Bilinear Mixed-Effects Models for Dyadic Data

Peter D. HOFF

This article discusses the use of a symmetric multiplicative interaction effect to capture certain types of third-order dependence patterns often present in social networks and other dyadic datasets. Such an effect, along with standard linear fixed and random effects, is incorporated into a generalized linear model, and a Markov chain Monte Carlo algorithm is provided for Bayesian estimation and inference. In an example analysis of international relations data, accounting for such patterns improves model fit and predictive performance.

KEY WORDS: Balance; Generalized linear model; Inner product scaling; Social network.

1. INTRODUCTION

Dyadic data consist of measurements made on pairs of objects or under pairs of conditions, so that $y_{i,j}$ denotes the value of the measurement on the potentially ordered pair (i,j). Examples of dyadic data include social networks, "round-robin" experiments in psychology, genetic intercross data, and comparative data in which $y_{i,j}$ might be a measure of similarity between units i and j.

In the social networks literature, modeling has focused on the binary case where $y_{i,j}$ is either 0 or 1, indicating the presence or absence of a "link" from i to j. This has led to the development of data analysis tools based on directed graphs and the study of exponentially parameterized random graph models (Wasserman and Pattison 1996). For valued (nonbinary) dyadic datasets, a perceived lack of statistical tools has sometimes led to ad hoc reductions of valued responses to binary data. However, ANOVA methods are available for valued dyadic data; the so-called "social relations" model (Warner, Kenny, and Stoto 1979; Wong 1982; Snijders and Kenny 1999) allows for the decomposition of the variance into sender- and receiver-specific effects, and also allows for correlation of responses within a dyad. Such a model has been studied in the context of a linear group symmetry model by Li (2002), and advances in variance component analysis have been made by Gill and Swartz (2001) and Li and Loken (2002). These models generally presume normally distributed data and additive effects, and thus also presume the absence of any sort of dependence beyond those specified by second-order moments. In contrast, many observed dyadic datasets exhibit certain forms of third-order dependence, and it often is of scientific interest to quantify these higher-order patterns.

In this article we propose a class of generalized additive models based on the social relations model, but incorporate third-order dependence via a bilinear effect. The bilinear effect for a pair (i,j) is simply the inner product of unobserved characteristic vectors $\mathbf{z}_i$ and $\mathbf{z}_j$, specific to units i and j. This approach is similar in spirit to the latent variable and latent distance methods proposed by Hoff, Raftery, and Handcock (2002) to capture transitivity in a social network dataset, but has some computational and conceptual advantages. The bilinear effect is also a type of multiplicative interaction (Gabriel 1978; Marasinghe and Johnson 1982; Oman 1991). The models presented in this article are similar to the generalized bilinear regression models studied by Gabriel (1998), who considered approximate maximum likelihood estimation in the context of factorial designs. In this article we show how a bilinear effect can be used to represent certain forms of dependence often seen in dyadic data and develop a Markov chain Monte Carlo (MCMC) algorithm based on Gibbs sampling, providing arbitrarily exact Bayesian inference. With some modifications, the algorithm can be used as a means of making Bayesian inference for a broad class of generalized bilinear regression models with mixed effects.

In the next section we discuss the basic linear mixed-effects model for dyadic data and the resulting dependence structure. In Section 3 we discuss types of third-order dependence often seen in network datasets and the use of a bilinear effect to capture such dependence. In Section 4 we give an MCMC algorithm that can be used to obtain samples from the posterior distribution of the parameters. We discuss model selection strategies in Section 5, and address practical data analysis and model interpretation issues in the context of an example analysis of international relations data in Section 6. We conclude with a discussion in Section 7.

2. LINEAR MIXED–EFFECTS MODELS FOR EXCHANGEABLE DYADIC DATA

Suppose that we are interested only in estimating the linear relationships between responses $y_{i,j}$ and a possibly vector-valued set of variables $\mathbf{x}_{i,j}$, which could include characteristics of unit i, characteristics of unit j, or characteristics specific to the pair. In this case we might consider the regression model

$$y_{i,j} = \boldsymbol{\beta}'\mathbf{x}_{i,j} + \epsilon_{i,j}, \qquad (1)$$

where $y_{i,i}$ is typically not defined. The generalized least squares estimate $\hat{\boldsymbol{\beta}}$ and its covariance matrix depend on the joint distribution of the $\epsilon_{i,j}$'s only through their covariance. It is often assumed in regression problems that the regressors $\mathbf{x}_{i,j}$ contain enough information so that the distribution of the errors is invariant under permutations of the unit labels. This assumption is equivalent to the $n \times n$ matrix of errors (with an undefined diagonal) having a distribution that is invariant under identical row and column permutations, so that $\{\epsilon_{i,j} : i \neq j\}$ is equal in distribution to $\{\epsilon_{\pi(i),\pi(j)} : i \neq j\}$ for any permutation π of $\{1, \dots, n\}$. This condition is called "weak row-and-column exchangeability" of an array. For undirected data, in which $y_{i,j} = y_{j,i}$ by design, such exchangeability implies a "random-effects" representation of the errors, in that $\epsilon_{i,j}$ is equal in distribution to $f(\mu, a_i, a_j, \gamma_{i,j})$, where $\mu, a_i, a_j, \gamma_{i,j}$ are independent random variables and f is a function to be specified (Aldous

Peter D. Hoff is Assistant Professor of Statistics, Departments of Statistics and Biostatistics and the Center for Statistics in Social Sciences, University of Washington, Seattle, WA 98195 (E-mail: *hoff@stat.washington.edu*). This research was supported by Office of Naval Research grant N00014-02-1-1011. The author thanks Mark Handcock and Michael Ward for helpful discussions, and an associate editor for comments that substantially improved this document.

© 2005 American Statistical Association
Journal of the American Statistical Association
March 2005, Vol. 100, No. 469, Theory and Methods
DOI 10.1198/016214504000001015

1985, thm. 14.11). If in addition to the foregoing invariance assumption we also model the errors as Gaussian with mean 0, then the joint distribution can be represented in terms of a linear random-effects model $\epsilon_{i,j} = a_i + a_j + \gamma_{i,j}$. In the more general case of directed observations, where $y_{i,j}$ and $\epsilon_{i,j}$ are potentially distinct from $y_{j,i}$ and $\epsilon_{j,i}$, we can represent the joint distribution of the $\epsilon_{i,j}$'s as

$$\epsilon_{i,j} = a_i + b_j + \gamma_{i,j}, \tag{2}$$

$$(a_i, b_i)' \sim \text{ multivariate normal [MVN]}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{ab}}),$$

$$\boldsymbol{\Sigma}_{\mathbf{ab}} = \begin{pmatrix} \sigma_{\mathbf{a}}^2 & \sigma_{\mathbf{ab}} \\ \sigma_{\mathbf{ab}} & \sigma_{\mathbf{b}}^2 \end{pmatrix},$$

$$(\gamma_{i,j}, \gamma_{j,i})' \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \begin{pmatrix} \sigma_{\boldsymbol{\gamma}}^2 & \rho\sigma_{\boldsymbol{\gamma}}^2 \\ \rho\sigma_{\boldsymbol{\gamma}}^2 & \sigma_{\boldsymbol{\gamma}}^2 \end{pmatrix},$$

with effects otherwise being independent. The covariance structure of the errors (and thus the observations) is

$$\begin{aligned} E(\epsilon_{i,j}^2) &= \sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{b}}^2 + \sigma_{\boldsymbol{\gamma}}^2, & E(\epsilon_{i,j}\epsilon_{i,k}) &= \sigma_{\mathbf{a}}^2, \\ E(\epsilon_{i,j}\epsilon_{j,i}) &= \rho\sigma_{\boldsymbol{\gamma}}^2 + 2\sigma_{\mathbf{ab}}, & E(\epsilon_{i,j}\epsilon_{k,j}) &= \sigma_{\mathbf{b}}^2, \\ E(\epsilon_{i,j}\epsilon_{k,l}) &= 0, & E(\epsilon_{i,j}\epsilon_{k,i}) &= \sigma_{\mathbf{ab}}, \end{aligned}$$

and so $\sigma_{\mathbf{a}}^2$ represents the dependence of observations having a common sender, $\sigma_{\mathbf{b}}^2$ represents the dependence of observations having a common receiver, and ρ represents the correlation of observations within a dyad (often interpreted as "mutuality" or "reciprocity"). This has been called the "social relations" or "round-robin" model (Warner et al. 1979; Wong 1982) and is related to a model for diallel cross-data used by Cockerham and Weir (1977). The model is a special case of a linear group symmetry model (Andersson and Madsen 1998), and has been studied in this context by Li (2002). Recent advances in variance component estimation have been made by Gill and Swartz (2001) and Li and Loken (2002).

