Permutation Models for Relational Data*

Carter T. Butts†‡

December 1, 2004
v0.1

Abstract

We here propose an exponential family of permutation models that is suitable for inferring the direction and strength of association among dyadic relational structures. A linear-time algorithm is shown for MCMC simulation of model draws, as is the use of simulated draws for maximum likelihood estimation (MCMC-MLE) and/or estimation of Monte Carlo standard errors. We also provide an easily performed maximum pseudo-likelihood estimation procedure for the permutation model family, which provides a reasonable means of generating seed models for the MCMC-MLE procedure. Use of the modeling framework is demonstrated via an application involving relationships among managers in a high-tech firm.

Keywords: relational data, discrete exponential families, permutation models, graph comparison

1 Introduction

A common problem in sociology, psychology, biology, geography, and management science is the comparison of dyadic relational structures (i.e., graphs). Where these structures are formed on a common set of elements, a natural question which arises is whether there is a tendency for elements which are strongly connected in one set of structures to be more – or less – strongly connected within another set. We may ask, for instance, whether there is a

*This paper is based upon work supported by the National Science Foundation under award NSF ITR #IIS-0331707.

†Department of Sociology and Institute for Mathematical Behavioral Sciences, University of California - Irvine; SSPA 2145, Irvine, CA 92697; buttsc@uci.edu

‡The author would like to thank Mark Handcock, David Hunter, Hal Stern, and Tom Snijders for their input and advice.

correspondence between golf games and business deals, trade and warfare, or spatial proximity and genetic similarity. In each case, the data for such comparisons may be continuous or discrete, and multiple relations may be involved simultaneously (e.g., when comparing multiple measures of international trade volume with multiple types of political interactions).

Because dyadic relational data can be represented in adjacency matrix form[1], the above questions can be viewed as problems of matrix comparison. Conventionally, such problems are attacked either via the use of bimatrix association statistics (such as Moran's I (Moran, 1950)) with known sampling properties, or else by fitting standard models (e.g., OLS) to the vectorized matrices and evaluating parameters via Monte Carlo or permutation tests (Mantel, 1967). This last approach was pursued extensively by Hubert (1987), and variants have been widely used in the social sciences and geography (see, e.g., Sokal, 1979; Gale et al., 1983; Nakao and Romney, 1984; Yeung and Martin, 2003; Gibbons and Olk, 2003). In the field of social network analysis, in particular, such methods are the dominant approach to graph comparison (having been popularized by Krackhardt, 1987b; 1988); this is both because the permutation test has an appealing substantive interpretation, and because such tests are robust to row, column, and block autocorrelation in the input data. Such autocorrelation arises naturally in data provided by human raters, and is problematic for conventional null hypothesis tests (Krackhardt, 1988).

Although matrix permutation tests are a powerful tool, their scope is necessarily limited. Extensions to the multivariate case have proven difficult: general procedures for testing single-parameter hypotheses are known only for OLS, and a procedure which is robust to third-variable effects has only recently become available (Dekker et al., 2003). Because many problems of relational comparison involve dichotomous matrices, OLS solutions are often difficult to interpret, and prediction is generally infeasible. More tailored, likelihood-based approaches to multivariate graph modeling have been proposed (see, e.g., Pattison and Wasserman, 1999), but these are currently restricted to the discrete case. Likewise, such models do not leverage the attractive substantive interpretation of the permutation distribution articulated by Hubert (1987), who recognized that matrix permutations can be viewed as *assignments* of objects to positions within a relational structure. By conditioning on the permutation distribution, we ask in effect whether there is a tendency for certain objects to be assigned to certain positions, holding the underlying structure constant. In many empirical contexts (e.g.,

[1] I.e., for a relation R on set V, a matrix A such that $A_{ij} = iRj$ for $i, j \in A$.

spatial layout, formal organizational structure), there are sound theoretical reasons to assume some or all of the relations under study to be exogenous to the assignment process. The special case of ranked comparison is one well-known example of such a context, where permutation models have been profitably applied to problems ranging from preference analysis (Luce, 1959) to ranking of sports teams (Stern, 1990; Holder and Nevill, 1997). Alternately, we may condition on underlying structures to avoid the problems inherent in modeling them directly, particularly where extant theory (especially regarding error processes) is limited.

Our objective in the present paper is to bring forward the logic of Hubert's approach (along with its theoretical and practical advantages) within a likelihood-based framework. Our general procedure is to define an exponential family of distributions on a set of permutation vectors, whose sufficient statistics are given by product moments of the matrices to be compared. Although this family involves an unknown normalizing factor, we use Markov chain Monte Carlo (MCMC) methods to simulate model draws and to estimate the normalized likelihood. Maximum likelihood is then employed to obtain parameter estimates, via refinement of initial estimates produced by a heuristic pseudo-likelihood method. After providing modeling details (and discussing associated procedures for assessing uncertainty), we demonstrate the use of the model on a multivariate relational data set involving advice-seeking behavior in a high-tech manufacturing firm. We close with some final comments regarding the relationship of the present model with exponential family random graph models in the statistical literature.

1.1 Preliminaries

We begin by assuming two sets of (possibly valued) graphs, G_Y and G_X, on common finite vertex tuple $V_Y = V_X = (v_1, \ldots, v_N)$. These graphs are represented here by their respective adjacency arrays, $\mathbf{Y}$ and $\mathbf{X}$, where $\mathbf{Y}_{ijk}$ and $\mathbf{X}_{ijk}$ are the values of the (v_j, v_k) edges in the ith graphs of G_Y and G_X (respectively). (Per standard practice (e.g., Wasserman and Faust, 1994), edges which are not present are assumed to be of zero value.) We denote a permutation (or relabeling) of V_Y by a vector $\mathbf{a}$ on the integers $1, \ldots, N$, such that $\mathbf{a}_i = j$ iff $\mathbf{a}$ maps vertex v_i onto vertex v_j. Obviously, the mapping implied by $\mathbf{a}$ must be one-to-one. Thus, the value of the (v_j, v_k) edge in the ith graph of G_Y under permutation $\mathbf{a}$ is given by $\mathbf{Y}_{i\mathbf{a}_j\mathbf{a}_k}$. Because our interest in this paper is in the correspondence of G_Y and G_X, we will focus only on the permutation of V_Y relative to V_X.

Let $\mathbb{A}$ be the set of realizable permutations of V_Y, and let $\mathbf{a} \in \mathbb{A}$ be a permutation vector; we then let $p(\mathbf{a})$ represent a pmf on $\mathbb{A}$. $\mathbb{A}$ may be equal to the permutation group on V_Y, or may be any subgroup thereof (e.g., if certain vertices in V_Y cannot be paired with certain vertices in V_X). We may imagine $p(\mathbf{a})$ to result from a hypothetical process which draws the realized $\mathbf{a}$ from $\mathbb{A}$ – that is, which assigns vertices to positions within G_Y – in a manner which depends upon the structures of G_Y and G_X. Our aim here, then, will be to parameterize $p(\mathbf{a})$ so as to capture the edgewise association between graph sets.

