Marginally Specified Hierarchical Models for Relational Data

Andrew C. Thomas and Joseph K. Blitzstein

October 23, 2009

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 Networks and Relational Data

A network, defined as a collection of individuals (or "nodes") who are connected in a pairwise fashion (with undirected "edges" or directed "arcs"), is a powerful tool for describing many types of complex systems across many fields of science, nature, technology and society. Because of the relative ease in perceiving this type of system, network constructions have been used both to describe systems that literally meet these criteria: computer networks have members that are individual machines (literally known as nodes), connected by wire or wirelessly to each other in pairwise fashion, often through hubs with high connectivity. As a result, solutions for the most economical configuration of nodes and connections for the circumstance is, without exaggeration, a multi-billion-dollar industry.

For other systems, a network model is often imposed onto the ensemble of individual components and their interactions. In correlational models, two individuals that have a high correlation in their outcomes are said to be connected – say, two people have breakfast at the same restaurant over a weeklong period – and a network is formed by considering all correlations between pairs of nodes. In this case, the network model may be a poor approximation if there is substantial

interaction at a level higher than pairwise; one person invites two friends to breakfast, and the three eat together, but the friends are unconnected except for their common acquaintance.

This work considers all classes of valued, directed networks under the heading of *relational data*: n individual units (or nodes) are connected pairwise by a maximum of $n(n-1)$ directed edges (or relations). When assigned value, the edges combine with the nodes into the ensemble configuration of a network. It is the specification of these edges, including their values and uncertainties, that form the basis of the investigation that follows.

Whether the network has a physical basis, or is simply an approximation for the entire system of interaction, the same categories of scientific and statistical interest apply:

- Statistical Description. In a single instance of a network, there are of order n^2 quantities in terms of nodes and edges. Is there a set of sufficient statistics that parsimoniously summarizes the construction of the network in a parametric family?

- Generative Inference and Prediction. This applies to both the story of how the network comes to be, both in terms of the time evolution of nodes up until the time of observation (the single observation case) and how networks from the same family would evolve (the super-population case). What does "asymptotic behaviour" mean in the case of networks – a single network growing in size with the same generative properties, a single network whose properties scale in some way with increasing size, or a growing series of networks with identical properties?

- Nodal behaviour. One of the over-arching goals of modelling a network is to show how it impacts the individuals within it. Therefore, any network constructed must ultimately be connected back to nodal properties and outcomes in order to have meaning. Modelling uncertainty in network ties therefore has a directly measurable impact on the nodes themselves.

Additional background on these modelling questions can be found in Goldenberg et al. [2009] and Kolaczyk [2009].

We begin by reviewing the development of stochastic network modelling from the perspective of statistics and sociology, from the Erdős-Rényi-Gilbert model to the Exponential Random Graph model, and introduce other mechanisms that have been designed to account for unexplained connections between individuals such as latent spaces and membership models. We then introduce the piecewise development of the general approach we prescribe, in terms of each of the three levels of assembly: general, node-specific and edge-specific quantities. This is then followed by the application of the method to binary networks produced by the "small-world" method of Watts and Strogatz [1998] and the preferential attachment mechanism of Barabasi and Albert [1999] as well as the assortative mixing measure of Newman [2002].

2 Edge-Specific Binary Network Models

2.1 Erdős-Rényi-Gilbert Random Graphs

While graph theory has provided many insights into the construction of deterministic networks, much of the interest in the use of networks in a nondeterministic setting began with a series of papers from Paul Erdős and Alfred Rényi [Erdos and Renyi, 1959, 1960, 1961; Erdos, 1959, 1961]. These papers put forth the notion of the canonical "classical random graph". In this simple model there is a community of N individuals, and $\binom{N}{2}$ potential undirected two-person relations, k of which are known to exist. There is an ensemble of graphs of size $\binom{\binom{N}{2}}{k}$ with this particular property, and remarkably, the properties common to the members of the ensemble are known to depend on the fraction of edges per node, $\lambda = k/N$, namely that graphs with $\lambda > 1$ are completely connected in the asymptotic limit.[1]

Our starting point is the closely related paper of Gilbert [1959], which takes a binomial starting point: each tie has probability of existing, independent of every other possible tie, with some fixed probability p. This is immediately extensible to the case where there are $N(N-1)$ potential *directed* edges that exist independently and with common probability p, or that for the edge between individuals labelled i and j, the tie is defined as $Y_{ij} \sim Be(p)$.

Other schemes for generating connected structures have since arisen. In the past 10 years, many researchers in the computer science, physics, and machine learning communities have followed these examples in proposing models for evolutions of complex networks with simple underlying properties. Among others, explorations in unified structure across classes [Airoldi and Carley, 2005], hierarchical sub-grouping [Clauset et al., 2008], and self-similar "Kronecker power" models [Leskovec et al., 2005] have brought new insights into the growth of complex systems from simple roots.

2.2 Distinguishing Individuals

It is of great interest in many scientific applications to include individual-specific information when considering the formation of networks among them. In particular, there are insights provided by mixed effects modelling that show how to include both covariate information and latent effects at each level (such as individuals or their communities), neither of which is possible under the classes of models previously specified.

Binary relational ties are modelled according to the probability of their presence. An extremely

[1] There is a substantial literature on using ensemble methods for random graphs, where each member of the ensemble has the same statistical specification and each has an equal probability of observation. Blitzstein and Diaconis [2006], Newman et al. [2001] and Handcock and Morris [2006] all deal with the generation of random graphs with fixed degree distribution, such that the number of ties for each individual is identical. A comparison of these methods appears in Section 7.

popular method of connecting binary outcomes to factors that affect their probabilities is the use of a link function, whose main purpose is to transform a potentially unbounded quantity to a value between zero and one. The most popular link functions for this purpose are the Gaussian cumulative distribution function (the use of which gives "probit", or probability unit, modeling), and the logistic distribution, which gives effects in terms of log-odds ratios [Berkson, 1951].

2.3 Joint-Conditional Specifications: The p-class of models

Several models have been proposed where the explicit probability model for a network's edges is in the joint distribution, so that a specification for a tie is made conditionally on the rest of the network. A description of the opposite approach is given in Section 5; here a series of models are detailed from the original approach, with their dependences constructed and measured.

The p_1 model

A groundbreaking model was published by Holland and Leinhardt [1981] and named "p_1" in part due to its simplicity. The model suggests the presence of three types of relations: the propensity with which an individual will be outgoing, or "gregariousness"; the propensity with which an individual will attract others, known as "popularity" and/or "attractiveness"; and the degree to which a connection in one direction will be reciprocated in the other direction compared to what would normally be expected, or "reciprocity".

To demonstrate this model, consider a trivial network of two individuals. With two possible directed edges, there are four mutually exclusive outcomes that can be observed. As specified by Holland and Leinhardt [1981] (with a slight modification), each of their joint probabilities is specified in the 2-by-2 table:

	Y_{21}=0	Y_{21}=1
Y_{12}=0	$\mathrm{P}((Y_{12}, Y_{21}) = (0,0)) = n$	$\mathrm{P}((Y_{12}, Y_{21}) = (0,1)) = b$
Y_{12}=1	$\mathrm{P}((Y_{12}, Y_{21}) = (1,0)) = a$	$\mathrm{P}((Y_{12}, Y_{21}) = (1,1)) = m$

The probabilities represent a *mutual*, *asymmetric* (individual 1 sends, a, or receives, b), or *null* connection respectively. This suggests a quadrinomial specification for any dyad $(1, 2)$:

$$P(Y_{12}, Y_{21}|m, n, a, b) = m^{Y_{12}Y_{21}} a^{Y_{12}(1-Y_{21})} b^{(1-Y_{12})Y_{21}} n^{(1-Y_{12})(1-Y_{21})},$$

noting that the probabilities sum to unity: $n + a + b + m = 1$. When respecified in terms of logarithms of the probabilities, the expression becomes

$$P(Y_{12}, Y_{21}|m, n, a, b) = \exp\left(Y_{12}Y_{21}\log\frac{mn}{ab} + Y_{12}\log\frac{a}{n} + Y_{21}\log\frac{b}{n} + \log n\right).$$