To analyze responses in particular sample spaces, such as binary or count data, the error structure described in (2) can be added to a linear predictor in a generalized linear model,

$$\theta_{i,j} = \boldsymbol{\beta}'\mathbf{x}_{i,j} + a_i + b_j + \gamma_{i,j}, \tag{3}$$

$$E(y_{i,j}|\theta_{i,j}) = g(\theta_{i,j}),$$

$$p(y_{1,2}, y_{1,3}, \ldots, y_{n,n-1}|\theta_{1,2}, \theta_{1,3}, \ldots, \theta_{n,n-1}), = \prod_{i \neq j} p(y_{i,j}|\theta_{i,j}).$$

This is a generalized linear mixed-effects model with inverse-link function $g(\boldsymbol{\theta})$. For example, binary data can be modeled by letting $p(y_{i,j} = 1|\theta_{i,j}) = e^{\theta_{i,j}}/(1 + e^{\theta_{i,j}})$, and count data can be accommodated by letting $p(y_{i,j}|\theta_{i,j})$ be the Poisson distribution with mean $\exp\{\theta_{i,j}\}$. In this random-effects setting, the observations are modeled as conditionally independent given the random effects but are unconditionally dependent. The covariance pattern for the observations is given approximately as

$$\begin{aligned} &\text{cov}(y_{i_1,j_1}, y_{i_2,j_2}) \\ &\quad = E[\text{cov}(y_{i_1,j_1}, y_{i_2,j_2}|\theta_{i_1,j_1}, \theta_{i_2,j_2})] \\ &\qquad + \text{cov}[E(y_{i_1,j_1}|\theta_{i_1,j_1}), E(y_{i_2,j_2}|\theta_{i_2,j_2})] \\ &\quad = E[0] + \text{cov}[g(\theta_{i_1,j_1}), g(\theta_{i_2,j_2})] \\ &\quad \approx \text{cov}(\theta_{i_1,j_1}, \theta_{i_2,j_2}) \times g'(\boldsymbol{\beta}'\mathbf{x}_{i_1,j_1})g'(\boldsymbol{\beta}'\mathbf{x}_{i_2,j_2}), \end{aligned}$$

where the pattern for $\text{cov}(\theta_{i_1,j_1}, \theta_{i_2,j_2})$ is the same as that for the $\epsilon_{i,j}$'s given earlier. However, unlike the linear regression case, $\hat{\boldsymbol{\beta}}$ is not given by linear combinations of the observations, and $E(\hat{\boldsymbol{\beta}})$ and $\text{cov}(\hat{\boldsymbol{\beta}})$ are not functions of only the first- and second-order moments of the data. Our inference on $\boldsymbol{\beta}$ will be affected by model lack of fit and third-order and higher-order dependence. Indeed, it is these higher-order patterns of dependence that are often of interest and may also provide information useful for predictive inference.

3. MODELING THIRD–ORDER DEPENDENCE PATTERNS

Some dependence patterns commonly seen in dyadic datasets have been given the descriptive titles of transitivity, balance, and clusterability. In the context of binary data, graph-theoretic definitions of these concepts have been given by Wasserman and Faust (1994, chap. 6). We modify these definitions to conceptualize third-order dependence in a regression setting. Suppose that one has fit a linear regression model with row and column effects to a dyadic dataset and has obtained the residuals $\{\hat{\xi}_{i,j} : i \neq j\}$. A set of three residuals among a triad of units i, j, k is called a *cycle* if there is one residual per pair of units. For example, $\{\hat{\xi}_{i,j}, \hat{\xi}_{j,k}, \hat{\xi}_{k,i}\}$ is one such set, as is $\{\hat{\xi}_{i,j}, \hat{\xi}_{i,k}, \hat{\xi}_{j,k}\}$. Such a cycle is called *balanced* if the product of the three residuals is positive; that is, there are zero or two negative residuals in the set. A cycle is called *clusterable* if it is balanced or if all three of the residuals are negative. Balanced and clusterable cycles of residuals are shown graphically in Figure 1. We note that for directed data, a given triad i, j, k has eight such cycles of residuals. The model that we propose does not distinguish among these, although a model extension discussed in Section 7 does.

For general signed relations among units, many theories of social systems suggest that the relationships within triads tend to be balanced or clusterable (see Heider 1979 for a review). The idea is that if the relationship between i and j is "positive," then i and j will relate to another unit k in a manner similar to one another. In a regression setting, this suggests that if $\hat{\xi}_{i,j} > 0$, then $\hat{\xi}_{j,k}$ and $\hat{\xi}_{i,k}$ are either both positive or both negative. If all triads among a population of units have balanced cycles, then the population can be divided into two groups, with $\hat{\xi}_{i,j} > 0$ if i and j are in the same group and $\hat{\xi}_{i,j} < 0$ if they are in opposite groups (Harary 1953). Clusterability is a relaxation of the concept of balance. If all the triads of a system are clusterable, then the units can be divided into two or more groups with positive residuals between units in the same group and negative residuals between units in different groups (Davis 1976).

In practice, a dataset will display varying degrees of balance or clusterability. Often it is found that there are more balanced or clusterable residual cycles than would be expected under models (2) or (3). Balance would also be indicated if the average value of $\hat{\xi}_{i,j} \times \hat{\xi}_{j,k} \times \hat{\xi}_{k,i}$ is substantially larger than 0, the expected value presumed by model (2). Schweinberger and Snijders (2003) used latent ultrametric distances to represent transitive structures in dyadic data. Hoff et al. (2002) used simple functions of latent characteristic vectors $\mathbf{z}_1, \ldots, \mathbf{z}_n \subset \mathbb{R}^K$

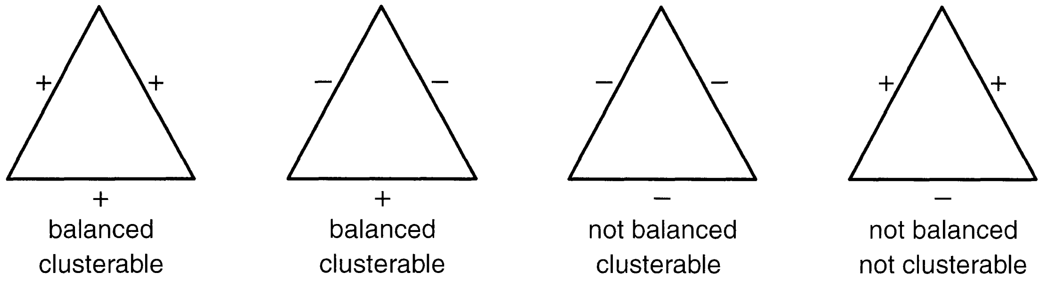

Figure 1. Balance and Clusterability of Cycles. In a regression setting, the "+" or "−" denotes the sign of the residual between two nodes, with the nodes represented by corners of the triangles. Cycles having one of the first two patterns from the left are balanced; cycles having one of the first three patterns from the left are clusterable.

in a fixed-effects setting to capture some forms of transitivity, balance, and clusterability. For example, they considered models in which $\theta_{i,j} = \boldsymbol{\beta}'\mathbf{x}_{i,j} + f(\mathbf{z}_i, \mathbf{z}_j)$, where $f(\mathbf{z}_i, \mathbf{z}_j) = -|\mathbf{z}_i - \mathbf{z}_j|$ ("the distance model") or $f(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{z}_i'\mathbf{z}_j/|\mathbf{z}_j|$ ("the projection model"). In what follows, we consider a similar approach using the inner-product kernel $f(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{z}_i'\mathbf{z}_j$, and give random-effects and fixed-effects interpretations.

Adding the bilinear effect $\mathbf{z}_i'\mathbf{z}_j$ to the linear random effects in models (2) and (3) gives

$$\begin{aligned} \epsilon_{i,j} &= a_i + b_j + \gamma_{i,j} + \xi_{i,j}, \qquad (4) \\ \xi_{i,j} &= \mathbf{z}_i'\mathbf{z}_j, \end{aligned}$$

where the random effects a_i, b_j, and $\gamma_{i,j}$ are modeled with the multivariate normal distributions described earlier. Note that the set of effects $\{\xi_{i,j} : i \neq j\}$ are able to represent balance and clusterability as we defined: If the dimension K of the $\mathbf{z}$'s is one, then $\xi_{i,j} \times \xi_{j,k} \times \xi_{k,j} = (\mathbf{z}_i \times \mathbf{z}_j \times \mathbf{z}_k)^2 > 0$. In this case, all cycles of the ξ's among triads are balanced, allowing for a clustering of the units into two groups with positive ξ's within groups and negative ξ's between groups. The number of possible such groupings increases as the dimension of the $\mathbf{z}$'s increases; in K dimensions, we can find $2K$ vectors having pairwise inner products that are less than or equal to 0 or $K + 1$ vectors having pairwise inner products that are strictly negative. In practice, we would not expect populations to be completely balanced or clusterable, and by having distinct $\mathbf{z}$'s among the n units we can allow for various proportions of cycles in each of the four categories shown in Figure 1.

We have also written $\xi_{i,j} = \mathbf{z}_i'\mathbf{z}_j$ to suggest the interpretation of $\mathbf{z}_i'\mathbf{z}_j$ as a mean-0 random effect. If the $\mathbf{z}$'s are modeled as independent K-dimensional multivariate normal random vectors with mean 0 and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{z}}$, then the resulting distribution for the ξ's has the following moment properties:

- $E(\xi_{i,j}) = 0$
- $E(\xi_{i,j}^2) = \operatorname{trace} \boldsymbol{\Sigma}_{\mathbf{z}}^2$
- $E(\xi_{i,j}\xi_{j,k}\xi_{k,i}) = \operatorname{trace} \boldsymbol{\Sigma}_{\mathbf{z}}^3$,

with all other second- and third-order moments equal to 0. Note that an orthogonal transformation of the $\mathbf{z}$'s leaves $\mathbf{z}_i'\mathbf{z}_i$ invariant, so we can assume that $\boldsymbol{\Sigma}_{\mathbf{z}}$ is a diagonal matrix (otherwise, the off-diagonal terms are nonidentifiable). For simplicity, we focus on the case where $\boldsymbol{\Sigma}_{\mathbf{z}} = \sigma_{\mathbf{z}}^2 \mathbf{I}_{K \times K}$, for which the foregoing moments are $0, K\sigma_{\mathbf{z}}^4$, and $K\sigma_{\mathbf{z}}^6$. With $\xi_{i,j}$ added to the error term, the nonzero second- and third-order moments are

$$\begin{aligned} E(\epsilon_{i,j}^2) &= \sigma_{\mathbf{a}}^2 + \sigma_{\mathbf{b}}^2 + \sigma_{\gamma}^2 + K\sigma_{\mathbf{z}}^4, \\ E(\epsilon_{i,j}\epsilon_{i,k}) &= \sigma_{\mathbf{a}}^2, \\ E(\epsilon_{i,j}\epsilon_{j,i}) &= \rho\sigma_{\gamma}^2 + 2\sigma_{\mathbf{ab}} + K\sigma_{\mathbf{z}}^4, \\ E(\epsilon_{i,j}\epsilon_{k,j}) &= \sigma_{\mathbf{b}}^2, \\ E(\epsilon_{i,j}\epsilon_{k,i}) &= \sigma_{\mathbf{ab}}, \\ E(\epsilon_{i,j}\epsilon_{j,k}\epsilon_{k,i}) &= K\sigma_{\mathbf{z}}^6. \end{aligned}$$

Thus the multiplicative term $\xi_{i,j} = \mathbf{z}_i'\mathbf{z}_j$ can be interpreted as a mean-0 random effect able to induce particular forms of third-order dependence often found in dyadic datasets.