2 An Exponential Family of Models for Permutations on Graph Sets

Given our two sets of graphs, G_Y and G_X, we now develop a model for the edgewise association between the two sets. Although we have taken the respective graph edge sets as fixed, we entertain the possibility that vertices might be assigned to various positions within G_Y relative to G_X; some of these assignments can be expected to yield adjacency structures which correspond more closely to those of G_X, while others can be expected to induce structures which diverge. In particular, consider the matrix product moment statistic $t(\mathbf{Y}_{i\cdot\cdot}, \mathbf{X}_{j\cdot\cdot}, \mathbf{a}) = \sum_{g=1}^{N} \sum_{h=1}^{N} \mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h} \mathbf{X}_{jgh}$. t provides a natural measure of the extent of association between the graphs i and j of G_Y and G_X under $\mathbf{a}$, being proportional to the matrix covariance when $\mathbf{Y}_{i\cdot\cdot}$ and $\mathbf{X}_{j\cdot\cdot}$ are mean-centered, and equal to the number of edge matchings in the dichotomous case. Summing t across all i, j given a particular $\mathbf{a}$ thus provides a measure of the extent of edgewise association between G_Y and G_X under $\mathbf{a}$.

If we now consider the distribution of $\mathbf{a}$, one fairly straightforward approach is to parameterize $p(\mathbf{a})$ via an exponential family having t as the set of sufficient statistics. In particular, define $\Theta \in \mathbb{R}^{M \times P}$, and let

$$p(\mathbf{a}|\mathbf{Y},\mathbf{X},\Theta) = \frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij} t(\mathbf{Y}_{i\cdot\cdot},\mathbf{X}_{j\cdot\cdot},\mathbf{a})\right)}{\sum_{\mathbf{a}'\in\mathbb{A}}\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij} t(\mathbf{Y}_{i\cdot\cdot},\mathbf{X}_{j\cdot\cdot},\mathbf{a}')\right)}\mathbb{I}_{\mathbb{A}}(\mathbf{a}) \tag{1}$$

$$= \frac{\exp\left(\mathbb{A}\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}{\sum_{\mathbf{a}'\in\mathbb{A}}\exp\left(\mathbb{A}\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}'_g\mathbf{a}'_h}\mathbf{X}_{jgh}\right)}\mathbb{I}_{\mathbb{A}}(\mathbf{a}) \tag{2}$$

$$= \frac{\exp\left(\mathbb{A}\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}{\chi_{\mathbb{A}}(\Theta)}\mathbb{I}_{\mathbb{A}}(\mathbf{a}), \tag{3}$$

where $\chi_{\mathbb{A}}$ is a normalizing factor, and $\mathbb{I}_{\mathbb{A}}$ is an indicator function for $\mathbb{A}$. Note that the use of discrete exponential families to parameterize permutations has a reasonably long history (see, e.g., DeGroot et al., 1971; Plackett, 1975), with typical models being parameterized in terms of conditional probabilities of "victory" in n-adic contests (Plackett, 1975) or order statistics (Henery, 1981). The approach taken here is closest in spirit to models for the broken sample problem (DeGroot et al., 1971; Goel, 1975), of which the present problem may be seen as an inverted version (i.e., we infer latent association parameters from matchings, rather than inferring a latent matching from an assumed association). Given G_Y, G_X, Equation 3 can also be seen as a joint distribution on the corresponding graph sets; we will return to this issue in Section 5.

Clearly, $\Theta_{ij} > 0$ implies a tendency towards positive association between graphs i and j of G_Y and G_X (respectively), while $\Theta_{ij} < 0$ implies the opposite tendency. This is perhaps most easily seen by considering the probability ratio between two hypothetical permutations, $\mathbf{a}$ and $\mathbf{a}'$. Using Equation 3 above, we have

$$\frac{p(\mathbf{a}'|\mathbf{Y},\mathbf{X},\Theta)}{p(\mathbf{a}|\mathbf{Y},\mathbf{X},\Theta)} = \exp\left(\sum_{i=1}^{P}\sum_{j=1}^{M}\sum_{g=1}^{N}\sum_{h=1}^{N}\Theta_{ij}\left(\mathbf{Y}_{i\mathbf{a}'_g\mathbf{a}'_h}\mathbf{X}_{jgh} - \mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)\right) \tag{4}$$

$$= \exp\left(\sum_{i=1}^{P}\sum_{j=1}^{M}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\left(\mathbf{Y}_{i\mathbf{a}'_g\mathbf{a}'_h} - \mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\right)\mathbf{X}_{jgh}\right). \tag{5}$$

Thus, Θ_{ij} acts as a weight on the edgewise product moment difference. In the dichotomous case, this is simply the change in the number of edge matches as one moves from $\mathbf{a}$ to $\mathbf{a}'$; each unit increase in the number of edge matches between the ith and jth graphs of G_Y and G_X, then, multiplies the probability ratio by $\exp(\Theta_{ij})$.

One obvious difficulty with this model is the calculation of the normalizing factor, $\chi_{\mathbb{A}}(\Theta)$. Since $|\mathbb{A}|$ will generally be on the order of $N!$, direct enumeration of $\chi_{\mathbb{A}}(\Theta)$ is rarely feasible; given this, we resort to Monte Carlo methods, as shown in Section 3.2. Fortunately, knowledge of $\chi_{\mathbb{A}}(\Theta)$ is not required for simulation of draws from the permutation distribution. It is to this problem that we now turn.

2.1 Simulation via the Metropolis Algorithm

Although the distribution of Equation 3 contains an unknown normalizing factor, this factor depends only on $\mathbb{A}$ and Θ. Thus, as shown in Equation 5, the probability ratio between any two permutations of V_Y depends only on the numerator of Equation 3, a number which is much more easily calculated. This immediately suggests the use of a Metropolis algorithm (see, e.g., Gamerman, 1997) to simulate draws from the permutation distribution, since the Metropolis algorithm is based exclusively on such ratios. In particular, let us imagine that on the nth iteration of the algorithm, we consider a move from permutation $\mathbf{a}$ to candidate permutation $\mathbf{a}'$. The probability of accepting the $\mathbf{a}'$ move is then given by Equation 5 if $\frac{p(\mathbf{a}'|\mathbf{Y},\mathbf{X},\Theta)}{p(\mathbf{a}|\mathbf{Y},\mathbf{X},\Theta)} \leq 1$, or 1 otherwise.

