Statistical Science
2021, Vol. 36, No. 1, 34–50
https://doi.org/10.1214/19-STS757

Additive and Multiplicative Effects Network Models

Peter Hoff

Abstract. Network datasets typically exhibit certain types of statistical patterns, such as within-dyad correlation, degree heterogeneity, and triadic patterns such as transitivity and clustering. The first two of these can be well represented with a social relations model, a type of additive effects model originally developed for continuous dyadic data. Higher-order patterns can be represented with multiplicative effects models, which are related to matrix decompositions that are commonly used for matrix-variate data analysis. Additionally, these multiplicative effects models generalize other popular latent feature network models, such as the stochastic blockmodel and the latent space model. In this article, we review a general regression framework for the analysis of network data that combines these two types of effects, and accommodates a variety of network data types, including continuous, binary and ordinal network relations.

Key words and phrases: Bayesian, factor model, generalized linear model, latent variable, matrix decomposition, mixed effects model.

1. INTRODUCTION

Network data provide quantitative information about relationships among objects, individuals or entities, which we refer to as nodes. Most network data quantify pairwise relationships between nodes. A pair of nodes is referred to as a *dyad*, and a quantity that is measured or observed for multiple dyads is called a *dyadic variable*. Common sample spaces for dyadic variables include continuous, discrete, dichotomous and ordinal spaces, among others. Examples of dyadic variables include quantitative measures of trade flows between countries, communications among people, binding activity among proteins, and structural connections among regions of the brain, to name just a few.

Measurements of a dyadic variable on a population of n nodes may be summarized with a *sociomatrix*, an $n \times n$ square matrix $\mathbf{Y}$ with an undefined diagonal, where entry $y_{i,j}$ denotes the value of the relationship between nodes i and j from the perspective of node i, or in the direction from i to j. If $y_{i,j}$ is equal to $y_{j,i}$ by design, then $\mathbf{Y}$ is necessarily symmetric and we say the variable is undirected. If $y_{i,j}$ is potentially different from $y_{j,i}$ then the variable is directed. In what follows, we assume that dyadic variables are directed unless stated otherwise.

Peter Hoff is Professor, Department of Statistical Science, Duke University, P.O. Box 90251, Durham, North Carolina 27708-0251, USA (e-mail: peter.hoff@duke.edu; URL: pdhoff.github.io).

Analysis of an observed sociomatrix $\mathbf{Y}$ often proceeds in the context of one or more statistical models, with which a data analyst may evaluate competing theories of network formation, describe patterns in the network, estimate effects of other variables on dyadic relations, or impute missing values. Much of the literature on dyadic data analysis has focused on binary network data for which the sociomatrix $\mathbf{Y}$ can be viewed as the adjacency matrix of a graph. Many statistical random graph models are motivated by intuitive, preconceived notions of how networks may form, particularly social networks. For example, preferential attachment models view an observed network as the end result of a social process in which nodes are sequentially introduced into a population of existing nodes (Price, 1976). As another example, the parameters in commonly used exponential family random graph models have interpretations as node-level preferences for certain links patterns (Wasserman and Pattison, 1996).

However, many types of dyadic data are not dichotomous in their raw form, in which case the sociomatrix $\mathbf{Y}$ does not correspond to a graph. For dichotomous and nondichotomous data alike, an alternative foundation upon which to build a statistical model for $\mathbf{Y}$ is from its inherent structure as a sociomatrix, that is, as a data matrix whose row labels are the same as its column labels. Such an approach can build upon familiar, well-developed statistical methodologies such as ANOVA, linear regression, matrix decompositions, factor analysis and linear and generalized linear mixed effects models, and can be

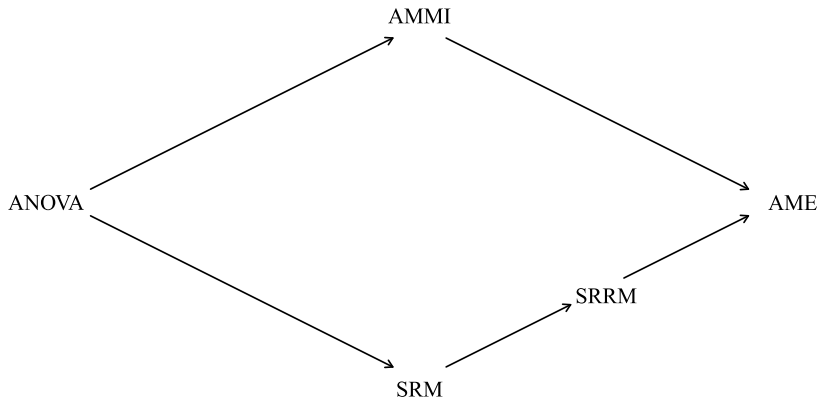

FIG. 1. *A graph describing the relationships between models. An arrow is drawn from one model to another if the former can be viewed as a submodel of the latter.*

applied to a wide variety of dyadic data types. In this article, we review such a framework for network data analysis using these tools, starting with simple ANOVA-style decompositions of sociomatrices and ending with additive and multiplicative random effects regression models for continuous, binary, ordinal and other types of dyadic network data.