This is an exponential family specification, yielding a set of natural parameters: $\theta_{12} = \log\frac{a}{n}$ and $\theta_{21} = \log\frac{b}{n}$ are expressions for the probability of an outbound or inbound connection relative to a null connection, and $\rho_{12} = \frac{mn}{ab} = \rho_{21}$ reflects the likelihood of two individuals holding identical views on their relationship, rather than disagreement. Replication of this particular dyad k times would produce the joint distribution

$$P(Y_{12,1}, Y_{21,1}, ..., Y_{12,k}, Y_{21,k}|m, n, a, b)$$

$$= \exp\left(\sum_{j=1}^{k}(Y_{12,j}Y_{21,j})\log\frac{mn}{ab} + \sum_{j=1}^{k}Y_{12,j}\log\frac{a}{n} + \sum_{j=1}^{k}Y_{21,j}\log\frac{b}{n} + k\log n\right).$$

In practice, however, only one replicate of each dyad is observed in a network setting. When expanding this full network of N people, each probability in the quadrinomial is labelled according to the dyad in question: the quantities (n, a, b, m) become $(n_{ij}, a_{ij}, a_{ji}, m_{ij})$, and the natural parameters change accordingly to $\theta_{ij} = \log\frac{a_{ij}}{n_{ij}}$ and $\rho_{ij} = \log\frac{m_{ij}n_{ij}}{a_{ij}a_{ji}}$. Because $3\binom{N}{2}$ terms are impossible to estimate with $2\binom{N}{2}$ data points, Holland and Leinhardt [1981] simplify this expression to reflect a global tendency for reciprocation, and individual-specific effects for gregariousness and attractiveness:

$$\rho_{ij} = \rho, \tag{1}$$

$$\theta_{ij} = \theta + \alpha_i + \beta_j. \tag{2}$$

Additionally, the baseline tendency for tie formation stipulates that the sender and receiver effects are set with respect to a reference point. In the p_1 case, each of the effects sums to zero:

$$\sum_i \alpha_i = 0; \tag{3}$$

$$\sum_j \beta_j = 0. \tag{4}$$

When this is expanded for all dyads simultaneously, the p_1 specification for a graph $Y = \{Y_{ij}, \forall i, j < N, i \neq j\}$ is

$$p_1(Y) = \exp\left(\theta(\sum_{i,j} Y_{ij}) + \sum_i \alpha_i(\sum_j Y_{ij}) + \sum_j \beta_j(\sum_i Y_{ij}) + \rho\sum_{i,j}(Y_{ij}Y_{ji})\right)\prod_{i<j} n_{ij},$$

which can be used to draw corresponding random graphs given the appropriate parameters, which in this case are taken to be fixed effects. Maximum likelihood estimation is the method recommended by Holland and Leinhardt [1981] to estimate the effects within and generate simulated random graphs for comparison.

Differential Reciprocity in $\mathbf{p}_1$

An immediate extension to the p_1 model is proposed by Fienberg and Wasserman [1981], by extending the specification of the reciprocity term to be

$$\rho_{ij} = \rho + \rho_i + \rho_j,$$

where each ρ_i represents the additional tendency for individual i to reciprocate a relationship over a baseline level ρ, and all terms sum to zero, $\sum_i \rho_i = 0$.

As this is an extension of the general model, a likelihood ratio test can be performed to see if the additional dispersion in reciprocity is necessary for model fit.

Partial pooling of sender-receiver terms in $\mathbf{p}_1$

Another extension of the model is into Bayesian territory. In the treatment of Wong [1987], the sender and receiver effects are partially pooled and jointly modelled, such that

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

In this specification, a relationship between the "gregariousness" and the "popularity" of each individual is modelled as a correlation, taken to be identical for each individual. Empirical Bayes methods are then used to estimate the common variance and correlation.

Addition of covariates: $\mathbf{p}_2$

One specification that follows from the original is the addition of the impact of covariates to the terms for the baseline, sender, receiver and reciprocity terms. As specified in van Duijn et al. [2004], the model takes the form

$$\alpha_i = X_{1i}\gamma_1 + A_i \quad (5)$$
$$\beta_i = X_{2i}\gamma_2 + B_i \quad (6)$$
$$\mu_{ij} = \mu + Z_{1ij}\delta_1 \quad (7)$$
$$\rho_{ij} = \rho + Z_{2ij}\delta_2 \quad (8)$$

where X_1, X_2, Z_1 and Z_2 are matrices of covariates to be considered; γ_1, γ_2, δ_1 and δ_2 are the corresponding vectors of coefficients; and A_i and B_i are respective intercept terms with common variances σ_A^2 and σ_B^2. μ and ρ are the common parameters for all arc means and correlations.

2.4 Geometric/Topologically Specified Models

A class of models that find their origins in the logistic specification are Exponential Random Graph models, or p-star (p*) models, which were primarily conceived for purposes in which individuals are not explicitly differentiated by their characteristics, only by the network level structure: sociological patterns are encoded as statistics within the likelihood function, and the ensemble of individuals and links is modeled jointly. The method has its origins in the Markov Graph models of Frank and Strauss [1986], in relation to Markov Random Fields and the connection of edges that share a common node; the method is detailed in [Wasserman and Pattison, 1996; Anderson et al., 1999].

In short, the likelihood of the graph is given in the form of the ensemble $\mathbf{Y}$, where Y_{ij} is a binary directed edge, and the statistical measures of interest to the investigator, such as counts for 3-cycles, $C = \sum_{i<j<k} Y_{ij}Y_{jk}Y_{ki}$, and transitive triples, $T = \sum_{i,j,k} Y_{ij}Y_{jk}Y_{ik}$, are placed directly into the likelihood

$$p(\mathbf{Y}|\alpha,\beta) = \frac{1}{\kappa(\beta)} \exp\left(\alpha + \beta_1 C(\mathbf{Y}) + \beta_2 T(\mathbf{Y})\right)$$

such that the constant $\kappa(\beta)$ normalizes the likelihood. Due to the allowance of dyadic dependence, this constant is notoriously difficult to compute exactly as network size increases. Even a maximal simultaneous dependence of three dyads on a modestly-sized 30 node network leads to $\binom{\binom{30}{2}}{3} = 13,624,345$ total terms to calculate.

While the specification is popular with sociologists, due mainly to the specification of structures with known explanations, it has several deficiencies for our purposes, mainly that its parameters are not directly interpretable for the behavior of the individuals. Additionally, while individual characteristics such as covariates can be added to the likelihood through summary statistics, it is not clear that these coefficients will be interpretable in combination with motif-based measures. As a result, there is significant motivation to pursue an approach through which one can measure

and recover the presence of these quantities of interest in a meaningful and interpretable way.

3 Latent Characteristic Modelling Approaches

Many of the models previously discussed have the capacity to include information on the nodes of a network in the likelihood of tie formation, particularly in terms of known covariates. These methods are expanded on in Chapter 4 in greater detail. This section contains background on a different approach: methods for inferring the presence of a latent geometry or latent nodal characteristics as a means of explaining connectivity.

One of the simplest modes of node identification is that of membership in a group. The principle of stochastic equivalence in relational data requires that the probability of a tie between members of two groups (or, two members in the same group) depends only on the label of the group(s); that is, all members of a group are essentially identical as far as tie formation is concerned. (For a more detailed explanation, see Hoff [2007b].)

If the cluster memberships are known, the system can be analyzed using the stochastic blockmodel method of Fienberg and Wasserman [1981]. With K subgroups, there are $(K + \binom{K}{2})$ intra- and inter-component factors to estimate; dyadic independence then extends upward to component independence allowing for faster analysis. In the examples provided by Fienberg and Wasserman [1981], each component factor is estimated under the p_1 framework, though the methodology is not restricted to this interpretation. This is extended to the full directed graph case in Wang and Wong [1987]

This method is almost immediately extensible to models with multiple group assignments. However, since these multiple groups may overlap with each other, the estimation procedure for component factors cannot be performed in parallel.