While the above is an improvement over $N!$ complexity, it still leaves something to be desired: the complexity of calculating Equation 3 is quadratic in N, which is potentially prohibitive for large graphs. To obtain a more efficient algorithm, we first consider the special case in which $\mathbf{a}'$ is formed by exchanging two entries of $\mathbf{a}$ (i.e., a dyadic swap). For such a dyadic exchange of the j and kth entries of $\mathbf{a}$, we may write the changescore from Equation 5 as

$$\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\Big[\mathbf{Y}_{i\mathbf{a}'_g\mathbf{a}'_h}\mathbf{X}_{jgh} - \mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\Big] \equiv \sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\Delta_{ij}(\mathbf{a},k,l,\mathbf{Y},\mathbf{X}), \tag{6}$$

where Δ reflects the change in the corresponding sufficient statistic. Given that only two entries of $\mathbf{a}$ have been altered, Δ can then be expressed as

$$\Delta_{ij}(\mathbf{a},k,l,\mathbf{Y},\mathbf{X}) = \left(\sum_{h\in\{1,\ldots,N\}\setminus\{k,l\}} \Big[(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_h}\mathbf{X}_{jkh} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_h}\mathbf{X}_{jkh}) + (\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_h}\mathbf{X}_{jlh} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_h}\mathbf{X}_{jlh}) \right.$$
$$\left. + (\mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_l}\mathbf{X}_{jhk} - \mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_k}\mathbf{X}_{jhk}) + (\mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_k}\mathbf{X}_{jhl} - \mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_l}\mathbf{X}_{jhl}) \Big] \right)$$
$$+ \Big[(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_l}\mathbf{X}_{jkk} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_k}\mathbf{X}_{jkk}) + (\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_k}\mathbf{X}_{jkl} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_l}\mathbf{X}_{jkl})$$
$$+ (\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_l}\mathbf{X}_{jlk} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_k}\mathbf{X}_{jlk}) + (\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_k}\mathbf{X}_{jll} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_l}\mathbf{X}_{jll}) \Big] \tag{7}$$

$$= \left(\sum_{h\in\{1,\ldots,N\}\setminus\{k,l\}} \Big[\mathbf{X}_{jkh}\left(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_h} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_h}\right) + \mathbf{X}_{jlh}\left(\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_h} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_h}\right) \right.$$
$$\left. + \mathbf{X}_{jhk}\left(\mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_l} - \mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_k}\right) + \mathbf{X}_{jhl}\left(\mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_k} - \mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_l}\right) \Big] \right)$$
$$+ \Big[\mathbf{X}_{jkk}\left(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_l} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_k}\right) + \mathbf{X}_{jkl}\left(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_k} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_l}\right) + \mathbf{X}_{jlk}\left(\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_l} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_k}\right)$$
$$+ \mathbf{X}_{jll}\left(\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_k} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_l}\right) \Big] \tag{8}$$

$$= \left(\sum_{h\in\{1,\ldots,N\}\setminus\{k,l\}} \Big[\left(\mathbf{X}_{jkh} - \mathbf{X}_{jlh}\right)\left(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_h} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_h}\right) \right.$$
$$\left. + \left(\mathbf{X}_{jhk} - \mathbf{X}_{jhl}\right)\left(\mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_l} - \mathbf{Y}_{i\mathbf{a}_h\mathbf{a}_k}\right) \Big] \right) .$$
$$+ \Big[\left(\mathbf{X}_{jkk} - \mathbf{X}_{jll}\right)\left(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_l} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_k}\right) + \left(\mathbf{X}_{jkl} - \mathbf{X}_{jlk}\right)\left(\mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_k} - \mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_l}\right)$$
$$+ \mathbf{X}_{jlk}\left(\mathbf{Y}_{i\mathbf{a}_k\mathbf{a}_l} - \mathbf{Y}_{i\mathbf{a}_l\mathbf{a}_k}\right) \Big] \tag{9}$$

The complexity of this calculation is linear in N, a substantial improvement. Using Equation 9 in conjunction with the procedure shown in Algorithm 1 permits efficient simulation of draws from the permutation model, even for graphs of reasonably large size (e.g., hundreds of vertices). Convergence of the Markov Chain can be monitored by examination of the sufficient statistics, e.g., using the methods of Geweke (1992) or Gelman and Rubin (1992). In some circumstances, it may be desirable to consider moves which are larger than single dyadic exchanges; in these cases, dyadic exchanges followed by updating of statistics using Equation 9 (lines 2.1–2.1) can be applied multiple times prior to the keep or reject decision (lines 2.1–2.1). Note that multiple dyadic exchanges per iteration may be required if $\mathbb{A}$ is a restricted subgroup of the full permutation group, to ensure irreducibility of the Markov chain. While this does increase the computational complexity of each iteration, the overall result is still linear in N for any fixed number of exchanges.

Algorithm 1 A Dyadic-Exchange Metropolis Algorithm for the Permutation Model

Let $\mathbf{a}^{(0)} := (1, \ldots, N)$
Let $i := 1$
repeat
  Let $\mathbf{a}^{(i)} := \mathbf{a}^{(i-1)}$
  Draw j, k uniformly from $(1, \ldots, N)$
  Let $l := \mathbf{a}^{(i)}_j$
  Let $\mathbf{a}^{(i)}_j := \mathbf{a}^{(i)}_k$
  Let $\mathbf{a}^{(i)}_k := l$
  Draw u from $U(0, 1)$
  if $u > \frac{p(\mathbf{a}^{(i)}|\mathbf{Y},\mathbf{X},\Theta)}{p(\mathbf{a}^{(i-1)}|\mathbf{Y},\mathbf{X},\Theta)}$ **then**
    Let $\mathbf{a}^{(i)} := \mathbf{a}^{(i-1)}$
  end if
  Let $i := i + 1$
until $\mathbf{a}^{(\cdot)} \sim p(\mathbf{a}|\mathbf{Y}, \mathbf{X}, \Theta)$
return $\mathbf{a}^{(\cdot)}$

3 Parameter Estimation

Thus far, we have treated Θ as a known matrix of parameters, expressing the extent to which the elements of G_Y match those of G_X. Here, we assume $\mathbf{a}$ to be known (and, without loss of generality, equal to the identity permutation on V_Y), and seek instead to conduct inference for an unknown Θ. Our approach proceeds in two stages: first, we demonstrate an easily calculated maximum pseudo-likelihood estimator (MPLE) for Θ; second, we use the Markov chain Monte Carlo maximum likelihood approach of Geyer and Thompson (1992) to obtain an approximate maximum likelihood estimator (the MCMC-MLE). The properties of the MPLE are not guaranteed, but it serves as a useful seed value for the MCMC-MLE calculation. Finally, we discuss the estimation of standard errors (and selected null hypothesis tests) for the MCMC-MLE.