4. PARAMETER ESTIMATION

In the maximum likelihood setting, approximate estimation for generalized linear mixed-effects models often proceeds via Taylor expansions and iteratively reweighted least squares for the fixed effects, along with approximate restricted maximum likelihood estimation for the variance components (Schall 1991; Breslow and Clayton 1993; Wolfinger and O'Connell 1993; McGilchrist 1994). The accuracy of these approximate methods is generally dependent on the sample size (see Booth and Hobert 1998 for a discussion). Gabriel (1998) suggested an algorithm along these lines for the generalized bilinear regression model. Alternatively, Zeger and Karim (1991), Gelfand, Sahu, and Carlin (1996), and Natarajan and Kass (2000) have proposed Gibbs sampling approaches to parameter estimation for generalized linear mixed-effects models. However, estimation is more difficult for the complicated dependence structure of the random effects in the invariant normal model (2). Gill and Swartz (2001) have proposed a Gibbs sampling scheme for estimation of random effects in the linear case with the identity link, although we have found that their algorithm does not mix well when covariates are included, due to a weak identifiability of the unit-level random effects and certain regression coefficients. As discussed by Gelfand, Sahu, and Carlin (1995), the random effects $\mathbf{a}$ and $\mathbf{b}$ will be confounded to a degree with each other and to regression parameters associated with predictors that do not vary across receivers (i.e., sender-specific effects) or across senders (receiver-specific effects). For example, a population-level intercept is one such parameter. To obtain a "cleaner" partition of the variance and a more efficient MCMC sampling scheme, we decompose $\mathbf{x}_{i,j}$ into $\mathbf{x}_{i,j} = (\mathbf{x}_{d,i,j}, \mathbf{x}_{\mathbf{s},i}, \mathbf{x}_{\mathbf{r},j})$, that is, into dyad-specific regressors $\mathbf{x}_{d,i,j}$, sender-specific regressors $\mathbf{x}_{\mathbf{s},i}$ and receiver-specific regressors $\mathbf{x}_{\mathbf{r},j}$. We then rewrite the generalized bilinear model as

$$\theta_{i,j} = \boldsymbol{\beta}'_d \mathbf{x}_{d,i,j} + (\boldsymbol{\beta}'_{\mathbf{s}} \mathbf{x}_{\mathbf{s},i} + a_i) + (\boldsymbol{\beta}'_{\mathbf{r}} \mathbf{x}_{\mathbf{r},j} + b_j) + \gamma_{i,j} + \mathbf{z}'_i \mathbf{z}_j$$

or, equivalently,

$$\begin{aligned} \theta_{i,j} &= \boldsymbol{\beta}'_d \mathbf{x}_{d,i,j} + s_i + r_j + \gamma_{i,j} + \mathbf{z}'_i \mathbf{z}_j, \\ s_i &= \boldsymbol{\beta}'_{\mathbf{s}} \mathbf{x}_{\mathbf{s},i} + a_i, \qquad \text{and} \\ r_i &= \boldsymbol{\beta}'_{\mathbf{r}} \mathbf{x}_{\mathbf{r},i} + b_i. \end{aligned}$$

Note that an intercept can be thought of as either a sender-specific or receiver-specific effect. For symmetry, we include the constant 1/2 at the beginning of each $\mathbf{x}_{\mathbf{s},i}$ and $\mathbf{x}_{\mathbf{r},j}$ vector, and estimate the first components of $\boldsymbol{\beta}_{\mathbf{s}}$ and $\boldsymbol{\beta}_{\mathbf{r}}$ as being equal. This parameterization for the linear unit-level effects is similar to the "centered" parameterizations suggested by Gelfand et al. (1995, 1996). The relative gains in MCMC efficiency, along with other efficient parameterizations, have been discussed by Papaspiliopoulos, Roberts, and Sköld (2003).

Using the foregoing reparameterization for $\theta_{i,j}$, we estimate the parameters for the generalized bilinear regression model by constructing a Markov chain in $\{\boldsymbol{\beta}_d, \boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{ab}}, \mathbf{Z}, \sigma^2_{\mathbf{z}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\}$, where $\mathbf{Z}$ denotes the $K \times n$ matrix of latent vectors, having $p(\boldsymbol{\beta}_d, \boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{ab}}, \mathbf{Z}, \sigma^2_{\mathbf{z}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} | \mathbf{Y})$ as the invariant distribution. We obtain this via an algorithm based on Gibbs sampling, which also samples $\mathbf{s}$, $\mathbf{r}$, and the $\boldsymbol{\theta}$'s. The basic algorithm involves iterating the following steps:

1. Sample linear effects:
   a. Sample $\boldsymbol{\beta}_d, \mathbf{s}, \mathbf{r} | \boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{ab}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}, \boldsymbol{\theta}$, and $\mathbf{Z}$ (linear regression).
   b. Sample $\boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}} | \mathbf{s}, \mathbf{r}$, and $\boldsymbol{\Sigma}_{\mathbf{ab}}$ (linear regression).
   c. Sample $\boldsymbol{\Sigma}_{\mathbf{ab}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ from their full conditionals.
2. Sample bilinear effects:
   a. For $i = 1, \dots, n$, sample $\mathbf{z}_i | \{\mathbf{z}_j, j \neq i\}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{s}, \mathbf{r}, \boldsymbol{\Sigma}_{\mathbf{z}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ (linear regression).
   b. Sample $\boldsymbol{\Sigma}_{\mathbf{z}}$ from its full conditional.
3. Sample dyad-specific parameters. Update $\{\theta_{i,j}, \theta_{j,i}\}$ using a Metropolis–Hastings step:
   a. Propose $\binom{\theta^*_{i,j}}{\theta^*_{j,i}} \sim \mathrm{MVN}\left(\binom{\boldsymbol{\beta}'\mathbf{x}_{i,j} + a_i + b_j + \mathbf{z}'_i\mathbf{z}_j}{\boldsymbol{\beta}'\mathbf{x}_{j,i} + a_j + b_i + \mathbf{z}'_j\mathbf{z}_i}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\right)$.
   b. Accept $\binom{\theta^*_{i,j}}{\theta^*_{j,i}}$ with probability $\frac{p(y_{i,j}|\theta^*_{i,j}) p(y_{j,i}|\theta^*_{j,i})}{p(y_{i,j}|\theta_{i,j}) p(y_{j,i}|\theta_{j,i})} \wedge 1$.

Various combinations of the foregoing steps can be used to estimate different models. The steps in 1 alone provide a Bayesian estimation procedure for the linear regression problem with an error covariance as in (2). Bayesian estimation of the normal bilinear model with the identity link could proceed by replacing each $\theta_{i,j}$ with $y_{i,j}$ and iterating only steps 1 and 2. For nonnormal data, estimation of a generalized bilinear mixed-effects model proceeds by iterating steps 1, 2, and 3. The full conditional distributions required to perform steps 1 and 2 are given later.

Note that the $\boldsymbol{\theta}$'s are essentially unrestricted in the sampling scheme. At this level, the fit is saturated and does not depend on the regressors, at least to the degree that the prior for $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is diffuse. What the MCMC algorithm provides is essentially a saturated fit for the $\boldsymbol{\theta}$'s (although somewhat smoothed by the common variance) and an ANOVA-like decomposition of the $\boldsymbol{\theta}$'s into regressor, sender, receiver, and inner-product effects.

4.1 Conditional Distributions for the Linear-Effects Components

Noting that $\theta_{i,j} - \mathbf{z}'_i\mathbf{z}_j = \boldsymbol{\beta}'_d \mathbf{x}_{i,j} + s_i + r_j + \gamma_{i,j}$, we see that conditional on the $\boldsymbol{\theta}$'s and $\mathbf{z}$'s, the other parameters can be sampled using a standard Bayesian normal-theory regression approach, although with a complicated covariance structure.

Full Conditional of $(\boldsymbol{\beta}_d, \mathbf{s}, \mathbf{r})$. Similar to Wong's (1982) approach to the invariant normal model, we let $u_{i,j} = \theta_{i,j} + \theta_{j,i} - 2\mathbf{z}'_i\mathbf{z}_j$ and $v_{i,j} = \theta_{i,j} - \theta_{j,i}$ for $i < j$. We then have