On the other side are algorithms that infer group membership based on relational ties. Wasserman and Anderson [1987]; Snijders and Nowicki [1997]; Nowicki and Snijders [2001] describe methods and algorithms for estimating a flexible number of clusters, and assignment to each cluster with a particular probability.

The trick in this case is that clusters are measured on people, whereas observations are made on ties. There is a clear advantage to this approach, in that clusters represent sociological phenomena. In particular, it is apparent that a cluster represents a community, or other such group, that has a tangible meaning or benefit to its members, leading to interpretability of the results both for connections and for individual outcomes.

There is an extension to the "mixed-membership" case, in which each individual can simultaneously belong to multiple blocks to various degrees. Airoldi et al. [2007, 2008] give an overview as well as algorithms to model mixed membership with Bayesian methods, in particular the use of a variational Expectation-Maximization (v-EM) algorithm to approximate the posterior distri-

bution [Jordan et al., 1999]. To summarize the generative process of the model, consider a mixed membership vector $\vec{\pi}_p$ for each individual p, which describes the strength of association with each block. (Any individual's respective strengths sum to 1.)

A tie existence is generated by sampling a single group membership for individuals p and q. Given these groups (G_p, G_q), the probability of a tie is given by the block matrix B_{G_p,G_q}. Note that this method also allows for links within clusters to be *less* likely than those between clusters, a property inherited from the observable cluster model of Fienberg and Wasserman [1981].

3.1 Latent Trait Models

The notion of latent cluster membership can be extended to the notion of latent traits. In this conception [Hoff, 2005; Nickel, 2006] the underlying traits are not confined to sum to one; the tendency of individuals to interact is instead perceived to be the inner product of latent trait vectors with respect to an underlying weight matrix.

The generative model for tie formation gives the probability of a tie as

$$logit(p_{ij}) = \mu + z_i^t C z_j,$$

where C is the weight matrix, typically diagonal, and (z_i, z_j) are the latent trait vectors. Note that for identifiability, either the elements of C must be fixed in magnitude or the latent trait vectors are somehow constrained.

The value of this method is not simply in accounting for unexplained tie formation, but also as a hypothesis-generating mechanism. For those traits where $C_{ii} < 0$, there is a latent trait heterophily; likewise, a latent-trait homophily is observed when $C_{ii} > 0$. Given that the model can pick up one of these phenomena, it may represent an unobserved covariate on the nodes worth discovering, rather than simply a useful mathematical curiosity.

3.2 Latent Space Models

An alternative explanation for social connectivity lies in considering a latent geometry within which all nodes take a position. The stochastic model postulated by Hoff et al. [2002] places nodes within a latent Cartesian space. The tendency for any two individuals to be connected by an edge is then driven by the distance between the individuals in the space. For example, let d_i be the position of individual i in the latent space. The probability of connection is then governed by

$$logit(p_{ij}) = \mu - \gamma ||d_i - d_j||_2$$

so that an increased distance implies a decreasing likelihood of connection. It is worth noting that

as expressed, the connection probability will always decrease from the maximum value governed by μ. This approach was integrated with the blockmodel approach in Handcock et al. [2007], and with sender and receiver effects in Krivitsky et al. [2009].

Another approach was considered in Linkletter [2007], so that rather than using only unexplained variance to formulate and space, a functional nonparametric method that begins with known covariates uses these to produce a latent space model. This is one step along the way to integrating the entire approach into a functional data analysis method.

4 Framework: Modelling Relational Data With Marginally Specified Hierarchical Models

Most of the approaches just listed are based on a notion of conditional dyadic independence, or the construction that 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. While this is by no means a certainty in many real-world observable sitations – for example, a person can only have two biological parents (and indeed, *must* have them) – this in no way limits the practical benefits of this modelling approach to other data sets. 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 [Hoff, 2007b; Hoff and Ward, 2003; Hoff, 2005; Krivitsky et al., 2009] – 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. Next, three of the properties explored in the last section are brought into the

current framework: latent spaces, latent characteristics and the behaviour of assortative mixing on degree. Following this, the extension of the standard Gaussian to a multivariate Student-t is discussed, converting the "probit"-type analysis to a ''robit" [Liu, 2004]. Finally, we conclude with a discussion on the comparisons between log-linear models and their marginal equivalents, including several arguments why the marginal model should be used in the base case for network analysis.

4.1 Marginal Specification and Extension of $\mathbf{p}_1$

The original p_1 model 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.

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

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

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:

$$\begin{bmatrix} Y_{ij} \\ Y_{ji} \end{bmatrix} \mid \begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix} = \begin{bmatrix} \mathbb{I}(Z_{ij} > 0) \\ \mathbb{I}(Z_{ji} > 0) \end{bmatrix}; \tag{9}$$

Parameter	Draw Type	Distribution	Method
$\begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix}$	Parallel	$\text{TN}_2(\begin{bmatrix} \mu+\alpha_i+\beta_j \\ \mu+\alpha_j+\beta_i \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \begin{bmatrix} Y_{ij} \\ Y_{ji} \end{bmatrix})$	Direct Draw
α_i	Sequential	Normal	Direct Draw
β_j	Sequential	Normal	Direct Draw
μ	Sequential	Normal	Direct Draw
σ_α	—	Inv-Gamma	Direct Draw
σ_β	—	Inv-Gamma	Direct Draw
ρ	—	$p(\mathbf{Z}\|\rho)p(\rho)$	Grid approx.

Table 1: An RCMS table summary for computing the GLM version of p_1. Further definitions are in Appendix A.

$$\begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix} |\alpha, \beta, \rho \;\sim\; N_2\left(\begin{bmatrix} \mu+\alpha_i+\beta_j \\ \mu+\alpha_j+\beta_i \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \tag{10}$$

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 Wong [1987],

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta \\ \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{bmatrix}\right) \tag{11}$$

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

A Gibbs sampling scheme, as inspired by Albert and Chib [1993], is relatively easy to put together. Following the method derived in Appendix B, there is a Gibbs sampling algorithm as given in Table 1. 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:

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

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 11, make the conditional draw for the parameter as natural as from a standard distribution.

It is also notable that the addition of other terms to the formula, as specified in Table 2 and

described later in this section, do not affect the form of these draws when conditioned on; the quadratic form is preserved.

4.2 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.

4.3 Differential Reciprocity Adjustments

Fienberg and Wasserman [1981] 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,

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

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.

4.4 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

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

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 $d_2 = (1, 0, ..., 0)$. 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.

A reliable RCMS profile can be built by fixing the coordinate of one point, say $z_1' = (1, 1, ..., 1)$, and allowing all other points and the mixing matrix C to vary relatively, exploring these via Metropolis steps. As in the latent position model, an optimization by simulated annealing may prove to be the most efficient way of determining a reliable starting point.

4.5 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.[2]

The approach of Newman [2002] 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

[2]In 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.

of assortative mixing, the parameter χ is introduced and an additional term is included, directly proportional to the popularity and gregariousness of the individuals:

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

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.

This form is also easily computable given the conditional maximization and sampling framework. In the Gibbs sampling formulation, conditional on χ and β_j, the sampling for α_i remains straightforward, as the full conditional posterior distribution is still a quadratic form. This remains true for the sampling of any one parameter, conditional on the other two.

4.6 Robust Analyses with the Multivariate t Distribution

Stability is often a concern in binary modelling with the probit framework, due to the light tails of the underlying normal distribution. A mechanism for allowing heavier tails, hence gaining resilience against outliers and robustness to the assumption of latent normality, is provided in "robit" regression [Liu, 2004]. The addition of extra variability on the latent normal is performed with a data augmentation step and is complementary to the Gibbs sampling and ECM approaches used to this point.