3.1 Maximum Pseudo-Likelihood

To obtain an initial estimate of the parameter matrix, Θ, it is tempting to make use of our ability to calculate probability ratios for dyadic exchanges in linear time. In particular, let $\mathbf{a}$ be the observed vertex ordering, and let $\mathbf{a}^{lk}$ represent the permutation in which the k and lth vertices have been exchanged (with all others remaining in the original order); for clarity of notation, we designate the original permutation in this context by $\mathbf{a}^{kl}$. From Equation 5, then, we have

$$\frac{p(\mathbf{a}^{kl}|\mathbf{Y},\mathbf{X},\Theta)}{p(\mathbf{a}^{lk}|\mathbf{Y},\mathbf{X},\Theta)} = \exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\Delta_{ij}(\mathbf{a}^{kl},l,k,\mathbf{Y},\mathbf{X})\right). \quad (10)$$

Although this ratio is exact, it clearly depends upon the entire permutation. Consider, however, the restricted case in which we evaluate the probability ratio for $\mathbf{a}^{lk}$ versus $\mathbf{a}^{kl}$ conditional on the other elements of $\mathbf{a}$ (which we denote by $\mathbf{a}^{-kl}=\mathbf{a}^{-lk}$). Conditional on $\mathbf{a}^{-kl}$, either $\mathbf{a}^{lk}$ or $\mathbf{a}^{kl}$ must obtain. It follows, then, that the above can be rewritten in the conditional case as

$$\frac{p(\mathbf{a}^{kl}|\mathbf{a}^{-kl},\mathbf{Y},\mathbf{X},\Theta)}{1-p(\mathbf{a}^{kl}|\mathbf{a}^{-kl},\mathbf{Y},\mathbf{X},\Theta)} = \exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\Delta_{ij}(\mathbf{a}^{kl},l,k,\mathbf{Y},\mathbf{X})\right), \quad (11)$$

which gives us the conditional dyadic probability

$$p(\mathbf{a}^{kl}|\mathbf{a}^{-kl},\mathbf{Y},\mathbf{X},\Theta) = \left(1+\exp\left(-\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\Delta_{ij}(\mathbf{a}^{kl},l,k,\mathbf{Y},\mathbf{X})\right)\right)^{-1}. \quad (12)$$

Under the assumption that $\mathbf{a}^{kl}$ represents the observed permutation, $p(\mathbf{a}^{kl}|\mathbf{a}^{-kl},\mathbf{Y},\mathbf{X},\Theta)$ is the conditional likelihood of the k,l dyad labeling, given the labelings of the other vertices in V_Y (i.e., the chance that the identities of k and l would not have been switched, given the identities of the other vertices). Although this represents only a portion of the information contained within $\mathbf{a}$, it is natural to ask whether an estimator might be constructed from such conditional likelihoods. If we consider the set of all $\binom{N}{2}$ possible dyadic exchanges, one estimator of Θ is that which maximizes the product of conditional likelihoods, i.e.

$$\hat{\Theta} = \arg \max_{\Theta \in \mathbb{R}^{M \times P}} \prod_{k=1}^{N} \prod_{l=k+1}^{N} p(\mathbf{a}^{kl} | \mathbf{a}^{-kl}, \mathbf{Y}, \mathbf{X}, \hat{\Theta}). \tag{13}$$

This product of conditional likelihoods is a pseudo-likelihood for the full permutation model (in the sense of Besag, 1975), making $\hat{\Theta}$ as given in Equation 13 a maximum pseudo-likelihood estimator. Given Equation 12, the MPLE can be readily calculated using standard optimization methods; since calculation of the pseudo-likelihood involves $\binom{N}{2}$ computations of order N, it follows that the total complexity of the pseudo-likelihood itself is order N^3. Although this scaling is not favorable, the fact that the pseudo-likelihood is generally calculated only a small number of times in practice (as opposed to the large number of likelihood ratio calculations needed for a typical Metropolis simulation) makes it computationally attractive in most cases.

It should be noted that the asymptotic properties of this MPLE are not known, and there is no guarantee that the estimator will closely approximate the MLE in the general case. For this reason, direct use of the MPLE is not recommended. The main value of the MPLE lies in its potential to act as an easily calculated heuristic approximation to the MLE, which can be used as the seed value for a Monte Carlo-based approach; it is to this last that we now turn.

3.2 Monte Carlo Maximum Likelihood

To obtain a true maximum likelihood estimate for Θ, we must maximize the probability given in Equation 3; as already noted, however, this expression contains a normalizing factor for which direct calculation is infeasible. Following Geyer and Thompson (1992), our approach to this problem is to use our ability to simulate draws from the permutation distribution to generate an importance sample which can, in turn, be used to approximate the desired likelihood. (Such an approach is now widely used in the modeling of relation data – see, e.g., Crouch et al., 1998; Snijders, 2002; Hunter and Handcock, 2004, .) As a practical matter, we begin with a baseline model (denoted here as simply Θ) and attempt to identify a parameter matrix $\hat{\Theta}$

such that $\frac{p(\mathbf{a}|\hat{\Theta},\mathbf{Y},\mathbf{X})}{p(\mathbf{a}|\Theta,\mathbf{Y},\mathbf{X})}$ is maximized. Such a matrix is obviously an MLE for the permutation model.

For a given permutation vector, $\mathbf{a} \in \mathbb{A}$, we may write the likelihood ratio to be maximized as

$$\frac{p(\mathbf{a}|\hat{\Theta},\mathbf{Y},\mathbf{X})}{p(\mathbf{a}|\Theta,\mathbf{Y},\mathbf{X})} = \frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}\frac{\chi_{\mathbb{A}}\left(\Theta\right)}{\chi_{\mathbb{A}}\left(\hat{\Theta}\right)} \tag{14}$$

$$= \exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\left(\hat{\Theta}_{ij}-\Theta_{ij}\right)\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)\frac{\chi_{\mathbb{A}}\left(\Theta\right)}{\chi_{\mathbb{A}}\left(\hat{\Theta}\right)}. \tag{15}$$

While the first of these factors is easily computable, the ratio of normalizing factors is obviously problematic. Fortunately, an excellent approximation to this ratio may be obtained via Monte Carlo quadrature. To see this, we begin by noting the following standard result:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(i)}\mathbf{a}_h^{(i)}}\mathbf{X}_{jgh}\right)}{p(\mathbf{a}^{(k)})}$$

$$\to \sum_{\mathbf{a}'\in\mathbb{A}}\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}'_g\mathbf{a}'_h}\mathbf{X}_{jgh}\right) \tag{16}$$

$$= \chi_{\mathbb{A}}\left(\hat{\Theta}\right), \tag{17}$$

where $\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(n)}$ are independent draws from some stochastic process with density p on support $\mathbb{A}$. (This is simply a Monte Carlo quadrature with importance density p.) Now, let the above importance density be the permutation model with known parameter matrix Θ. Substitution from

