [9] Peter D. Hoff and Michael D. Ward. Modeling dependencies in international networks. Technical report, University of Washington, 2003.
- [10] P.W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–65, 1981.
- [11] Chris A. J. Klaassen and Jon A. Wellner. Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, 3(1):55–77, 1997.
- [12] Pavel N. Krivitsky, Mark S. Handcock, Adrian E. Raftery, and Peter D. Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(3):204–213, 2009.
- [13] C. Liu. Robit regression: A simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pages 227–238. Wiley, 2004.
- [14] S Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.
- [15] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89:208701, 2002.
- [16] Michael Pitt, David Chan, and Robert Kohn. Efficient bayesian inference for gaussian copula regression models. *Biometrika*, 93(3):537–554, 2006.
- [17] W. T. Shaw and K.T.A. Lee. Copula methods vs canonical multivariate distributions: the multivariate student t distribution with general degrees of freedom. Technical report, King’s College London, 2007.
- [18] H. A Simon. On a class of skew distribution functions. *Biometrika*, 42:425, 1955.
- [19] D.J. Watts and S.H. Strogatz. Collective dynamics of “small-world” networks. *Nature*, 393:440–442, 1998.
- [20] George Y. Wong. Bayesian models for directed graphs. *Journal of the American Statistical Association*, 82(140-148):397, 1987.