The original latent variable formation is $Z_{ij}|\mu_{ij} \sim N(\mu_{ij}, 1)$. To convert this to a t-distribution, the constant variance is replaced by a random variable. In particular, consider a Gamma variate $G_{ij} \sim 2/\nu Gam(\nu/2)$, so that $EG_{ij} = 1$ and $E\frac{1}{G} = \frac{\nu}{\nu-2}$. A representation for the standard t distribution is

$$T_{ij}|G_{ij} \sim N\left(0, \frac{1}{G_{ij}}\right),$$

yielding the marginal distribution

$$T_{ij} \sim t_\nu.$$

As the variance for this distribution is $\frac{\nu}{\nu-2}$, the underlying variate can be restored to unit variance with an added scale factor, and given the original mean shift μ_{ij}, the latent distribution takes the form

$$T_{ij}|\mu ij, G_{ij} \sim N\left(\mu_{ij}, \frac{\nu-2}{\nu G_{ij}}\right).$$

Given the augmented variance, the probability of a positive Bernoulli outcome is now

$$P(T_{ij} > 0|G_{ij}, \mu_{ij}) = \Phi\left(\frac{(\nu-2)\mu_{ij}}{\nu G_{ij}}\right);$$

conditional on the augmented variance term G_{ij}, the existing Gibbs sampling structure may still be used for the augmented data $T_{ij}|G_{ij}$ and the terms included in μ_{ij}. Given a fixed degrees of freedom ν, the Gibbs sampler for each G_{ij} draws from the distribution for the gamma:

$$\log(p(Z_{ij}|\mu_{ij}, G_{ij})p(G_{ij})) = c + \frac{1}{2}\log G_{ij} - G_{ij}\frac{(Z_{ij}-\mu_{ij})^2\nu}{2(\nu-2)} + (\nu/2-1)\log G_{ij} - \nu G_{ij}/2;$$

$$G_{ij}|Z_{ij}, \mu_{ij} \sim \frac{2}{\nu(1+\frac{(Z_{ij}-\mu_{ij})^2}{\nu-2})} Gam\left(\frac{\nu+1}{2}\right).$$

One additional benefit of the robit model is the ability to tune the degrees of freedom. In particular, Liu [2004] suggests that setting $\nu = 7$ gives the robit model contours that approximate the logistic distribution quite well, with the benefit of heavier tails. This means that for the addition of computing power, a Gibbs sampling model can be built that will approximate the commonly used logistic model very well, while still maintaining the benefits of heavy-tailed distributions.

4.7 Integration with Other Generalized Linear Model Forms, and Further Extensions

All the recipes listed in this section have so far been defined on the probit model for two main reasons. First, as these methods have been defined and developed primarily for the analysis of binary data, it is essential that any extensions that are subsequently developed can be applied to that domain. Second, the computational tools developed for other analyses of binary data are not themselves limited to this class of outcomes.

The model-building strategies laid out to this point for probit-type models are equally applicable to other classes of data, with differences only in the parametrizations of each of these model families.

4.7.1 Normal-family Data

Because the binary data strategy has been derived from latent normal-family distributions, the extension to this family of data is immediate. Consider the general form for a bivariate normal

distribution,

$$\begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix} | \begin{bmatrix} \mu_{ij}, \mu_{ji}, \sigma_{ij}, \\ \sigma_{ji}, \rho_{ij} \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \mu_{ij} \\ \mu_{ji} \end{bmatrix}, \begin{bmatrix} \sigma_{ij} & 0 \\ 0 & \sigma_{ji} \end{bmatrix} \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{ij} & 0 \\ 0 & \sigma_{ji} \end{bmatrix} \right),$$

which has three groups of terms that can be expanded: mean values, variances and correlation. The mean term can be decomposed just as in the previous examples, with a grand mean, sender and receiver effects, assortative mixing and latent parameters. Correlation between arcs in a dyad can be handled as in Section 4.3.

The inclusion of standard deviation terms in the parametrization, particularly the notion of differential variance, allows for additional flexibility in modelling. Consider a decomposition of the form

$$\log \sigma_{ij} = \log \mu_\sigma + \log \sigma_i + \log \sigma_j,$$

so that the variance of each edge depends on both the mean degree of variance in the system and on characteristics of each node in the dyad.

4.7.2 Partial Correlations as Network Ties

Section 4.3 refers to the modelling strategies for reciprocity between ties as a correlation function. While these are instances of modelling latent quantities, the same methodology can be applied to correlations that are estimated between units.

If the correlation between units represents the total impact of one unit's fluctuations on the other, than partial correlations between units represent direct effects, represented as network ties. The Fisher transformation

$$q_{ij} = \log \frac{1 + \rho'_{ij}}{1 - \rho'_{ij}}$$

is used to bring the data to the normal scale; subsequently, this can be modelled as if each transformed correlation is a normal random variable.

4.7.3 Count Data

If data are integer-valued counts, $Y_{ij} \sim Poisson(\lambda_{ij})$ is a natural interpretation for tie strengths, which can then be modelled as

$$\log \lambda_{ij} = \mu + \alpha_i + \beta_j$$

for sender and receiver effects respectively. The Negative Binomial distribution can be modelled as in Zheng et al. [2006] by adding an additional overdispersion term,

$$\log \lambda_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where $\gamma_{ij} \sim \omega Gamma(1/\omega)$; if necessary, this itself can be expanded so that the overdispersion is hierarchically modelled and necessarily positive,

$$\log \omega_{ij} = \log \omega + \log \omega_i + \log \omega_j.$$

4.7.4 Finite Ordinal Data

In applications of the measurement of network ties by surveys, many measurements are taken on an ordinal scale. The work of Rasch [1960] suggests that this may be accomplished through a latent variable estimation method. While the original method was proposed in terms of the logistic distribution, the definition is equally palatable in terms of the normal distribution.

Consider k ordinal levels of a particular response to a survey question; the typical example is a five-point scale in the style of Likert [1932], where a statement can be treated as {strongly disagree, disagree, neither, agree, strongly agree}, rescaled to $\{1, 2, 3, 4, 5\}$. While these values can be taken directly as numerical scores, it is also possible to consider these to be the manifestation of a latent normal random variable with break points, in this case $\{\beta_1, ...\beta_4\}$, such that

$$P(X = i) = \Phi(\beta_i) - \Phi(\beta_{i-1}),$$

where $\beta_0 = -\infty$ and $\beta_k = \infty$ for the sake of completeness. This can be directly expressed in terms of the latent normal,

$$X = \sum_i i\mathbb{I}(\beta_{i-1} < Z < \beta_i)$$

so that inference is then taken on the breakpoints. Given that the latent variable is normal, and that survey questions to one individual regarding another is a description of a directed arc, it then remains to model each arc as the manifestation of the latent variable, $Z_{ij}|\mu_{ij} \sim N(\mu_{ij}, 1)$ as in the binary and continuous normal cases.

4.8 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

$\mathbf{Y}\|\mathbf{T}$	$= f(T^1, T^2, ...)$: 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)	$z_i' C z_j$
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 2: 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 $f(T^1, T^2, ...)$ may be deterministic or stochastic.

terms. These terms are summarized in Table 2.

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 [Klaassen and Wellner, 1997; Pitt et al., 2006; Shaw and Lee, 2007; Hoff, 2007a].

5 Discussion: Comparing Marginal Specifications to Conditional Models

Section 2 gives a number of examples where the arcs in a dyad were treated jointly in the modelling process. This section demonstrates some of the options available when arcs were analyzed marginally first, with the interaction of the two arcs in a dyad considered as a secondary concern. The choice to refer instead to marginal models for ties has a number of motivations, which are discussed in the upcoming subsections. Two main areas are discussed: first, the fundamental unit of analysis in each case, being the arc or the dyad; second, the expandability of marginal models versus their joint counterparts.

5.1 The Fundamental Unit: The Arc or the Dyad

If there is dyadic independence and the object in question is undirected, there is no issue of model choice, as the two are functionally equivalent: the arc is the dyad, and vice versa, and the parametrization is essentially a matter for the investigator, whether a choice between logit and probit analysis, or a parametric versus nonparametric model, or any other choice that may come up in model selection. This also applies if the two arcs in the model are conditionally independent; if there is no need to account for the other object, it is as if there are two undirected dyads across the same pair of nodes.

The issue becomes the case of dependence, and the unit of interest in each case. While the dyad is the equivalent of the experimental unit in this case (since it is perceived to be conditionally independent of all others), the arc is the unit of observation, as it focuses on one individual first. These data sets are typically collected by survey on the individual, or the observation of communications to or from an individual, so that while the reciprocal behaviour is potentially full of information, a model can easily be composed without the need for corresponding mutuality of observation.