Equation 3 then gives us

$$\frac{1}{n}\sum_{k=1}^{n}\frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(k)}\mathbf{a}_h^{(k)}}\mathbf{X}_{jgh}\right)}{p(\mathbf{a}^{(k)})} =$$

$$\frac{1}{n}\sum_{k=1}^{n}\left[\frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(k)}\mathbf{a}_h^{(k)}}\mathbf{X}_{jgh}\right)}{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(k)}\mathbf{a}_h^{(k)}}\mathbf{X}_{jgh}\right)}\chi_{\mathbb{A}}\left(\Theta\right)\right], \quad (18)$$

and pulling out $\chi_{\mathbb{A}}\left(\Theta\right)$ yields

$$= \chi_{\mathbb{A}}\left(\Theta\right)\left[\frac{1}{n}\sum_{k=1}^{n}\frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}\right]. \quad (19)$$

Returning to the limit expressed in Equation 16, we may substitute from Equation 19 to get

$$\lim_{n\to\infty}\chi_{\mathbb{A}}\left(\Theta\right)\left[\frac{1}{n}\sum_{k=1}^{n}\frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)}\right] \to \chi_{\mathbb{A}}\left(\hat{\Theta}\right). \quad (20)$$

Dividing through by $\chi_{\mathbb{A}}\left(\Theta\right)$ then gives us

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\hat{\Theta}_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(k)}\mathbf{a}_h^{(k)}}\mathbf{X}_{jgh}\right)}{\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\Theta_{ij}\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(k)}\mathbf{a}_h^{(k)}}\mathbf{X}_{jgh}\right)} \to \frac{\chi_{\mathbb{A}}\left(\hat{\Theta}\right)}{\chi_{\mathbb{A}}\left(\Theta\right)}, \quad (21)$$

which is the reciprocal of the ratio of normalizing factors contained in Equation 15. Putting all this together, then, gives us the useful result that

$$\lim_{n\to\infty}\left[\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\left(\hat{\Theta}_{ij}-\Theta_{ij}\right)\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{jgh}\right)\right.$$
$$\left.\left[\frac{1}{n}\sum_{k=1}^{n}\exp\left(\sum_{i=1}^{M}\sum_{j=1}^{P}\left(\hat{\Theta}_{ij}-\Theta_{ij}\right)\sum_{g=1}^{N}\sum_{h=1}^{N}\mathbf{Y}_{i\mathbf{a}_g^{(k)}\mathbf{a}_h^{(k)}}\mathbf{X}_{jgh}\right)\right]^{-1}\right] \to \frac{p(\mathbf{a}|\hat{\Theta},\mathbf{Y},\mathbf{X})}{p(\mathbf{a}|\Theta,\mathbf{Y},\mathbf{X})} \quad (22)$$

where the permutation vectors $\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(n)}$ are drawn from the permutation model of Equation 3 with parameter matrix Θ. Given this, we calculate the approximate MLE by first taking a sample of draws from the initial model, and then by maximizing the Monte Carlo approximation to $\frac{p(\mathbf{a}|\hat{\Theta},\mathbf{Y},\mathbf{X})}{p(\mathbf{a}|\Theta,\mathbf{Y},\mathbf{X})}$ using standard methods, (e.g., Newton-Rapheson or the like). We refer to this estimate as the MCMC-MLE.

Although Equation 22 shows only that the sum over the importance sample converges as $n \to \infty$, a finite sum will (in practice) provide a good estimate of the likelihood ratio so long as n is sufficiently large and the candidate parameter matrix, $\hat{\Theta}$, is reasonably close to the baseline parameter matrix, Θ. (The adequacy of this approximation for any given sample can be assessed by utilizing the asymptotic normality of the Monte Carlo estimator to obtain stochastic error bounds.) A more serious concern is the selection of the baseline parameter matrix. If Θ is a poor approximation to the MLE, then the importance ratio of Equation 16 will place a great deal of weight on a small number of draws, and convergence will be slow. At the same time, the computational burden of calculating Θ should not be excessive. Given both of these concerns, we here employ the MPLE as the baseline value for Θ.

3.3 Uncertainty and Fit Assessment

While raw parameter estimates are of obvious interest, they are of dubious value without associated measures of uncertainty. Similarly, it is useful to have some sense of the substantive fit of the model to data, above and beyond comparison with null hypotheses of no association. Here, we sketch some basic approaches to each issue, beginning with the estimation of standard errors for model parameters.

3.3.1 Standard Errors and Null Hypothesis Tests

Given that our present model family belongs to the class of regular exponential families (Barndorff-Nielsen, 1978), it is tempting to seek to apply standard asymptotic results regarding the sampling properties of the MLE (see, e.g., Johansen, 1979). In particular, a result with the general form $\hat{\Theta} \to \lim_{N\to\infty} \mathcal{N}\left(\Theta, J\left(\Theta\right)^{-1}\right)$ (where J is the Fischer information matrix) would lead immediately to conventional standard errors, confidence intervals, and z-tests for hypotheses on Θ. This matter is subtle, however, in that convergence in N is not closely akin to convergence under repeated independent trials; one approach would be conjecture that for almost all sequences of adjacency array/permutation vector triples, $(\mathbf{Y}^{(i)}, \mathbf{X}^{(i)}, \mathbf{a}^{(i)})$, of increasing order with $\mathbf{a}^{(i)}$ drawn from a distribution with mass $p\left(\mathbf{a}^{(i)} \middle| \Theta, \mathbf{Y}^{(i)}, \mathbf{X}^{(i)}\right)$,

$\hat{\Theta}^{(i)} \to \lim_{N\to\infty} \mathcal{N}\left(\Theta, \left(J\left(\Theta\right)^{(i)}\right)^{-1}\right)$. While this conjecture seems not immediately unreasonable,[2] its proof or denial is not currently available. Thus, asymptotic arguments regarding the sampling behavior of MLEs for the permutation model should be considered heuristic at present.

An alternative to heuristic use of asymptotics lies in our capacity to simulate draws from the permutation model using Algorithm 1. In particular, we may estimate the sampling distribution of $\hat{\Theta}$ by drawing an MCMC sample $\left(\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(n)}\right)$ from $p\left(\mathbf{a} \left|\Theta, \mathbf{Y}, \mathbf{X}\right.\right)$ and then fitting the model to each draw to obtain $\left(\hat{\Theta}^{(1)}, \ldots, \hat{\Theta}^{(n)}\right)$. Assuming that the original MCMC sample has converged, the resulting sample of $\hat{\Theta}$ values is drawn from the sampling distribution of the MLE. Confidence intervals and standard errors can be obtained directly from this distribution. For testing of point hypotheses, another alternative is to repeat the above procedure after simulating draws from $p\left(\mathbf{a} \left|\Theta^H, \mathbf{Y}, \mathbf{X}\right.\right)$, where Θ^H is the hypothesized parameter vector. The observed parameter quantiles in the replicated $\hat{\Theta}$ distribution then provide Monte Carlo p-values against the hypothesis that $\Theta = \Theta^H$ (see also Besag and Clifford, 1989). For nested hypotheses, it may also be of interest to compare the maximized log-likelihood to its sampling distribution. For any given $\hat{\Theta}$ and importance sample $\left(\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{n}\right)$ drawn using parameter matrix Θ, $p\left(\mathbf{a} \middle| \mathbf{Y}, \mathbf{X}, \hat{\Theta}\right)$ via the quadrature