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{\mathbf{u}} \\ \mathbf{X}_{\mathbf{v}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_d \\ \mathbf{s} \\ \mathbf{r} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{u}} \\ \boldsymbol{\delta}_{\mathbf{v}} \end{pmatrix}, \qquad (5)$$

where $\mathbf{X}_{\mathbf{u}}$ and $\mathbf{X}_{\mathbf{v}}$ are the appropriate design matrices and $\boldsymbol{\delta}_{\mathbf{u}}$ and $\boldsymbol{\delta}_{\mathbf{v}}$ are vectors of independent error terms with variances $\sigma^2_{\mathbf{u}} = 2\sigma^2_{\boldsymbol{\gamma}}(1 + \rho)$ and $\sigma^2_{\mathbf{v}} = 2\sigma^2_{\boldsymbol{\gamma}}(1 - \rho)$. The full conditional distribution of $(\boldsymbol{\beta}_d, \mathbf{s}, \mathbf{r})$ is then proportional to $p(\mathbf{u}, \mathbf{v} | \boldsymbol{\beta}_d, \mathbf{s}, \mathbf{r}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) \times p(\mathbf{s}, \mathbf{r} | \boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{ab}}) \times p(\boldsymbol{\beta}_d)$. For a multivariate normal $(\boldsymbol{\mu}_{\boldsymbol{\beta}_d}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_d})$ prior distribution on $\boldsymbol{\beta}_d$, the term in the exponent of the full conditional is

$$\begin{aligned} &\boldsymbol{\phi}' \left[\begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_d} \boldsymbol{\mu}_{\boldsymbol{\beta}_d} \\ \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} \mathbf{X}_{\mathbf{sr}} \boldsymbol{\beta}_{\mathbf{sr}} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{u}/\sigma^2_{\mathbf{u}} + \mathbf{X}'_{\mathbf{v}}\mathbf{v}/\sigma^2_{\mathbf{v}} \right] \\ &\qquad - \frac{1}{2}\boldsymbol{\phi}' \left[\begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_d} & 0 \\ 0 & \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{X}_{\mathbf{u}}/\sigma^2_{\mathbf{u}} + \mathbf{X}'_{\mathbf{v}}\mathbf{X}_{\mathbf{v}}/\sigma^2_{\mathbf{v}} \right] \boldsymbol{\phi}, \end{aligned}$$

where $\boldsymbol{\phi}' = (\boldsymbol{\beta}'_d \mathbf{s}' \mathbf{r}')$, $\mathbf{X}_{\mathbf{sr}}$ and $\boldsymbol{\beta}_{\mathbf{sr}}$ are the combined design matrix and regression parameters for $\mathbf{s}$ and $\mathbf{r}$, and $\boldsymbol{\Sigma}_{\mathbf{sr}}$ is the covariance matrix of $(\mathbf{s}'\mathbf{r}')'$, which is easily derived from $\boldsymbol{\Sigma}_{\mathbf{ab}}$. The conditional distribution is thus multivariate normal $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \left[\begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_d} \boldsymbol{\beta}_{d0} \\ \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} \mathbf{X}_{\mathbf{sr}} \boldsymbol{\beta}_{\mathbf{sr}} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{u}/\sigma^2_{\mathbf{u}} + \mathbf{X}'_{\mathbf{v}}\mathbf{v}/\sigma^2_{\mathbf{v}} \right]$$

and

$$\boldsymbol{\Sigma} = \left[\begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_d} & 0 \\ 0 & \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} \end{pmatrix} + \mathbf{X}'_{\mathbf{u}}\mathbf{X}_{\mathbf{u}}/\sigma^2_{\mathbf{u}} + \mathbf{X}'_{\mathbf{v}}\mathbf{X}_{\mathbf{v}}/\sigma^2_{\mathbf{v}} \right]^{-1}.$$

Note that the inverse of $\boldsymbol{\Sigma}_{\mathbf{sr}}$ is given by

$$\begin{aligned} \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} &= \begin{pmatrix} (\sigma^2_{\mathbf{b}}/\Delta)\mathbf{I}_{n \times n} & -(\sigma_{\mathbf{ab}}/\Delta)\mathbf{I}_{n \times n} \\ -(\sigma_{\mathbf{ab}}/\Delta)\mathbf{I}_{n \times n} & (\sigma^2_{\mathbf{a}}/\Delta)\mathbf{I}_{n \times n} \end{pmatrix}, \\ \Delta &= \sigma^2_{\mathbf{a}}\sigma^2_{\mathbf{b}} - \sigma^2_{\mathbf{ab}}. \end{aligned}$$

Full Conditional of $(\boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}})$. The full conditional of $(\boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}})$ is proportional to $p(\mathbf{s}, \mathbf{r} | \boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{ab}}) \times p(\boldsymbol{\beta}_{\mathbf{s}}, \boldsymbol{\beta}_{\mathbf{r}})$. Assuming a multivariate normal $(\boldsymbol{\mu}_{\boldsymbol{\beta}_{\mathbf{sr}}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\mathbf{sr}}})$ prior distribution for the combined regression parameters, the full conditional is a multivariate normal distribution with mean and variance $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given by

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \left[\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_{\mathbf{sr}}} \boldsymbol{\mu}_{\boldsymbol{\beta}_{\mathbf{sr}}} + \mathbf{X}_{\mathbf{sr}} \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} \begin{pmatrix} \mathbf{s} \\ \mathbf{r} \end{pmatrix} \right]$$

and

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_{\mathbf{sr}}} + \mathbf{X}'_{\mathbf{sr}} \boldsymbol{\Sigma}^{-1}_{\mathbf{sr}} \mathbf{X}_{\mathbf{sr}} \right)^{-1}.$$

Full Conditional of $\boldsymbol{\Sigma}_{\mathbf{ab}}$. The full conditional of $\boldsymbol{\Sigma}_{\mathbf{ab}}$ is proportional to $p(\mathbf{s},\mathbf{r}|\boldsymbol{\beta}_{\mathbf{s}},\boldsymbol{\beta}_{\mathbf{r}},\boldsymbol{\Sigma}_{\mathbf{ab}})p(\boldsymbol{\Sigma}_{\mathbf{ab}})$. Using a prior distribution of $\boldsymbol{\Sigma}_{\mathbf{ab}} \sim$ inverse Wishart$(\boldsymbol{\Sigma}_{\mathbf{ab}0},\nu)$ [parameterized so that $E(\boldsymbol{\Sigma}_{\mathbf{ab}}) = \boldsymbol{\Sigma}_{\mathbf{ab}0}/(\nu-3)$], the full conditional of $\boldsymbol{\Sigma}_{\mathbf{ab}}$ is $\boldsymbol{\Sigma}_{\mathbf{ab}}|\mathbf{a},\mathbf{b} \sim$ inverse Wishart$(\boldsymbol{\Sigma}_{\mathbf{ab}0} + (\mathbf{ab})'(\mathbf{ab}), \nu+n)$, where $\mathbf{a} = (\mathbf{s} - \mathbf{X}_{\mathbf{s}}\boldsymbol{\beta}_{\mathbf{s}})$ and $\mathbf{b} = (\mathbf{r} - \mathbf{X}_{\mathbf{r}}\boldsymbol{\beta}_{\mathbf{r}})$.

Full Conditional of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Using prior distributions of $\sigma^2_{\mathbf{u}} \sim$ inverse gamma$(\alpha_{\mathbf{u}1},\alpha_{\mathbf{u}2})$ and $\sigma^2_{\mathbf{v}} \sim$ inverse gamma$(\alpha_{\mathbf{v}1},\alpha_{\mathbf{v}2})$, the full conditionals are given by $\sigma^2_{\mathbf{u}}|\mathbf{u} \sim$ inverse gamma$(\alpha_{\mathbf{u}1} + \frac{1}{2}\binom{n}{2}, \alpha_{\mathbf{u}2} + \frac{1}{2}\sum[u_i - \hat{u}_{i,j}]^2)$ and $\sigma^2_{\mathbf{v}}|\mathbf{v} \sim$ inverse gamma$(\alpha_{\mathbf{v}1} + \frac{1}{2}\binom{n}{2}, \alpha_{\mathbf{v}2} + \frac{1}{2}\sum[v_i - \hat{v}_{i,j}]^2)$, where $\hat{u}_{i,j} = E[u_{i,j}|\boldsymbol{\beta}_d,\mathbf{x}_{i,j},s_i,r_j] = \boldsymbol{\beta}'_d(\mathbf{x}_{i,j} + \mathbf{x}_{j,i}) + s_i + s_j + r_i + r_j$, and $\hat{v}_{i,j}$ is given similarly. The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ can be reconstructed from $\sigma^2_{\mathbf{u}}$ and $\sigma^2_{\mathbf{v}}$ via $\sigma^2_{\gamma} = (\sigma^2_{\mathbf{u}} + \sigma^2_{\mathbf{v}})/4$ and $\rho = (\sigma^2_{\mathbf{u}} - \sigma^2_{\mathbf{v}})/(\sigma^2_{\mathbf{u}} + \sigma^2_{\mathbf{v}})$.

4.2 Conditional Distributions for the Bilinear Effects Component

Let $e_{i,j} = (\theta_{i,j} + \theta_{j,i} - \hat{u}_{i,j})/2$, the residual of the symmetric part of the matrix of $\boldsymbol{\theta}$'s after fitting the linear effects, and let $\delta_{\mathbf{u},i,j} = \gamma_{i,j} + \gamma_{j,i}$. Considering the full conditional of $\mathbf{z}_i$, we have

$$\begin{aligned} e_{i,1} &= \mathbf{z}'_i\mathbf{z}_1 + \delta_{\mathbf{u},i,1}/2, \\ e_{i,2} &= \mathbf{z}'_i\mathbf{z}_2 + \delta_{\mathbf{u},i,2}/2, \\ &\vdots \\ e_{i,n} &= \mathbf{z}'_i\mathbf{z}_n + \delta_{\mathbf{u},i,n}/2, \end{aligned}$$

and we see that sampling $\mathbf{z}_i$ from its full conditional is equivalent to a (Bayesian) linear regression problem. Modeling the $\mathbf{z}$'s as a priori independent multivariate normal $(\mathbf{0},\boldsymbol{\Sigma}_{\mathbf{z}})$ variables, the full conditional of $\mathbf{z}_i$ is multivariate normal $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = 4\boldsymbol{\Sigma}\mathbf{Z}_{-i}\mathbf{e}_{i,-i}/\sigma^2_{\mathbf{u}}$$

and

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}^{-1}_{\mathbf{z}} + 4\mathbf{Z}_{-i}\mathbf{Z}'_{-i}/\sigma^2_{\mathbf{u}}\right)^{-1},$$

where $\mathbf{Z}_{-i}$ denotes the $K \times (n-1)$ matrix obtained by removing the ith column of $\mathbf{Z}$ and $\mathbf{e}_{i,-i}$ denotes the vector of residuals $\{e_{i,j} : j \neq i\}$. Using an inverse Wishart$(\boldsymbol{\Sigma}_{\mathbf{z}0},\nu)$ prior, the full conditional of $\boldsymbol{\Sigma}_{\mathbf{z}}$ is inverse Wishart$(\boldsymbol{\Sigma}_{\mathbf{z}0} + \mathbf{Z}\mathbf{Z}', \nu+n)$. Alternatively, if we restrict $\boldsymbol{\Sigma}_{\mathbf{z}}$ to be $\sigma^2_{\mathbf{z}}\mathbf{I}_{K\times K}$ and use an inverse gamma(α_0,α_1) prior for $\sigma^2_{\mathbf{z}}$, then the full conditional is given by $\sigma^2_{\mathbf{z}}|\mathbf{Z} \sim$ inverse gamma$(\alpha_0 + nK/2, \alpha_1 + \text{trace}(\mathbf{Z}'\mathbf{Z})/2)$.