5.2 Expandability: Beyond the Binary Case

It has been demonstrated that marginal and joint methods can both model conditionally independent binary data. When expanding to edge types beyond the binary, there is more room for debate about which method is preferable.

Consider first the case of ordinal data in the uncountably infinite case, where each arc is valued along the real line or a subset thereof. When the choice is between modelling the probability of a dyad value in an infinitesimal area, or of first modelling each arc probability within an infinitesimal length, separating this problem from that of the correlation of the two arc values. While a functional data approach might be able to model the probabilities in terms of a joint functional distribution, such methods are less suitable for parametric models, or even Generalized Additive Models, whose multidimensional equivalents have typically been defined to be interpreted in terms of the marginal distributions, such as the multivariate normal.

The issue becomes slightly murkier when arcs are measured as classes of finite, categorical data. While the quadrinomial model can be extended to a general k^2-nomial framework for k categories in each arc, there is a disadvantage when only one correlation parameter is presented between the two outcomes, yielding models with only $2k + 1$ degrees of freedom as opposed to $k^2 - 1$ in the multinomial framework.

However, the issue is strictly one of interpretability, as there is no strict requirement that only one interaction term is necessary. Imai and van Dyk [2005] propose a latent variable formulation for the multinomial probit, using a latent multivariate normal whose result is determined by

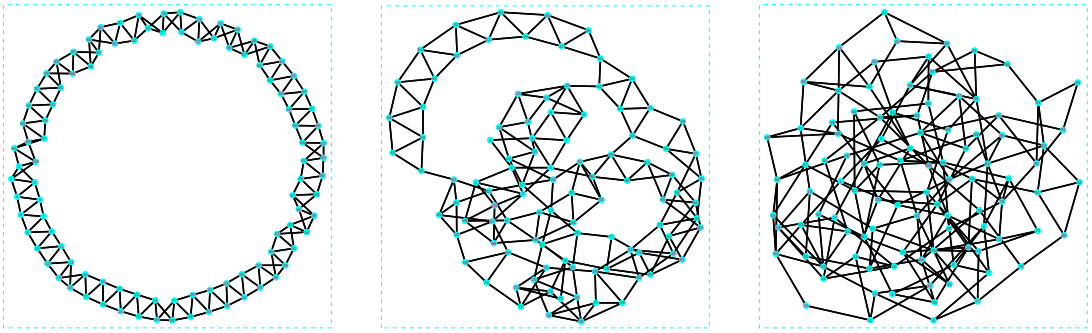

Figure 1: A series of networks formed by the Watts-Strogatz small-world algorithm. Left, an "ordered" lattice, in which every node is connected to its two nearest neighbours. Middle, a small number of rewirings is permitted, maintaining the close connection of neighbours but decreasing the geodesic path lengths. Right, continued rewiring of connections at random.

the component with the maximum value. The expansion of this multivariate normal with the appropriate correlations included would allow for the marginal model to be implemented in this context as well; these modifications will be implemented in future research.

6 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 those models from a different perspective. In the process, there is no commitment to a single generative story for these systems, only to demonstrate that there are multiple ways of constructing the same networks that give rise to multiple plausible explanations for their generation; in particular, that there are other simple explanations for small-world and preferential-attachment graphs. The remaining section deals with a class of data that can be put together with similar simple explanations: the modelling of correlational data, often used to propose binary network structures.

6.1 Watts-Strogatz "Small-World" Networks

While studying the mechanisms of coupled harmonic oscillations in biological networks, Watts and Strogatz [1998] 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 [Milgram, 1967], 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. The original algorithm took the following form.

1. Create n nodes around a two-dimensional ring. Connect edges between nodes if they are within a certain distance along the ring; in particular, the original paper proposed that a node was connected to each of its two nearest neighbours on each side, for a degree of four for each node.

2. Select a rewiring proportion p, and choose $4np$ edge endpoints uniformly.

3. For each selected endpoint (with node identity i), choose a node uniformly at random from all nodes in the lattice (k) except for the corresponding endpoint of the current edge (j); rewire the edge such that $Y_{ij} = 0$ and $Y_{kj} = 1$.

An example of rewiring at two scales is given in Figure 1 for an example. This model has proven extremely inspiring to researchers across fields; a thorough review of the literature is available [Dorogovtsev and Mendes, 2002]. This model is easy to code, and to visually process. It also has the advantage of being extremely fast in the large n case (2.25×10^5, in the movie-actors example). However, the generation process of the model assumes a stark bifurcation between two classes of relationships: close neighbours and everyone else. Additionally, while the story of "random rewiring" of a fixed number of connections is easy to explain the topological properties of a system, it may prove to be an implausible model for the way in which the connections in a networked system may develop.

The original small-world generative model takes three parameters: n, the number of nodes in the system: k, the number of neighbours on each side to which a node is initially tied, and p, the probability that any end of a tie is rewired at random. The GLM method for constructing a small-world-type graph takes a similar input, with a slight redefinition of terms. While n is still the number of nodes in the system, let d be the total edge density (equal to $2k/(n-1)$ in the original case) to allow for a wider range of densities. p becomes a measure of the influence of longer-distance nodes.

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.[3] Let Y_{ij} be drawn from a Bernoulli $\{0, 1\}$ random variable with probability of success as the sum of two pieces.

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

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

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

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

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

As will be shown in the next section, 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.

To demonstrate the characteristics of the GLM models in comparison to the original small-world generation, each model is simulated for a series of cases with 200 nodes, in which the expected number of ties for each node is 6. The cases take various degrees of perturbation to demonstrate the presence of "small-world characteristics" in both the Watts-Strogatz and GLM generative models, namely

- The mean path length for all pairs of points.
- The Watts-Strogatz clustering statistic, defined to be the mean of the fraction of observed triangles from all those possible for each node, $K_k = \frac{\sum_i \sum_j Y_{ki} Y_{kj} Y_{ij}}{\sum_i \sum_j Y_{ki} Y_{kj}}$, averaged over all nodes k.

These characteristics are on displayed at various levels of p in Figure 2. 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 Watts and Strogatz [1998].

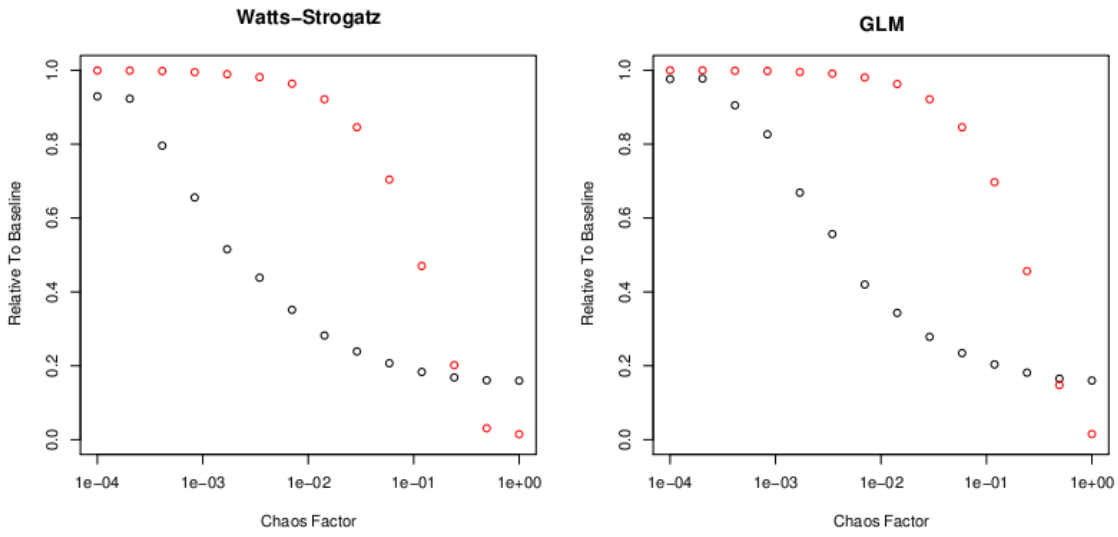

Figure 2: The median path length and clustering statistics for small-world graphs generated by the Watts-Strogatz and GLM methods. Each display the region between the curves where the graphs have small-world properties

6.2 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 Barabasi and Albert [1999], 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.