$$p\left(\mathbf{a} \middle| \mathbf{Y}, \mathbf{X}, \hat{\Theta}\right) \approx \exp\left(\sum_{j=1}^{M}\sum_{k=1}^{P} \hat{\Theta}_{jk} \sum_{g=1}^{N}\sum_{h=1}^{N} \mathbf{Y}_{j\mathbf{a}_g\mathbf{a}_h}\mathbf{X}_{k\mathbf{a}_g\mathbf{a}_h}\right)$$
$$\times \frac{\sum_{i=1}^{n} \exp\left(-\sum_{j=1}^{M}\sum_{k=1}^{P} \Theta_{jk} \sum_{g=1}^{N}\sum_{h=1}^{N} \mathbf{Y}_{j\mathbf{a}_g^{(i)}\mathbf{a}_h^{(i)}}\mathbf{X}_{k\mathbf{a}_g^{(i)}\mathbf{a}_h^{(i)}}\right)}{N! \sum_{i=1}^{n} \exp\left(\sum_{j=1}^{M}\sum_{k=1}^{P} \left(\hat{\Theta}_{jk} - \Theta_{jk}\right) \sum_{g=1}^{N}\sum_{h=1}^{N} \mathbf{Y}_{j\mathbf{a}_g^{(i)}\mathbf{a}_h^{(i)}}\mathbf{X}_{k\mathbf{a}_g^{(i)}\mathbf{a}_h^{(i)}}\right)}. \tag{23}$$

The quantile of the maximized log-likelihood for $\mathbf{a}$ under $\hat{\Theta}$ in the distribution of maximized log-likelihoods based on the Θ^H sample then provides a Monte Carlo p-value for nested hypothesis Θ^H.

As a practical matter, it should be noted that computation of Monte Carlo standard errors/p-values can be expensive in practice, since each

[2] Although a necessary condition for the above conjecture is that, for fixed Θ, the probability of $t\left(\mathbf{Y}^{(i)}, \mathbf{X}^{(i)}, \mathbf{a}^{(i)}\right)$ lying within the convex hull of the sample statistics approaches 1 as N approaches infinity. This would seem to be a fairly strong condition, casting some doubt on the conjecture.

model draw requires a corresponding importance sample. One means of reducing the computational burden is to reuse the original importance sample when computing the replicated models; provided that the replicated MLE is sufficiently "close" to the seed model, this procedure will produce reasonable estimates. In the event that this condition does not obtain (e.g., if the t values for the replicated data do not lie within the convex hull of the t values from the original importance sample), the replicated MCMC-MLE may fail to exist. As is generally the case with such methods, they are most efficacious when the model provides a good fit to the observed data.

3.3.2 Mean Value Parameterization and Substantive Interpretation of Relationship Strength

Quite apart from the issue of statistical significance, it is generally desirable to translate the strength of association between the graphs of G_Y and G_X into substantively interpretable terms. While the log-odds interpretation of Θ under hypothetical dyadic exchanges is one such translation, the practical impact of Θ may also be understood by examining the predictive distribution of the t statistics under Θ. Let $\mathbf{E}_\Theta t\left(\mathbf{Y},\mathbf{X},\mathbf{a}\right)$ be the expectation of t given Θ. Following Barndorff-Nielsen (1978),

$$\tau\left(\Theta\right) \equiv \mathbf{E}_\Theta t\left(\mathbf{Y},\mathbf{X},\mathbf{a}\right) \tag{24}$$

is then the mean value parameterization of $p(\mathbf{a})$. Let $\tau_{ij} = \tau(\Theta)_{ij}$ be the mean value parameter corresponding to Θ_{ij}. Then it is a standard result that the MLE of the mean value parameter is given by $\hat{\tau_{ij}} = \mathbf{E}_{\hat{\Theta}} t_{ij}$ (Johansen, 1979). This has an immediate and natural interpretation for many applications, particularly where $\mathbf{X}$ and $\mathbf{Y}$ have been centered/scaled such that t_{ij} corresponds to a matrix correlation or covariance statistic. For instance, let $\mathbf{Y}^\rho_{i\cdot\cdot} = \frac{\mathbf{Y}_{i\cdot\cdot}-\mu_{\mathbf{Y}_{i\cdot\cdot}}}{\sigma_{\mathbf{Y}_{i\cdot\cdot}}\sqrt{N^2-1}}, \mathbf{X}^\rho_{j\cdot\cdot} = \frac{\mathbf{X}_{j\cdot\cdot}-\mu_{\mathbf{X}_{j\cdot\cdot}}}{\sigma_{\mathbf{X}_{j\cdot\cdot}}\sqrt{N^2-1}}$, where $\mu_\cdot$ and $\sigma_\cdot$ are the appropriate matrix means and standard deviations (respectively). Them t_{ij} is a matrix correlation, and $\hat{\tau_{ij}}$ is the expected matrix correlation between $\mathbf{Y}_{i\cdot\cdot}$ and $\mathbf{X}_{j\cdot\cdot}$ under the MLE. Although it is a standard result that $\hat{\tau}$ is equal to the vector of observed sufficient statistics at the MLE, less trivial insights can be gleaned from examining other properties of the predictive distribution.

One such family of useful properties is the set of predictive quantiles of t under $\hat{\Theta}$. Such quantiles can be interpreted as providing the range of association statistic values which would be expected under hypothetical replications of the data, given $\hat{\Theta}$. In the same manner, $\mathrm{Var}_\Theta t\left(\mathbf{Y},\mathbf{X},\mathbf{a}\right)$ may be used to succinctly describe the predictive variability of t at the MLE. This last may be estimated directly from MCMC draws (using Algorithm 1) or approximated from the estimated Hessian of the log-likelihood at the MLE (Johansen, 1979). Where the fitted model suggests a consistently strong association between graphs, we can then infer that the relationship in question is correspondingly pronounced; by contrast, a model which divides probability relatively evenly across positive and negative associations suggests a relationship of questionable practical importance, even if Θ_{ij} is highly significant.

4 Example: Advice-Seeking Behavior in a High-Tech Firm

To illustrate the use of the permutation model, we now turn to a data set collected by David Krackhardt (1987a) regarding interpersonal relations among managers in a high-tech manufacturing firm. In particular, we are here interested in the question of whether (and to what extent) managers' patterns of reported advice-seeking behavior are related to differences in status characteristics such age, level, and tenure within the firm, formal structural properties such as department membership and reporting relationships, and informal relationships (such as friendship). Given the underlying (i.e., unlabeled) structures of these networks, does there appear to be a tendency for managers to occupy positions in the advice network which induce consonance with (or dissonance from) these other structures? To determine this, we fit two permutation models to the Krackhardt data: first, a model using the untransformed matrices; and second, an equivalent model on the centered and rescaled matrices (making t equivalent to the matrix correlation statistic).