4.3 Posterior Analysis of Z

Note that the probability model depends on $\mathbf{Z}$ only through the matrix of inner products $\mathbf{Z}'\mathbf{Z}$, which is invariant under rotations and reflections of $\mathbf{Z}$. Therefore, $\log\Pr(\mathbf{Y}|\mathbf{Z},\boldsymbol{\beta},\mathbf{X}) = \log\Pr(\mathbf{Y}|\mathbf{Z}^*,\boldsymbol{\beta},\mathbf{X})$ for any $\mathbf{Z}^*$, which is equivalent to $\mathbf{Z}$ under the operations of rotation or reflection. Values of $\mathbf{Z}$ sampled from the posterior distribution may at first seem highly variable, but perhaps are nearly rotations of each other and are thus not highly variable in terms of the resulting inner-product matrices. To appropriately compare sample values of $\mathbf{Z}$, we must first rotate them to a common orientation. This can be done using, for example, a "Procrustean" transformation (Sibson 1978) in which for each $\mathbf{Z}$ sampled from the posterior, we find the rotation $\mathbf{Z}^*$ of $\mathbf{Z}$ that has the smallest sum of squared deviations from an arbitrary fixed reference matrix $\mathbf{Z}_0$. The rotated matrix $\mathbf{Z}^*$ that minimizes the sum of squares is given by $\mathbf{Z}^* = \mathbf{Z}_0\mathbf{Z}'(\mathbf{Z}\mathbf{Z}'_0\mathbf{Z}_0\mathbf{Z}')^{-1/2}\mathbf{Z}$. Note that the posterior distribution of $\mathbf{Z}^{*\prime}\mathbf{Z}^*$ is equivalent to that of $\mathbf{Z}'\mathbf{Z}$ (see Hoff et al. 2002 for further discussion). Another approach to making inference on $\mathbf{Z}$ is by computing the posterior mean of $\mathbf{Z}'\mathbf{Z}$. Note that the resulting $n \times n$ matrix in general will not be representable by the inner products of K-dimensional vectors. However, one can obtain a least squares estimate by computing the eigen decomposition of $E(\mathbf{Z}'\mathbf{Z}|\mathbf{Y})$ and then obtaining $\hat{\mathbf{Z}}$ as the first K eigenvectors, each multiplied by the square root of their associated eigenvalue.

5. SELECTION OF *K*

Selection of the dimension K of the latent $\mathbf{z}$-vectors will generally depend on the goal of the analysis. For example, if the goal is descriptive (i.e., the desired end result is a decomposition of the variance into easily presentable components), then a choice of $K = 1, 2$, or 3 would allow for a simple graphical presentation of a multiplicative component of the variance.

In other situations, the purpose of the model is to predict unobserved data. For example, suppose that only a subset of the $n(n-1)$ responses were randomly chosen to be measured. As long as we have some measurements for each unit, we can estimate the effects $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{z}$ for each unit and make predictions for missing responses based on these estimates. To compare predictive performance across models with different values of K, we might use an L-fold cross-validation procedure as follows:

1. Randomly split the set of ordered pairs $\{(i,j) : i \neq j\}$ into L test sets $A_1,\ldots,A_L$.
2. For $K = 0, 1, 2, \ldots$:
   a. For $l = 1,\ldots,L$:
      (1) perform the MCMC algorithm using only $\{y_{i,j} : (i,j) \notin A_l\}$, but sample values of $\theta_{i,j}$ for all ordered pairs.
      (2) Based on the sampled values of $\theta_{i,j}$ compute the posterior mean $\hat{\theta}_{i,j}$ for $(i,j) \in A_l$ and the log predictive probability $\text{lpp}(A_l) = \sum_{(i,j)\in A_l} \log p(y_{i,j}|\hat{\theta}_{i,j})$.
   b. Measure the predictive performance for K as $\text{LPP}(K) = \sum_{l=1}^{L} \text{lpp}(A_l)$.

Many authors suggest using likelihood-based measures of fit for model selection. For the models in this article, one might consider obtaining estimates $\hat{\boldsymbol{\theta}}^{(K)}$ for different values of K and comparing $\log p(Y|\hat{\boldsymbol{\theta}}^{(K)}) = \sum_{i\neq j} \log p(y_{i,j}|\hat{\theta}^{(K)}_{i,j})$. However, such a quantity may not be ideal for selecting between models. As described at the beginning of Section 4, the model is essentially unrestricted in the $\boldsymbol{\theta}$'s, giving a nearly saturated fit that does not depend much on the choice of K or the regressors (provided that the prior for $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is sufficiently diffuse). A likelihood that is more appropriate is obtained from the marginal probability of data within a pair, $\log p(\mathbf{Y}|\boldsymbol{\beta},\mathbf{a},\mathbf{b},\mathbf{Z},\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) = \sum_{\{i,j\}} \log p(y_{i,j}, y_{j,i}|\boldsymbol{\beta}, a_i, b_j, a_j, b_i, \mathbf{z}_i, \mathbf{z}_j, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$, where the sum is over unordered pairs. Note that this treats the $\mathbf{a}$'s, $\mathbf{b}$'s, and $\mathbf{z}$'s as fixed effects. Also, computation of $\log p(y_{i,j}, y_{j,i}|\boldsymbol{\beta}, a_i, b_j, a_j,$

$b_i, \mathbf{z}_i, \mathbf{z}_j, \boldsymbol{\Sigma}_\gamma$) involves an integral over $\{\gamma_{i,j}, \gamma_{j,i}\}$ and needs to be approximated except in the case of the normal model with the identity link.

Having obtained posterior estimates $\hat{\psi}^{(K)} = \{\hat{\boldsymbol{\beta}}, \hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{Z}}, \hat{\boldsymbol{\Sigma}}_\gamma\}$ for a range of K, one can compare the values of $\log p(\mathbf{Y}|\hat{\psi}^{(K)})$ to assess model fit versus complexity. As a function of K, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are

$$\mathrm{AIC}(K) = -2\log p\big(\mathbf{Y}\big|\hat{\psi}^{(K)}\big) + c + [2n] \times K$$

and

$$\mathrm{BIC}(K) = -2\log p\big(\mathbf{Y}\big|\hat{\psi}^{(K)}\big) + c + \left[n \log\binom{n}{2}\right] \times K,$$

where the suggestion is to prefer the model with a lowest value of the criterion. However, the **a**'s, **b**'s, and **z**'s are being modeled as random effects, and treating them as fixed effects as in the foregoing criteria may lead to overestimating of model complexity. For hierarchical models, Spiegelhalter, Best, Carlin, and van der Linde (2002) suggested using the deviance information criterion (DIC), $\mathrm{DIC}(K) = -2\log p(\mathbf{Y}|\hat{\psi}^{(K)}) + 2 \times p_D^{(K)}$, where the penalty $p_D^{(K)}$ on the model complexity is given by

$$p_D^{(K)} = -2 \times \big\{E\big[\log p\big(\mathbf{Y}\big|\psi^{(K)}\big)\big|\mathbf{Y}\big] - \log p\big(\mathbf{Y}\big|\hat{\psi}^{(K)}\big)\big\}.$$

This expectation can be approximated by averaging over MCMC samples. The penalty term $p_D^{(K)}$ has been referred to as the "effective number of parameters," because it has this interpretation in normal linear models. However, for other models, the interpretation and adequacy of the penalization has been a matter of debate (see, e.g., the discussion of Spiegelhalter et al. 2002).

Finally, if one is interested in particular features of the data or in examining particular aspects of lack of fit, one can evaluate the model with posterior predictive checks. This is done by comparing the observed value of a statistic of interest $T(\mathbf{Y})$ with its posterior predictive distribution $p[T(\mathbf{Y}_{\mathrm{pred}})|\mathbf{Y}]$. The idea is that if $p[T(\mathbf{Y}_{\mathrm{pred}})|\mathbf{Y}]$ does not put much mass near $T(\mathbf{Y})$, then the model is not able to capture this feature of the data well (see Gelman, Carlin, Stern, and Rubin 1995, chap. 6; Gelman, Meng, and Stern 1996). In terms of selecting a value for K, one might proceed by selecting the smallest K for which there is no substantial lack of fit for a set of statistics of interest. For dyadic data, some statistics of interest might involve measures of third-order dependence as described earlier, within-unit variability, variability in outdegree or indegree, and others. Comparison of various statistics of interest to reference distributions has been a fundamental tool for inference and model evaluation in the social networks literature (Wasserman and Faust 1994, chaps. 13 and 14; Besag 2000).

6. DATA ANALYSIS: INTERNATIONAL RELATIONS IN CENTRAL ASIA

We analyze data on international relations in central Asia as recorded by the Kansas Event Data (KEDS) project (*http://www.ku.edu/~keds/project.html*) and described by Schrodt, Simpson, and Gerner (2001). News stories are downloaded from the Reuters Business Briefing Service on Afghanistan, Armenia, Azerbaijan, and the former Soviet Republics of Central Asia, and political interactions between countries are recorded and categorized. We take our response $y_{i,j}$ to be the total number of "positive" actions reportedly initiated by country i with target j from 1992 to 1999 (i.e., after the breakup of the Soviet Union), as recorded by the KEDS project. Positive actions here include such events as approval, endorsement, or praise of one government by another; military assistance; formation of alliances; promises of financial or policy support; and others (essentially all events having Goldstein scale values greater than 2.5, except cease-fire or ceding of power; see the KEDS project webpage for more details). We include in our population the 99 countries closest in geographic distance to Afghanistan, plus the United States, giving a total of $n = 100$ countries for analysis. We note that 17 of the 100 countries had 0 actions as either initiators or targets of actions over the 7-year period.