The symmetry of an Erdős-Rényi random graph may suggest simultaneous formation, due to the relatively even distribution of ties. In this class of network, the individuals with the longest memberships are more likely to have a higher number of ties due, effectively, to their longevity, amplified by an individual's tendency to connect with these more popular individuals. This is a process studied in Simon [1955] and de Solla Price [1976], and exemplifies the "rich get richer" phenomenon as a natural consequence. This is a mechanism that has been shown to explain the growth of the World Wide Web in its early years.

Notably, a preferential attachment model does not distinguish directly between its member individuals by label; when a complete graph is presented, the order of attachment is typically inferred from the degree distribution of the node. Thus, the model considers the notion of degree to be the identifying factor in the age of its presence in the network.

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:

$$\begin{aligned} Z_{ij} &\sim N(\beta_j, 1); \\ Y_{ij} &= \mathbb{I}(Z_{ij} > 0). \end{aligned}$$

A symmetrized version of this graph is produced to get the undirected equivalent, so that

$$Y_{ij} = \max\left(\mathbb{I}(Z_{ij} > 0), \mathbb{I}(Z_{ji} > 0)\right).$$

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.

6.2.1 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

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

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

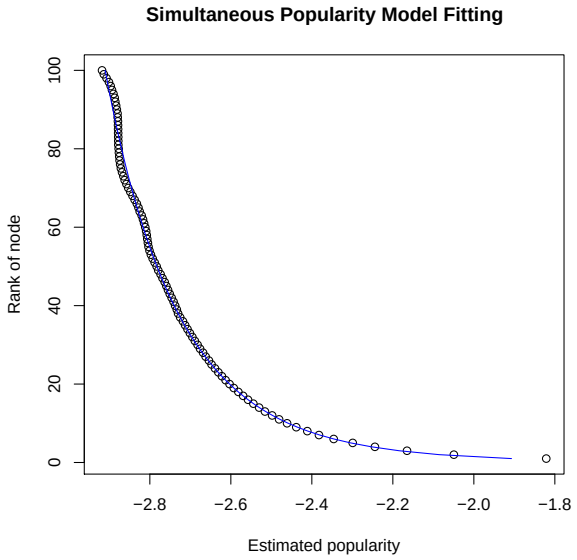

Figure 3: Fitting the expected degree of each node under a GLM model to the observed degrees of networks constructed by preferential attachment.

Quantity	Value
Visible Nodes	≈ 1000
"True" Nodes	≈ 5000
μ	-5.41
σ	0.974

Table 3: Results for a least-squares fit for a normal distribution fit to the estimated popularity measures. Note the much larger estimate of the number of true nodes present in the system that are not part of the visible giant component of the system.

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 3), 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. Rough estimates for a corresponding fit are given in table 3.

The most notable feature of the system is its high size, strongly indicating the presence of a great number of nodes that are not apparent in the main system, possibly in smaller components only internally connected.

A simulation from the GLM model is plotted in Figure 4 to show the similarities between it

and the original model. The largest component in this simulation has 906 nodes; there are an additional 1221 nodes that are connected in some way to others, leaving 3779 isolated nodes.

Finally, note that this system is not meant to be an exact duplication of the preferential attachment model. In particular, 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.

A Rotational Conditional Maximization and Sampling Methods

Throughout this work, the analyses proposed are executed by computer, given both the magnitudes of the data sets and the complexity of the models. As the models are composed, they are build with the ability to separate into conditional components. Additionally, the introduction of a latent structure with data augmentation allows for components that are easier to handle for sampling and maximization. For this reason, while many other algorithmic approaches are feasible to solve for the state of the system, this work advocates methods in which the likelihood or posterior distribution is maximized or sampled in terms of one parameter (or one parameter group) at a time. In the likelihood approach, this is done with the ECM algorithm [Meng and Rubin, 1993], which is already attractive due to the nature of the augmented data $\{Z_{ij}\}$.

In the Bayesian framework, the most common conditional sampler is the Gibbs sampler [Geman and Geman, 1984], one of the more commonly used Markov Chain Monte Carlo (MCMC) methods, in which each parameter (or sub-block of parameters) is sampled in sequence from the current full conditional distribution; the stationary state of the chain is then taken to be the posterior distribution in the inference. Approximations of the posterior distribution are also calculable though variational EM (vEM) methods (see Jordan et al. [1999] for examples), which can also be conditionally maximized by rotating through the model parameters.

Because both of these methods can be expressed in table form – the sequence of parameters from which to calculate expectation functions, find maxima or take draws – all distributions will be displayed in this form, a Rotational Conditional Maximization and/or Sampling (RMCS) table, for whatever method of inference is chosen to solve the system; a simple RCMS method is given in Table 4 for a bivariate normal distribution with known correlation. As a personal preference of the author, systems will be solved first through the maximization of the posterior distribution in the style of ECM (though here in the fully Bayesian context, all variables are considered random

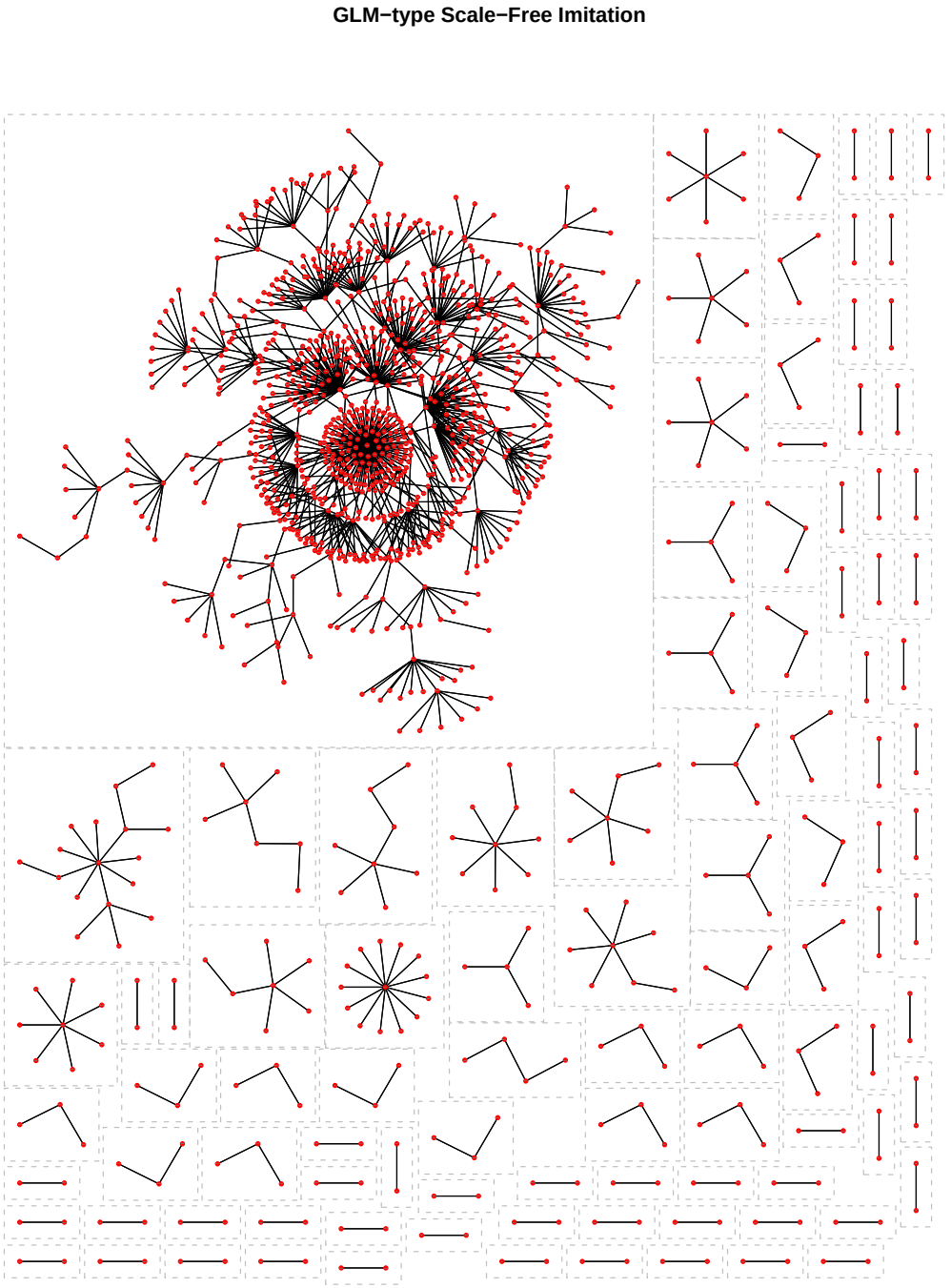

Figure 4: A representative simulation of the GLM model for scale-free-type networks. 1221 nodes with connections are shown, 906 of which are in the largest component.