Data for this analysis was obtained from a survey of firm members. A questionnaire was administered to each of the firm's 21 managers; this questionnaire included both personal information (age, tenure in the firm, etc.) and interpersonal items. Among the interpersonal information elicited were attributions of personal friendship ("whom do you consider to be a close, personal friend") and advice-seeking behavior ("whom do you go to for help or advice at work"). Information was also obtained regarding each manager's department and level, as well as formal reporting responsibility (i.e., who reports to whom). The friendship, advice-seeking, and reporting networks were coded as dichotomous adjacency matrices, with a value of 1 indicating the presence of an edge in the associated graph. Valued adjacency matrices were also formed based on age, tenure, and level, by taking signed differences; thus, for such a matrix $X_{i\cdot\cdot}$, based on attribute vector $\mathbf{x}$, $X_{ijk} = \mathbf{x}_j - \mathbf{x}_k$. Finally, a dichotomous adjacency matrix was constructed in which two vertices were coded as adjacent iff they belonged to the same department.

Given the above, a permutation model was fit taking the advice network as G_Y and the other networks as G_X. Initial parameter estimates were obtained using maximum pseudo-likelihood, with subsequent refinement using Monte Carlo maximum likelihood estimation. Adequacy of convergence was assessed by taking simulated draws from the fitted model using the Metropolis algorithm, and comparing the mean values of the sufficient statistics with the observed values; deviations of more than 1.96 standard errors from the observed values on a sample of 100,000 MCMC draws were taken to indicate inadequate convergence. Where convergence was not reached, the model was re-fit using current MCMC-MLE estimates as starting values, with this process being repeated until convergence was obtained. The final model fit was based on an importance sample of 750,000 MCMC draws, uniformly thinned from an initial sample of 150 million. (One million initial burn-in draws were discarded.) One dyadic exchange was made per candidate move, and $\mathbb{A}$ was taken to consist of all permutations on V_Y.

Table 1 shows parameter estimates and fit information for the permutation model on the untransformed data. For comparison, both the MPLE and MCMC-MLE values are shown; while both are fairly similar, the differences are large enough to be substantively meaningful. Effects for friendship, reporting, and difference in tenure are significant at the 0.01 level (two-sided asymptotic z-test), while the p-value for difference in level is just above 0.05. Effects for age difference and departmental co-membership do not approach significance. Overall, the strongest effect is the association with ReportsTo, which multiplies the likelihood ratio for an exchange by approximately 35 for each additional edge match added (or divides it by 35 for each edge match removed). By comparison, an increase in the signed tenure difference of 10 years induced by an exchange would reduce the exchange probability ratio by about 20%. (Note that the negative sign indicates a tendency of advice ties to coincide with negative tenure differences, i.e., mangers report seeking advice from those with more experience than themselves.) Putting this together, we may reasonably infer a strong tendency for managers to report seeking advice from their immediate superiors (with some suggestion of a tendency to seek advice from higher-level managers generally), as well as a weaker tendency for managers to report seeking advice from friends and from more senior members of the firm. Since the analysis in question

	Estimate ($\hat{\Theta}$)				
Effect	MPLE	MCMC-MLE	s.e.	$p(>\|Z\|)$	
Advice×ReportsTo	3.159	3.573	1.185	0.0026	**
Advice×Friendship	0.616	0.527	0.199	0.0080	**
Advice×TenureDiff	-0.027	-0.028	0.009	0.0030	**
Advice×LevelDiff	-0.354	-0.193	0.103	0.0624	+
Advice×AgeDiff	0.003	0.009	0.006	0.1391	
Advice×SameDept	-0.540	-0.323	0.276	0.2418	
Log-Likelihood: -31.05					
AIC: 74.10					

Table 1: Permutation Model Fit for Advice-Seeking Behavior

is symmetric, it is important to emphasize that the above must be taken as statements of association rather than causality: it is equally correct to say that managers tend to supervise those who report seeking advice from them, for instance.

In terms of overall goodness-of-fit, it should be noted that a null model of random association (i.e., $\Theta = 0$ for all parameters) results in a log-likelihood of approximately -45.38 for this data set. This is a significantly worse fit under the asymptotic likelihood-ratio χ^2 ($\chi^2 = 28.66$, df= 6, $p < 0.001$), although such a performance standard is obviously quite minimal.

For another view of the same data, we fit the above model to the centered, scaled adjacency matrices. Recall that this results in sufficient statistics which are equivalent to matrix correlations; thus, it is useful to interpret model fit using the mean value parameterization. Parameter estimates for the correlation model are shown in Table 2, with $\sigma\left(\hat{\Theta}\right) \equiv \sqrt{\mathrm{Var}_{\hat{\Theta}} t\left(\mathbf{Y}, \mathbf{X}, \mathbf{a}\right)}$. Although the data transformation does not greatly alter the overall pattern of results, it does provide another avenue for interpretation. For instance, Table 2 shows that the expected matrix correlation between advice and reporting to be approximately, 0.229, with a 95% prediction interval of 0.184 to 0.251 under hypothetical replications. Friendship, by contrast, has a prediction interval which ranges from 0.044 to 0.284. Thus, while both ReportsTo and Friendship are significant effects, reporting would be expected to have a more *consistent* positive marginal relationship in data gathered under similar circumstances.

In interpreting mean value parameters, it is important to bear in mind that parameter values reflect overall, marginal relationships. For instance,

	Estimate			MCMC t Quantiles			
Effect	$\hat{\Theta}$	$\tau\left(\hat{\Theta}\right)$	$\sigma\left(\hat{\Theta}\right)$	2.5%	97.5%	$p(>\lvert Z\rvert)$	
Advice×ReportsTo	162.5	0.229	0.020	0.184	0.251	0.0027	**
Advice×Friendship	48.0	0.174	0.064	0.044	0.283	0.0088	**
Advice×TenureDiff	-67.9	-0.293	0.055	-0.390	-0.174	0.0039	**
Advice×LevelDiff	-32.6	0.195	0.074	0.018	0.308	0.0533	+
Advice×AgeDiff	24.5	-0.045	0.067	-0.185	0.082	0.1467	
Advice×SameDept	-29.9	-0.037	0.038	-0.110	0.046	0.2806	
Log-Likelihood: -26.24							
AIC: 64.49							

Table 2: Permutation Model Fit for Advice-Seeking Behavior, Matrix Correlation Parameterization

Table 2 shows a negative Θ effect for LevelDiff, even through the corresponding τ is 0.195! This apparent contradiction is resolved by noting that Θ here describes *relative tendency* towards (dis)alignment, while τ describes the absolute level obtained. Conditional on other parameters within the model, the matrix correlation between Advice and LevelDiff is lower than would be expected if its corresponding Θ were equal to 0; thus, the LevelDiff effect is negative, despite a positive marginal association. Such relationships illustrate the importance of interpreting mean value parameters together with their associated natural parameters, as well as the risk of attempting to employ raw matrix correlations in a multivariate environment without a model such as the one discussed here. Taken in the appropriate context, an apparently confusing observation of a reverse level effect is seen to be an artifact of the association between LevelDiff and other variables, with the true effect estimated to be in the expected direction.