6.1 Data Description

The occurrence of an event between any two given countries in these data is rare, with 92.7% of the nondiagonal entries of the sociomatrix being equal to 0. Some descriptive plots of the raw data are given in Figure 2. Panel (a) plots jittered values of $\log(1 + \sum_{j:j\neq i} y_{i,j})$ versus $\log(1 + \sum_{j:j\neq i} y_{j,i})$ for each country i. The quantities $\sum_{j:j\neq i} y_{i,j}$ and $\sum_{j:j\neq i} y_{j,i}$ are typically called the outdegree and indegree of unit i. Note the strong correlation, which suggests a large value of $\sigma_{\mathbf{ab}}/(\sigma_{\mathbf{a}}\sigma_{\mathbf{b}})$ in the random-effects model being considered. Panel (b) plots the log of each country's outdegree plus 1, $\log(1 + \sum_{j:j\neq i} y_{i,j})$, versus log population, which suggests a positive relationship between response and population. (A plot of log-indegree versus population is similar.) Panel (c) plots the response on a log scale versus the geographic distance in thousands of miles between countries i and j. More precisely, this distance is the "minimum distance" between two countries, which is 0 if i and j share a border. On average, the number of events between two countries decreases as geographic distance increases. This pattern is made more clear by separating out the measurements involving the United States (which are circled).

6.2 Evidence of Third-Order Dependence

Before fitting a somewhat complicated bilinear Poisson regression model, we evaluate the necessity of such an effort by looking for evidence of balance and clusterability in the data. We do this by fitting a simple linear regression on the log-transformed data and examining the residuals for third-order dependencies of the types described in Section 3. More specifically, we obtain ordinary least squares estimates for the regression model $\log(y_{i,j} + 1) = \beta_0 + \beta_d x_{i,j} + a_i + b_j + \xi_{i,j}$, where $x_{i,j}$ is the geographic distance between i and j. For simplicity, we analyze the average residual within a dyad $\bar{\xi}_{i,j} = \frac{1}{2}(\hat{\xi}_{i,j} + \hat{\xi}_{j,i})$.

There are several indications of third-order dependence in these residuals:

1. Because the mean of the residuals is 0, independence of the residuals implies that the average value of the product $\bar{\xi}_{i,j}\bar{\xi}_{j,k}\bar{\xi}_{k,i}$ over triads also should be 0. As discussed in Section 3, a value larger than 0 would indicate some degree of balance. The empirical average over triads turns

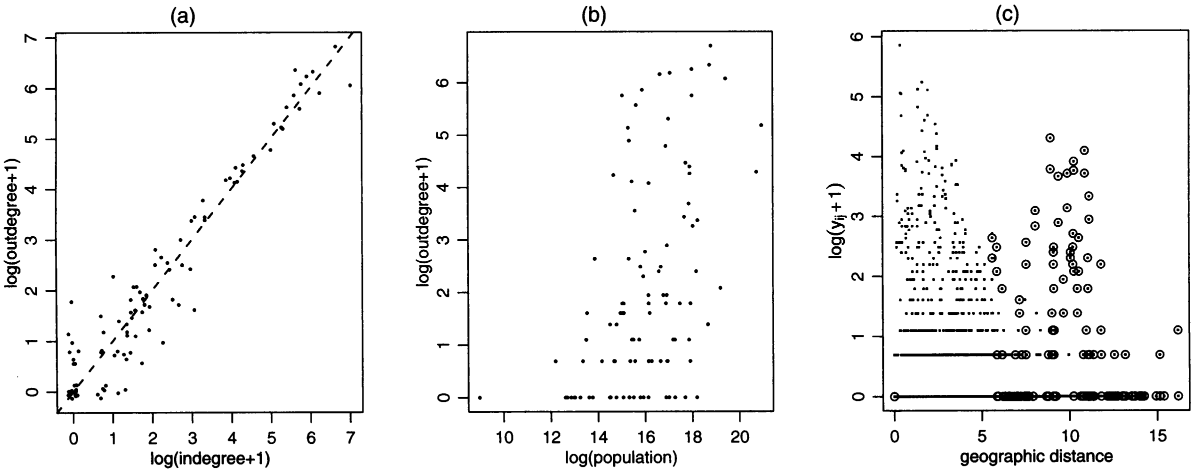

Figure 2. Relationships Between (a) Outdegree and Indegree, (b) Outdegree and Population, and (c) Response and Geographic Distance. Responses involving the United States are circled.

out to be .026. This is quite a bit larger than we would expect to see due to just noise. Under 1,000 random permutations of the residuals, the largest absolute value of this statistic was .0005.

2. The fraction of residuals that are positive is $p = .56$ (the distribution of residuals is not symmetric). Under independence, the proportions of cycles that we would expect in the two balanced categories shown in Figure 1 ($+ + +$ and $+ - -$) are $p^3 = .176$ and $3p(1-p)^2 = .325$, whereas the observed proportions are substantially higher, .306 and .448. The observed proportion in the unclusterable category ($+ + -$) is .163, which is substantially lower than the value expected under independence, $3p^2(1-p) = .414$. Thus we have many more balanced cycles and far fewer unclusterable cycles than we would expect under independence. (The expected proportion in the clusterable but unbalanced category was .085, about the same as the observed proportion, .083.)
3. As described in Section 3, in a balanced system we expect that if $\bar{\xi}_{i,j} > 0$, then $\bar{\xi}_{i,k}$ and $\bar{\xi}_{j,k}$ will have the same sign. Such a pattern is shown graphically in Figure 3, which for each pair $\{i,j\}$ plots $\bar{\xi}_{i,j}$ versus the proportion of other nodes k for which $\bar{\xi}_{i,k} \times \bar{\xi}_{j,k} > 0$. Although the distribution of residuals is far from normal (due to the skew of the data and the simplistic regression model on the log-transformed scale), we do see some indication of this type of third-order dependence. As we would expect from a balanced system, pairs $\{i,j\}$ for which $\bar{\xi}_{i,j}$ is less than 0 generally have dissimilar residuals to others [$\hat{\Pr}(\bar{\xi}_{i,k} \times \bar{\xi}_{j,k} > 0)$ tends to be less than .5], pairs $\{i,j\}$ for which $\bar{\xi}_{i,j}$ is greater than 0 generally have similar residuals to others [$\hat{\Pr}(\bar{\xi}_{i,k} \times \bar{\xi}_{j,k} > 0)$ tends to be greater than .5], and $\hat{\Pr}(\bar{\xi}_{i,k} \times \bar{\xi}_{j,k} > 0)$ is generally increasing in $\bar{\xi}_{i,j}$.

We also checked the foregoing patterns in residuals under the same regression model but with square-root and quarter-power transformations, and also under a Poisson regression model (using the working residuals). Results from these alternative regression models gave similar indications of third-order dependence.

6.3 Model and Priors

Note that the data are from an observational study and are not randomly sampled. Rather, we have defined a population of units based on geographic distance and have measurements on all pairs. For this analysis, we interpret a probability model primarily as a tool for describing the variance in the dataset and the

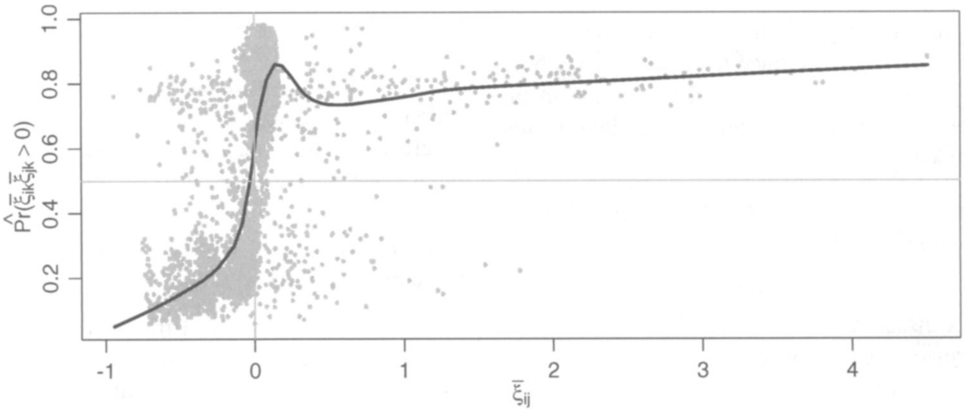

Figure 3. Balanced Residuals. As $\bar{\xi}_{i,j}$ increases, so does the average number of k for which $\bar{\xi}_{i,k}$ and $\bar{\xi}_{j,k}$ have the same sign.

regression coefficients as measures of the multiplicative, or log-linear, components of the relationship between response and regressors.

We fit the random-effects model (4) to the data using a Poisson distribution and the log-link, so that each response $y_{i,j}$ is assumed to have come from a Poisson population with mean $e^{\theta_{i,j}}$, and the $\mathbf{y}$'s are conditionally independent given the $\boldsymbol{\theta}$'s. We decompose the variance in the $\boldsymbol{\theta}$'s as

$$\theta_{i,j} = \beta_0 + \beta_d x_{i,j} + \beta_{\mathbf{s}} x_i + \beta_{\mathbf{r}} x_j + \epsilon_{i,j}$$

and

$$\epsilon_{i,j} = a_i + b_j + \gamma_{i,j} + \mathbf{z}_i'\mathbf{z}_j,$$

where $x_{i,j}$ is the geographic distance between i and j and x_i is the log population of i. For estimation of variance components, we model the random effects as having the following multivariate normal distributions: $(a_i, b_i)' \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{ab}})$, $(\gamma_{i,j}, \gamma_{j,i})' \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma})$, and $\mathbf{z}_i \sim \text{MVN}(\mathbf{0}, \sigma_{\mathbf{z}}^2 \mathbf{I}_{k\times k})$. Prior distributions of the parameters are taken to be

- $\boldsymbol{\beta} \sim \text{MVN}(\mathbf{0}, 100 \times \mathbf{I}_{4\times 4})$
- $\boldsymbol{\Sigma}_{\mathbf{ab}} \sim$ inverse Wishart$(\mathbf{I}_{2\times 2}, 4)$
- $\sigma_{\mathbf{u}}^2, \sigma_{\mathbf{v}}^2 \sim$ iid inverse gamma$(1, 1)$, $\sigma_{\gamma}^2 = (\sigma_{\mathbf{u}}^2 + \sigma_{\mathbf{v}}^2)/4$, $\rho = (\sigma_{\mathbf{u}}^2 - \sigma_{\mathbf{v}}^2)/(\sigma_{\mathbf{u}}^2 + \sigma_{\mathbf{v}}^2)$.