$$\begin{bmatrix} Z_{1i} \\ Z_{2i} \end{bmatrix} | \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right); \qquad \mu_1, \mu_2 \sim N(0, 1000^2)$$

Parameter	Draw Type	Distribution	Method
μ_1	Single	$\prod_i p(Z_{i1}, Z_{i2}\|\mu_1, \mu_2)p(\mu_1)$	Direct Draw (normal)
μ_2	Single	$\prod_i p(Z_{i1}, Z_{i2}\|\mu_1, \mu_2)p(\mu_2)$	Direct Draw (normal)

Table 4: A Rotational Conditional Maximation and/or Sampling table for a bivariate normal distribution with known covariance.

variables, not missing "data" per se), and probability intervals calculated through generalized Gibbs sampling methods about any modes discovered in the previous step.

B Marginal Specifications of Dyads under the Gaussian Framework

Older network methods have been based on log-linear modeling and logistic regressions, which have enjoyed a long history largely due to their ease of computation. As the logistic CDF has an easy closed form, probabilities can be calculated from these models quickly, and by hand, given values for the parameters. However, due to the recent rise in computational power, difficulties formerly associated with the Gaussian CDF are no longer an impediment to its adoption for these tasks. In particular, its natural multivariate specification makes it ideal for modeling dependence between ties and individuals. See Cox [1972] for a road map of multivariate binary analysis that is still valid today.

Because the Gaussian CDF represents the probability of an underlying normal random variable being above or below a particular point, I consider all of our modeling of binary ties to reflect such a latent scheme. The approach used is a class of data augmentation algorithm first proposed to solve computational problems by Tanner and Wong [1987], while the particular algorithm uses has its roots in the work of Albert and Chib [1993]. Others who have extended this approach for other classes of data are McCulloch and Rossi [1994]; Nobile [1998]; Imai and van Dyk [2005]. in this case, I extend the approach of Albert and Chib to consider multivariate binary models using latent multivariate Gaussians.

The use of latent continuous variables for modelling discrete outcomes has a history over a century old, including cases when the outcomes can be dependent. Pearson [1900] first described a procedure for interpreting a binary outcome as the dichotomization of a latent normal random variable, which are then extended to a pair of outcomes on a two-by-two contingency table with a latent bivariate normal.

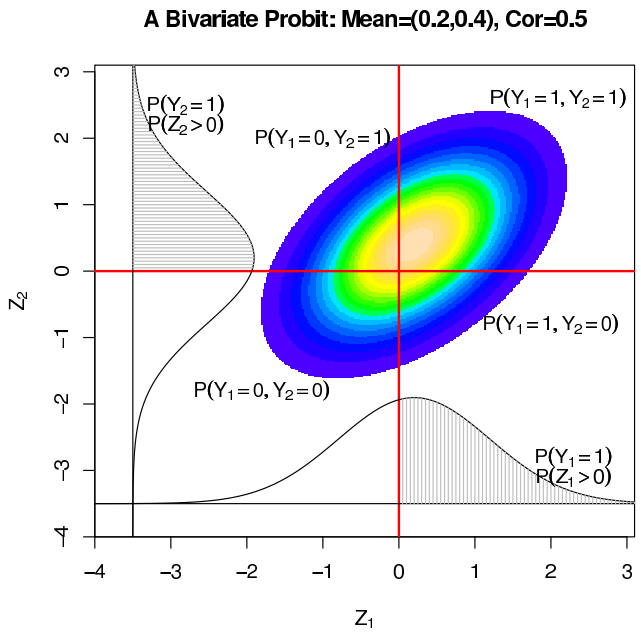

Figure 5: A latent bivariate normal random variable forms the basis for a bivariate probit model.

Consider a contingency table with the following cell and marginal probabilities for the joint binary variables (Y_1, Y_2):

Joint Probabilities	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	n	a	$1 - \pi_1$
$Y_1 = 1$	b	m	π_1
Total	$1 - \pi_2$	π_2	1

The layout of the table suggests a measure of dependence between the two outcome variables: if the joint probabilities in the table can all be expressed as the product of their marginal distributions, the rows and columns are independent. The mechanism suggested by Pearson [1900] is to parameterize dependence in terms of the latent bivariate normal.

The estimate of the correlation ρ between the latent normal variable pairs is what later became known as "tetrachoric correlation" (literally, the "four voices" in a two-by-two table). Due in no small part to the difficulty in calculating functions of the normal cumulative distribution function, this approach of analyzing contingency tables became less popular than the kappa technique for quantifying rater agreement, a concept with similar mathematical properties but unconcerned with latent traits, and odds-ratio estimation for off-diagonal probabilities, which have characteristics that have proven useful in the biometric literature.

Development of probit-based techniques later became widespread in biological studies by Bliss [1935], though the introduction of [Ashford and Sowden, 1970] reintroduced the concept of the

bivariate probit to the biometric community.[4]

In general, the bivariate specification is extendable to the multivariate case. Rather than estimating cell probabilities (as there is no easy expression for multivariate normal CDFs) I use the Pearson latent variable specification. Let $(Y_1, ..., Y_n)$ be a vector of binary responses. In general, the latent multivariate normal $\mathbf{Z}$ is defined as

$$\begin{aligned} \mathbf{Z} &\sim N_n(\mu, \mathbf{\Sigma}); \\ Var(Z_i) &= 1; \\ Cov(Z_i, Z_j) &= \rho_{ij}, i \neq j; \\ Y_i | Z_i &= \mathbb{I}(Z_i > 0). \end{aligned}$$

Marginally, each outcome Y_i is a Bernoulli outcome with success probability $\Phi(\mu)$. If there is no correlation between observations, the model reverts to a traditional probit analysis.

C Estimation Methods and Algorithms

The latent variable specification makes Bayesian inference on the resulting model significantly easier than in log-linear models. I use a modification of the algorithm proposed by Albert and Chib [1993] that capitalizes on the Gaussian framework, while adding sampling steps for the additional correlation terms.

Consider first a simple model where each observed outcome $i \in \{1, ..., n\}$ is bivariate binary, and all outcomes are conditionally independent:

$$\begin{aligned} \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} | \begin{bmatrix} Z_{i1} \\ Z_{i2} \end{bmatrix} &= \begin{bmatrix} \mathbb{I}(Z_{i1} > 0) \\ \mathbb{I}(Z_{i2} > 0) \end{bmatrix}; \\ \begin{bmatrix} Z_{11} \\ Z_{i2} \end{bmatrix} | \beta, \nu, \rho &\sim N_2 \left(\begin{bmatrix} \mathbf{X}_i \beta \\ \mathbf{X}_i \nu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \end{aligned}$$

where the last line is a bivariate general linear model with independent rows, and prior distributions are specified appropriately. All together, the latent variables are specified as a $(2n)-$variate normal:

[4] The authors make a key point that not all joint distributions whose margins are normal are themselves multivariate normal, such as the case $(Y \sim N(0,1), X = SY, S = 2Be(0.5) - 1)$ in which X and Y are uncorrelated but dependent. It is assumed throughout this work that the underlying model is the canonical multivariate normal distribution.