5 Final Comments

We have demonstrated here an exponential family of permutation models, which is suitable for inferring the direction and strength of association among dyadic relational structures. While this family forms a distinct class of distributions, it is important to note its close connection with another class of discrete exponential families which have been proposed for the analysis of multivariate relational data (see, e.g., Wasserman and Pattison, 1996; Pattison and Wasserman, 1999; Robins et al., 1999). These multivariate exponential random graph (MERG) models (also called multivariate $p*$ models) have the general form

$$p\left(G\,|\theta\right) \propto \exp\left(\theta^T s\left(\theta\right)\right) \tag{25}$$

where $G = (G_1, \ldots, G_n)$ is a graph set on common vertex set V, s is a vector of general real-valued functions on the support of G, and the support of G is generally taken to be the set of all possible n-tuples of unvalued graphs on V. The univariate exponential random graph (ERG, also called $p*$) model as a special case when $n = 1$. Clearly, this family is a general form for densities on the set of graph n-tuples having order $|V|$, and allows for complex specification of both univariate an multivariate structural properties. Indeed, in the dichotomous case, Equation 25 can be made equivalent to Equation 3 by taking $G = (G_Y, G_X)$, choosing s equivalent to t, and by restricting the support of the model to the setwise isomorphism classes of the original data (G_Y, G_X). The statistic used here is equivalent to the association statistic of Pattison and Wasserman (1999), and can be motivated by the application of the Hammersley-Clifford theorem to the family of multivariate Markov graphs. Thus, for dichotomous data, the present model is a subclass of the (appropriately conditioned) MERG family. On the other hand, the standard MERG is limited to the discrete case (see Robins et al. (1999) for the extension of the dichotomous model), whereas this is not true of the permutation model. Indeed, by virtue of conditioning on the structure of the input matrices, the permutation model is nonparametric with respect to the distribution of the cell values. This is a very useful property in exploratory contexts, especially where mixed data types are involved. The permutation model and the MERG, then, have a non-empty intersection, but have distinct scope. Both are useful for particular problems, and it is not the intention of this paper to present either as an all-purpose solution to the problem of relational comparison. Jointly, likelihood-based tools such as these constitute a viable alternative to matrix correlation/regression approaches in a wide range of circumstances, providing a principled approach to inference on properties of graph sets.

6 References

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

Besag, J. (1975). Statistical Analysis of Non-Lattice Data. *The Statistician*, 24(3):179–195.

Besag, J. and Clifford, P. (1989). Generalized Monte Carlo Significance Tests. *Biometrika*, 76(4):633–642.

Crouch, B., Wasserman, S., and Trachtenburg, F. (1998). Markov Chain Monte Carlo Maximum Likelihood Estimation for $p*$ Social Network Models. Paper presented at the XVIII International Sunbelt Social Network Conference, Sitges, Spain.

DeGroot, M. H., Feder, P. I., and Goel, P. K. (1971). Matchmaking. *The Annals of Mathematical Statistics*, 42(1):578–593.

Dekker, D., Krackhardt, D., and Snijders, T. (2003). Mulicollinearity Robust QAP for Multiple Regression. CASOS working paper, Carnegie Mellon University, Pittsburgh, PA.

Gale, N., Hubert, L. J., Tobler, W. R., and Golledge, R. G. (1983). Combinatorial Procedures for the Analysis of Alternative Models: An Example from Interregional Migration. *Papers of the Regional Science Association*, 53:105–115.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7:457–511.

Geweke, J. (1992). Evaluating the Acuracy of Sampling-based Approaches to the Calculation of Posterior Moments (with Discussion). In Berger, J. O., Bernardo, J. M., Dawid, A. P., and Smith A. F. M., editors, *Bayesian Statistics*, volume 4, pages 164–193. Oxford University Press, Oxford.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo Maximum Likelihood Calculations (with Discussion). *Journal of the Royal Statistical Society, Series C*, 54:657–699.

Gibbons, D. and Olk, P. M. (2003). Individual and Structural Origins of Friendship and Social Position Among Professionals. *Journal of Personality and Social Psychology*, 84(2):340–351.

Goel, P. K. (1975). On Re-Pairing Observations in a Broken Random Sample. *The Annals of Statistics*, 3(6):1364–1369.

Henery, R. J. (1981). Permutation Probabilities as Models for Horse Races. *Journal of the Royal Statistical Society, Series B*, 43(1):86–91.

Holder, R. L. and Nevill, A. M. (1997). Modelling Performance at International Tennis and Golf Tournaments: Is There a Home Advantage? *The Statistician*, 46(4):551–559.

Hubert, L. J. (1987). *Assignment Methods in Combinatorial Data Analysis*. Marcel Dekker, New York.

Hunter, D. R. and Handcock, M. S. (2004). Inference in Curved Exponential Family Models for Networks. Technical Report TR0402, Penn State Department of Statistics.

Johansen, S. (1979). *Introduction to the Theory of Regular Exponential Families*. Institute of Mathematical Statistics, University of Copenhagen, Copenhagen.

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

Krackhardt, D. (1987b). QAP partialling as a test of spuriousness. *Social Networks*, 9:171–186.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analyses of dyadic data. *Social Networks*, 10:359–382.

Luce, R. D. (1959). *Individual Choice Behavior*. John Wiley, New York.

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27:209–220.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37:17–23.

Nakao, K. and Romney, A. K. (1984). A Method for Testing Alternative Theories: An Example from English Kinship. *American Anthropologist*, 86:668–673.

Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52:169–193.

Plackett, R. L. (1975). The Analysis of Permutations. *Applied Statistics*, 24(2):193–202.

Robins, G., Pattison, P., and Wasserman, S. (1999). Logit models and logistic regressions for social networks, III. valued relations. *Psychometrika*, 64:371–394.

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

Sokal, R. R. (1979). Testing Statistical Significance of Geographic Variation Patterns. *Systematic Zoology*, 28:227–232.

Stern, H. (1990). Models for Distributions on Permutations. *Journal of the American Statistical Association*, 85(410):558–564.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and $p*$. *Psychometrika*, 60:401–426.

Yeung, K.-T. and Martin, J. L. (2003). The Looking Glass Self: An Empirical Test and Elaboration. *Social Forces*, 81(3):843–879.