Posterior calculations proceed as described in Section 4.

6.4 Selecting the Latent Dimension

We performed a four-fold cross-validation procedure as described in Section 5 for these data, and give the results in Table 1. The table also includes are the marginal probability criteria and the DIC penalty, p_D, in the third column. The cross-validation procedure suggests that models having a dimension of $K = 2, 3$, or 4 have roughly the same predictive performance. In terms of the marginal likelihood criterion, the biggest improvements in fit are in going from $K = 0$ to $K = 1$ and from $K = 1$ to $K = 2$. The improvements in fit in going from two to three dimensions and from three to four dimensions are smaller. Using the AIC criterion and penalizing the improvement in likelihood by the number of additional parameters (100 per additional dimension), we would choose $K = 2$. The BIC, with a higher penalty on the number of parameters, favors $K = 0$. In contrast, the DIC favors large K. Note that the increase in the DIC penalty (i.e., the estimated effective number of parameters) tends to decrease with increasing K, suggesting that the latent space is not being "fully used" for larger values of K. Additionally, although the random-effects nature of the model may prevent some overfitting for large K, a small value may be sufficient for predictive purposes, as suggested by the cross-validation results. (See the discussion in Richardson 2002 of Spiegelhalter et al. 2002 on the potential lack of parsimony of the DIC criterion.)

Table 1. Evaluation of K

K	LLP(K)	$\log p(\mathbf{Y} \mid \hat{\beta}, \hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{Z}}, \hat{\Sigma}_\epsilon)$	p_D
0	−3,558.78	−2,432.67	148.19
1	−3,351.76	−2,317.47	215.69
2	−3,078.79	−2,214.68	286.51
3	−3,076.73	−2,127.26	346.57
4	−3,077.30	−2,038.95	374.99

Based on these results (and our ability to plot in two dimensions), we choose to present the analysis of the $K = 2$ model in more detail.

6.5 Results for $K = 2$

We constructed two Markov chains of length 200,000 each using the algorithm described earlier. The first chain used starting values of 0 for all regression coefficients and country-specific intercepts, the identity matrix for $\boldsymbol{\Sigma}_{\mathbf{ab}}$ and $\boldsymbol{\Sigma}_{\gamma}$, a value of .1 for $\sigma_{\mathbf{z}}^2$, and components of $\mathbf{Z}$ sampled independently from a normal $(0, \sigma_{\mathbf{z}}^2)$ distribution. The second chain used starting values obtained from the following procedure: Maximum likelihood estimates of β_d, s, and r were obtained by fitting an ordinary generalized linear model using geographic distance as a regressor and sender and receiver labels as factor variables. Estimates of $\beta_0, \beta_s, \beta_r$, and $\boldsymbol{\Sigma}_{\mathbf{ab}}$ were obtained from the estimates of s and r. The iteratively reweighted least squares fitting procedure produces a matrix $\mathbf{R}$ of working residuals, with the off-diagonal elements undefined. An estimate $\hat{\mathbf{Z}}$ of $\mathbf{Z}$ was then obtained by approximating $\mathbf{R}$ with a matrix product of the form $\mathbf{Z}'\mathbf{Z}$. This can be done with an iterative least squares procedure, similar to the Gibbs sampling procedure outlined in Section 4.2 (see ten Berge and Kiers 1989 for more details on this problem). An estimate of $\boldsymbol{\Sigma}_{\gamma}$ is then obtained from $\mathbf{R} - \hat{\mathbf{Z}}'\hat{\mathbf{Z}}$.

Samples of parameter values were saved from the Markov chains every 100 iterations. Diagnostic plots (not shown here) suggest that both chains achieved stationarity well before 50,000 iterations, and so we base our inference on the saved samples after this point. Posterior means and standard deviations of the model parameters, based on the 3,000 saved MCMC samples (1,500 from each chain), are given in Table 2. As in the raw data, we see a negative relation between response and geographic distance ($E[\beta_d|\mathbf{Y}] = -.18$) and a positive relation between response and country populations ($E[\beta_s|\mathbf{Y}] = 1.00$, $E[\beta_r|\mathbf{Y}] = .94$). We also estimate a strong positive correlation of within-dyad responses as well as the within-country random effects a and b.

Next we analyze the posterior distribution of the $2 \times n$ matrix of latent vectors $\mathbf{Z}$ after rotating the samples to a common orientation. The resulting mean of $\mathbf{Z}^*$ is given in Figure 4. Marginal uncertainty in $\mathbf{Z}^*$ could be displayed by plotting sample $\mathbf{Z}^*$'s over the plot of the means, using colors to distinguish between countries.

Generally, two countries will be modeled as having $\mathbf{z}$'s in the same direction if they have large responses to one another relative to their total number of actions and covariate values, and/or if their responses involving other countries are similar (a model that can distinguish between these two phenomena is proposed in the discussion). For example, Croatia and Slovenia are each recorded as the initiator of an action with the other as a target,

Table 2. Posterior Means and Standard Deviations for K = 2

	β_d	β_s	β_r	$\sigma_{\mathbf{a}}^2$	$\sigma_{\mathbf{b}}^2$	$\sigma_{\mathbf{ab}}$	σ_{γ}^2	ρ	$\sigma_{\mathbf{z}}^2$
MEAN	−.18	1.00	.94	6.46	6.37	6.4	1.23	.95	1.99
SD	.04	.17	.17	1.23	1.2	1.21	.14	.01	.27

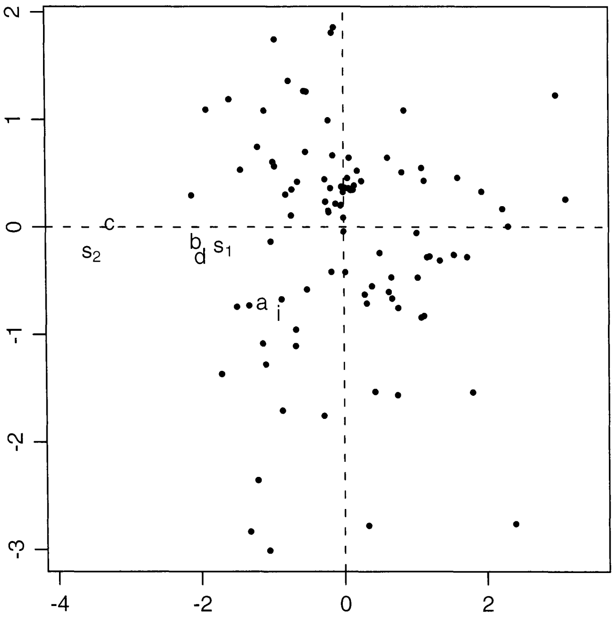

Figure 4. Posterior Mean of $\mathbf{Z}^*$. *The plotting characters for Azerbaijan, Bosnia-Herzegovina, Croatia, Denmark, Italy, Serbia, and Slovenia are "a," "b," "c," "d," "i," "*s_1*," and "*s_2*." The plotting characters for all other countries are black dots.*

and each initiates an action with Serbia as well. With the exception of one action from Slovenia to Italy, these are the only events recorded for Croatia and Slovenia, and so these countries are "similar" in that they have actions involving each other and to Serbia, and only one other action involving another country. Bosnia-Herzegovina and Denmark have no actions with Croatia or Slovenia, but like Croatia and Slovenia they each have one action with Serbia and very few actions otherwise (each has one action with Azerbaijan and no other actions), and thus are located in a similar direction. Serbia, although active with this group of countries (on the scale of their response rates), has actions with 10 other countries and thus is placed more toward the center. Of course, the posterior variances of the $\mathbf{z}$'s for Croatia, Slovenia, Bosnia-Herzegovina, and Denmark are quite high, because our information about them is coming primarily from the few nonzero responses among them. A figure attaching country names to the $\mathbf{Z}$-values of the other countries is available at my website.

Finally, we performed some model checking to evaluate whether these data were overdispersed or underdispersed relative to the Poisson model. This was done by comparing the observed overall variance and the observed within-sender variance to the posterior predictive distributions of these quantities. No lack of fit was detected, but perhaps this is not too surprising, because the error structure in this model is very flexible.

7. DISCUSSION

This article has presented an approach to modeling third-order dependence patterns often seen in dyadic datasets, such as social networks. The models are based on generalized linear mixed-effects models with the addition of a reduced-rank interaction term composed of inner products of latent characteristic vectors. Such an approach allows for the analysis of dyadic data using familiar regression tools, but also allows one to capture patterns, such as balance and clusterability, which are often of interest to social science researchers. Other approaches to capturing such dependence patterns have used metric distances (Hoff et al. 2002) and ultrametric distances (Schweinberger and Snijders 2003), although not in the presence of the covariance structure (2). Although such latent distance models may be easy to understand, the inner-product approach has some conceptual appeal, because the term $\mathbf{z}_i'\mathbf{z}_j$ can be viewed as a mean-0 random effect.

Spatial interaction data (Haynes and Fotheringham 1984; Banerjee, Gelfand, and Polasek 2000) measuring flows or transfers between pairs of locations can be seen as a type of dyadic data, although these data are typically modeled as depending on fixed, known locations. In contrast, if such data are only partially explained by known spatial locations, then it may be advantageous to model the data further with the approach outlined in this article. For example, in the data analysis example in Section 6, the geographic distances between countries only partially explained the observed data, and adding a latent inner-product effect resulted in an improved fit and increased predictive performance.