Parameter		Distribution	Method
$\begin{bmatrix} Z_{ij} \\ Z_{ji} \end{bmatrix}$	Parallel	Truncated N_2	Direct Draw
$\begin{bmatrix} \beta \\ \nu \end{bmatrix}$	Joint	N_{2p}	Direct Draw
ρ	—	$p(\mathbf{Z}\|\rho)p(\rho)$	Grid approximation

Table 5: An RMCS table for the bivariate normal with coefficients β and ν; full expressions for each term are given in the text. (See Section A for more on RCMS notation.)

$$\begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} |\beta, \nu, \rho \sim N_{2n} \left((I_2 \otimes X) \begin{bmatrix} \beta \\ \nu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes I_n \right)$$

In this specification, a Gibbs sampler in the form of that suggested by Albert and Chib [1993] can be executed to obtain draws from the posterior distribution (Table 5 gives the steps in order.) To draw the latent normal conditional on the coefficients β, μ, correlation ρ and outcomes $(\mathbf{Y_1}, \mathbf{Y_2})$, the truncated bivariate normal is used, where all n units can be drawn simultaneously:

$$\begin{bmatrix} \mathbf{Z}_{i1} \\ \mathbf{Z}_{i2} \end{bmatrix} |\beta, \nu, \rho, \mathbf{Y} \sim TN_2 \left(\begin{bmatrix} \mathbf{X}_i\beta \\ \mathbf{X}_i\nu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} \right).$$

For the coefficients (β, ν), the properties of the multivariate normal make the mean and variance estimates easier to calculate,

$$\begin{bmatrix} \beta \\ \nu \end{bmatrix} |\mathbf{Z}, \rho \sim N_{2p} \left((I_2 \otimes (X'X)^{-1}X') \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes (X'X)^{-1} \right).$$

This form is easily extendable to include a non-flat multivariate prior distribution for (β, ν).

Without a convenient conjugate form, the draw for ρ is best done by either a Metropolis-Hastings step or an approximation such as a grid-method draw. Since the value is bounded to the interval $(-1, 1)$, my personal preference is for a grid sample or a direct acceptance-rejection scheme, which would allow for more mixing than a Metropolis-Hastings method.

References

Airoldi, E. and Carley, K. (2005). Sampling algorithms for pure network topologies. *ACM SIGKDD Explorations Newsletter*, **7** 13–22.

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2007). Combining stochastic

block models and mixed membership for statistical network analysis. *Lecture Notes in Computer Science*, **4503**.

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, **9** 1981–2014.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88** 669–679.

Anderson, C. J., Wasserman, S. and Crouch, B. (1999). A p* primer: logit models for social networks. *Social Networks*, **21** 37–66.

Ashford, J. and Sowden, R. (1970). Multivariate probit analysis. *Biometrics*, **26** 535–546.

Barabasi, A. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286** 509–512.

Berkson, J. (1951). Why i prefer logits to probits. *Biometrics*, **7** 327.

Bliss, C. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, **22** 307–330.

Blitzstein, J. and Diaconis, P. (2006). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Unpublished manuscript.

Clauset, A., Moore, C. and Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, **453**.

Cox, D. (1972). The analysis of multivariate binary data. *Applied Statistics*, **21** 113–20.

de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, **27** 292–306.

Dorogovtsev, S. and Mendes, J. (2002). Evolution of networks. *Advances in Physics*, **51** 1079–1187.

Erdos, P. (1959). Graph theory and probability, i. *Canadian Journal of Mathematics*, **11** 34–38.

Erdos, P. (1961). Graph theory and probability, ii. *Canadian Journal of Mathematics*, **13** 346–352.

Erdos, P. and Renyi, A. (1959). On random graphs. *Publications Mathematicae*, **6** 290.

Erdos, P. and Renyi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute - Hungarian Academy of Science*, **5** 17.

Erdos, P. and Renyi, A. (1961). On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, **12** 261–267.

Fienberg, S. and Wasserman, S. (1981). Categorical data analysis of single sociometric relations. *Sociological methodology*, **11** 156–192.

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

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence*, **6** 721–41.

Gilbert, E. (1959). Random graphs. *Annals of Mathematical Statistics*, **30** 1141–1144.

Goldenberg, A., Zheng, A. X., Fienberg, S. E. and Airoldi, E. (2009). A survey of statistical network models. Unpublished manuscript.

Handcock, M. S. and Morris, M. (2006). A simple model for complex networks with arbitrary degree distribution and clustering. In *Statistical Network Analysis: Models*. 103.

Handcock, M. S., Raftery, A. E. and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A*, **170** 301354.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, **295** 286–295.

Hoff, P. D. (2007a). Extending the rank likelihood for semiparametric copula estimation. *Annals of Applied Statistics*, **1** 265–283.

Hoff, P. D. (2007b). Modeling homophily and stochastic equivalence in symmetric relational data. *Neural Information Processing Systems*, **20**.

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** 1090–1098.

Hoff, P. D. and Ward, M. D. (2003). Modeling dependencies in international networks. Tech. rep., University of Washington.

Holland, P. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, **76** 33–65.

Imai, K. and van Dyk, D. (2005). A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, **124** 311–334.

Jordan, M., Ghahramani, Z., Jaakkola, T. and Saul, L. (1999). Introduction to variational methods for graphical models. *Machine Learning*, **37** 183–233.

Klaassen, C. A. J. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, **3** 55–77.

Kolaczyk, E. (2009). *Statistical Analysis of Network Data*. Springer.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E. and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, **31** 204–213.

Leskovec, J., Chakrabarti, D., Kleinberg, J. and Faloutsos, C. (2005). Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *ACM SIGKDD*. 133.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, **22** 55.

Linkletter, C. (2007). *Spatial Process Models for Social Network Analysis*. Ph.D. thesis, Simon Fraser University.

Liu, C. (2004). Robit regression: A simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. Wiley, 227–238.

McCulloch, R. and Rossi, P. (1994). An exact likelihood analysis of the multinomial probit model. *J. Economet*, **64** 207–40.

Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, **80** 267–78.

Milgram, S. (1967). The small world problem. *Psychology Today*, **2** 60–67.

Newman, M. (2002). Assortative mixing in networks. *Physical Review Letters*, **89** 208701.

Newman, M., Strogatz, S. and Watts, D. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys Rev E Stat Nonlin Soft Matter Phys*, **64** 026118.

Nickel, C. L. M. (2006). *Random Dot Product Models: A Model for Social Networks*. Ph.D. thesis, Johns Hopkins University.

Nobile, A. (1998). A hybrid markov chain for the bayesian analysis of the multinomial probit model. *Statistics and Computing*, **8** 229–242.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic block-structures. *Journal of the American Statistical Association*, **96** 1077–1087.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. vii. on the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A*, **195** 1–47.

Pitt, M., Chan, D. and Kohn, R. (2006). Efficient bayesian inference for gaussian copula regression models. *Biometrika*, **93** 537–554.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Tech. rep., Danish Institute for Educational Research. From expanded edition (1980) with foreword and afterword by BD Wright.

Shaw, W. T. and Lee, K. (2007). Copula methods vs canonical multivariate distributions: the multivariate student t distribution with general degrees of freedom. Tech. rep., King's College London.

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

Snijders, T. A. B. and Nowicki, K. (1997). Estimation and prediction for stochastic block-structures for graphs with latent block structure. *Journal of Classification*, **14** 75–100.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82** 528–540.

van Duijn, M., Snijders, T. and Zijlstra, B. (2004). p2: a random effects model with covariates for directed graphs. *Statist. Neerland*, **58** 234–254.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, **82** 8–19.

Wasserman, S. and Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks*, **9** 1–36.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov random graphs and p-star. *Psychometrika*, **60** 401426.

Watts, D. and Strogatz, S. (1998). Collective dynamics of "small-world" networks. *Nature*, **393** 440–442.

Wong, G. Y. (1987). Bayesian models for directed graphs. *Journal of the American Statistical Association*, **82** 397.

Zheng, T., Salganik, M. J. and Gelman, A. (2006). How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association*, **101**.