Another dependence pattern often of interest to researchers is that of stochastic equivalence, in which two units i and j are said to be stochastically equivalent if their responses have the same probability distribution, that is, $p(y_{i,1},\dots,y_{i,n}) = p(y_{j,1},\dots,y_{j,n})$. The model considered in this article, as well as the latent distance approaches mentioned earlier, potentially confound stochastic equivalence patterns with those of clusterability and balance: two units will generally be estimated as having similar latent characteristic vectors if they have strong relations to each other, or have similar relations to others' units. However, in some datasets there may be clusters of units that relate similarly to others, but not strongly to each other. Nowicki and Snijders (2001) considered a latent class model that identified clusters of such stochastically equivalent units, but did not separately consider clustering based on strength of relations. A possible approach to modeling both types of patterns is to extend the bilinear effect discussed in this article to a more general asymmetric bilinear effect such as $\mathbf{z}_i'\mathbf{R}\mathbf{z}_j$, where $\mathbf{R}$ is a $K \times K$ matrix. Estimation of similar types of effects has been considered by Gabriel (1998), and least squares representations of an asymmetric matrix $\mathbf{Y}$ by $\mathbf{Z}'\mathbf{R}\mathbf{Z}$ has been considered by ten Berge and Kiers (1989), Kiers (1989), and Trendafilov (2002), among others. In the present application, the vector $\mathbf{z}_i$ could be interpreted as giving grades of membership for unit i to each of K classes, and R_{lm} could be interpreted as the response rate from class l to m. Interestingly, the restriction of each $\mathbf{z}_i$ to be unity at one component and 0 at the others components gives a representation of the latent class model of Nowicki and Snijders (2001). Unrestricted estimation of $\mathbf{z}_i'\mathbf{R}\mathbf{z}_j$, in the presence of the error structure (2), is a topic of my current research.

Finding a default selection method for the dimension K of the latent variables $\mathbf{Z}$ is difficult, as likelihood-based criteria such as AIC, BIC, and DIC can give conflicting results. I suggest using such criteria as guidelines, but also weighing the complexity of a model versus its lack of fit, measured via the posterior predictive distribution of summary statistics of interest. If one is interested in a particular statistical quantity, then a model's the ability to predict that quantity may be more relevant than how well the model fits the entire dataset based on a global likelihood-based measure. But if one is interested in predicting individual responses, then a cross-validation procedure may be appropriate for model selection. Unfortunately, such a procedure can be very computationally intensive. Another alternative is to use Bayes factors, in which K is chosen based on the prior predictive probability $\Pr(\mathbf{Y}|K)$. However, good approximations to this marginal probability (which involves integrating over all of the parameters) are also computationally intensive for the models in this article, and simple approximations to $\Pr(\mathbf{Y}|K)$, such as the harmonic mean estimator of Newton and Raftery (1994), can be unstable. An easy-to-compute measure of predictive performance for the models discussed in this article would be extremely useful.

The data analyzed in Section 6, along with R-functions for implementing the proposed methods for both directed and undirected data, are available at my website, *www.stat.washington.edu/hoff*.

[Received July 2003. Revised June 2004.]

REFERENCES

Aldous, D. J. (1985), "Exchangeability and Related Topics," in *École d'été de Probabilités de Saint-Flour, XIII—1983*, Berlin: Springer-Verlag, pp. 1–198.

Andersson, S., and Madsen, J. (1998), "Symmetry and Lattice Conditional Independence in a Multivariate Normal Distribution," *The Annals of Statistics*, 26, 525–572.

Banerjee, S., Gelfand, A. E., and Polasek, W. (2000), "Geostatistical Modelling for Spatial Interaction Data With Application to Postal Service Performance," *Journal Statistical Planning and Inference*, 90, 87–105.

Besag, J. (2000), "Markov Chain Monte Carlo for Statistical Inference," working paper, University of Washington, Center for Statistics and the Social Sciences.

Booth, J. G., and Hobert, J. P. (1998), "Standard Errors of Prediction in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 93, 262–272.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.

Cockerham, C. C., and Weir, B. S. (1977), "Quadratic Analyses of Reciprocal Crosses," *Biometrics*, 33, 187–204.

Davis, J. (1976), "Clustering and Structural Balance in Graphs," *Human Relations*, 20, 181–187.

Gabriel, K. R. (1978), "Least Squares Approximation of Matrices by Additive and Multiplicative Models," *Journal of the Royal Statistical Society*, Ser. B, 40, 186–196.

——— (1998), "Generalised Bilinear Regression," *Biometrika*, 85, 689–700.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995), "Efficient Parameterisations for Normal Linear Mixed Models," *Biometrika*, 82, 479–488.

——— (1996), "Efficient Parameterizations for Generalized Linear Mixed Models," in *Bayesian Statistics 5*, eds. J. Bernardo et al., New York: Oxford University Press, pp. 165–180.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Gelman, A., Meng, X.-L., and Stern, H. (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies" (with discussion), *Statistica Sinica*, 6, 733–807.

Gill, P. S., and Swartz, T. B. (2001), "Statistical Analyses for Round Robin Interaction Data," *The Canadian Journal of Statistics*, 29, 321–331.

Harary, F. (1953), "On the Notion of Balance of a Signed Graph," *Michigan Mathematics Journal*, 2, 143–146.

Haynes, K., and Fotheringham, A. S. (1984), *Gravity and Spatial Interaction Models*, New York: Sage.

Heider, F. (1979), "On Balance and Attribution," in *Perspectives on Social Network Research*, eds. P. W. Holland and S. Leinhardt, New York: Academic Press, pp. 11–23.

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.

Kiers, H. A. L. (1989), "An Alternating Least Squares Algorithm for Fitting the Two- and Three-Way DEDICOM Model and the IDIOSCAL Model," *Psychometrika*, 54, 515–521.

Li, H. (2002), "Modeling Through Group Invariance: An Interesting Example With Potential Applications," *The Annals of Statistics*, 30, 1069–1080.

Li, H., and Loken, E. (2002), "A Unified Theory of Statistical Analysis and Inference for Variance Component Models for Dyadic Data," *Statistica Sinica*, 12, 519–535.

Marasinghe, M. G., and Johnson, D. E. (1982), "A Test of Incomplete Additivity in the Multiplicative Interaction Model," *Journal of the American Statistical Association*, 77, 869–877.

McGilchrist, C. A. (1994), "Estimation in Generalized Mixed Models," *Journal of the Royal Statistical Society*, Ser. B, 56, 61–69.

Natarajan, R., and Kass, R. E. (2000), "Reference Bayesian Methods for Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 95, 227–237.

Newton, M. A., and Raftery, A. E. (1994), "Approximate Bayesian Inference With the Weighted Likelihood Bootstrap" (with discussion), *Journal of Royal Statistical Society*, Ser. B, 56, 3–48.

Nowicki, K., and Snijders, T. A. B. (2001), "Estimation and Prediction for Stochastic Blockstructures," *Journal of the American Statistical Association*, 96, 1077–1087.

Oman, S. D. (1991), "Multiplicative Effects in Mixed Model Analysis of Variance," *Biometrika*, 78, 729–739.

Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2003), "Non-Centered Parameterizations for Hierarchical Models and Data Augmentation," in *Bayesian Statistics 7*, eds. J. M. Bernardo et al., New York: Oxford University Press, pp. 307–326.

Richardson, S. (2002), Discussion of "Bayesian Measures of Model Complexity and Fit" by D. J. Spiegelhalter et al., *Journal of the Royal Statistical Society*, Ser. B, 64, 626–627.

Schall, R. (1991), "Estimation in Generalized Linear Models With Random Effects," *Biometrika*, 78, 719–727.

Schrodt, P. A., Simpson, E. M., and Gerner, D. J. (2001), "Monitoring Conflict Using Automated Coding of Newswire Sources," presented at the High-Level Scientific Conference on Identifying Wars, Uppsala University, Uppsala, Sweden; available at *http://www.pcr.uu.se/Schrodt_Uppsala.pdf.*

Schweinberger, M., and Snijders, T. A. B. (2003), "Settings in Social Networks: A Measurement Model," *Sociological Methodology*, 33, 307–342.

Sibson, R. (1978), "Studies in the Robustness of Multidimensional Scaling," *Journal of the Royal Statistical Society*, Ser. B, 40, 234–238.

Snijders, T. A. B., and Kenny, D. A. (1999), "The Social Relations Model for Family Data: A Multilevel Approach," *Personal Relationships*, 6, 471–486.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society*, Ser. B, 64, 583–639.

ten Berge, J. M. F., and Kiers, H. A. L. (1989), "Fitting the Off-Diagonal DEDICOM Model in the Least-Squares Sense by a Generalization of the Harman and Jones MINRES Procedure of Factor Analysis," *Psychometrika*, 54, 333–337.

Trendafilov, N. T. (2002), "GIPSCAL Revisited: A Projected Gradient Approach," *Statistics and Computing*, 12, 135–145.

Warner, R., Kenny, D. A., and Stoto, M. (1979), "A New Round-Robin Analysis of Variance for Social Interaction Data," *Journal of Personality and Social Psychology*, 37, 1742–1757.

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

Wasserman, S., and Pattison, P. (1996), "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and $p*$," *Psychometrika*, 61, 401–425.

Wolfinger, R., and O'Connell, M. (1993), "Generalized Linear Mixed Models: A Pseudo-Likelihood Approach," *Journal of Statistical Computation and Simulation*, 48, 233–243.

Wong, G. Y. (1982), "Round-Robin Analysis of Variance via Maximum Likelihood," *Journal of the American Statistical Association*, 77, 714–724.

Zeger, S. L., and Karim, M. R. (1991), "Generalized Linear Models With Random Effects: A Gibbs Sampling Approach," *Journal of the American Statistical Association*, 86, 79–86.