In the next section, we review an ANOVA-style decomposition of a sociomatrix known as the social relations model (SRM) (Warner, Kenny and Stoto, 1979, Wong, 1982), which corresponds to a particular additive nodal effects model for directed dyadic data. An extension of this model that includes covariates is also developed, which we call the social relations regression model (SRRM). The SRM and SRRM are able to describe network variances and covariances, but are unable to describe commonly observed third-order patterns involving triads of nodes, such as transitivity, balance, or clustering. In Section 3 we discuss how, for both directed and undirected data, such patterns can be represented by a multiplicative nodal effects model, in which the relationship between two nodes depends on how similar they are in terms of unobserved node-specific latent features. From a matrix decomposition perspective, the combination of additive and multiplicative nodal effects corresponds to an "additive main effects, multiplicative interaction" (AMMI) matrix model (Gollob, 1968, Bradu and Gabriel, 1974). Combining an AMMI model with a social relations covariance model for the additive effects yields what we call an additive and multiplicative effects (AME) network model. A graph describing the relationships between all of these models is shown in Figure 1.

The components of an AME model are related to linear regression, linear mixed effects models and matrix decompositions—methods that are most appropriate for continuous data consisting of a signal of interest plus Gaussian noise. In contrast, many dyadic variables are discrete, ordinal, binary or sparse. In Section 4, we extend the AME framework to accommodate these and other types of dyadic variables using a Gaussian transformation model. In Section 5, we compare the multiplicative effects component of an AME model with two other latent feature models, the stochastic blockmodel (Nowicki and Snijders, 2001) and the latent space model (Hoff, Raftery and Handcock, 2002). We review results showing that these latter two models can be viewed as submodels of the multiplicative effects model. This is no coincidence—the blockmodel and distance model were the first nonadditive latent feature network models and are the precursors to the AME models presented here. More specifically, a simple multiplicative effects model for binary data appeared in Hoff, Raftery and Handcock (2002), along with some other nonadditive nodal effects models. Symmetric multiplicative effects were combined with a social relations generalized linear model for binary and count data in Hoff (2005). The basic form of the AME model developed in the present article first appeared in Hoff et al. (2013), and is implemented in the R software package `amen` (Hoff et al., 2012), the current version of which is available at https://pdhoff.github.io/amen/.

Some recent applications of AME models include analyses of data from Syrian rebel groups (Gade, Hafez and Gabbay, 2019), bird populations (Genovart et al., 2019), Nigerian conflicts (Dorff, Gallop and Minhas, 2019), and social networks of Nicaraguan horticulturalists (Koster, 2018). A primary motivation for the AME framework in these applications is to provide inference for parameters describing relationships between the sociomatrix and observed nodal and dyadic covariates, while accounting for certain types of second- and third-order dependence patterns often observed in network data. However, an alternative thread of research (Young and Scheinerman, 2007, Hoff, 2008, 2009b, Rohe, Chatterjee and Yu, 2011) has focused on estimation of the latent features themselves, as these provide an embedding of the nodes in a low-dimensional space that can assist with visualization of network structure, or help uncover "communities"

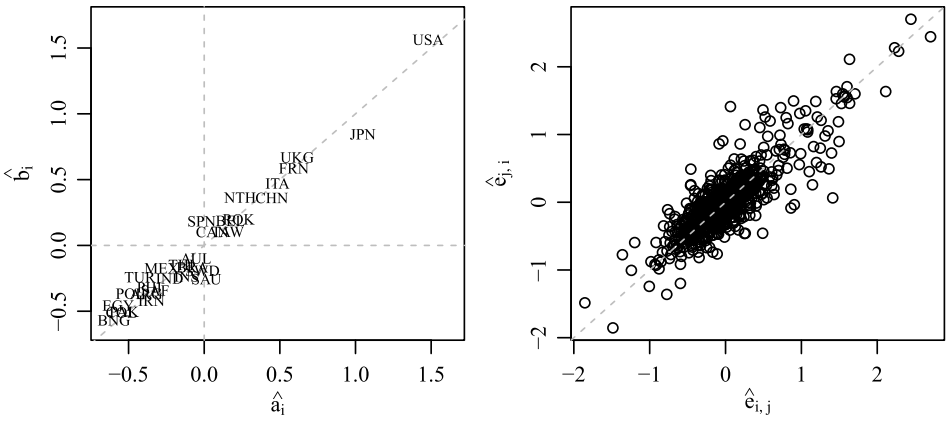

FIG. 2. *Left panel*: *Scatterplot of country-level export effects versus import effects*. *Right panel*: *Scatterplot of dyadic residuals*.

of nodes that behave similarly or have high within-group rates of interaction.

An alternative popular approach to describing network structure and dependence patterns is with exponentially parameterized random graph models (ERGMs) (Wasserman and Pattison, 1996). Connections between AME models and ERGMs are discussed in Section 5, along with a summary of the limitations of AME models in terms of evaluating certain types of network dependencies. Section 6 presents a Markov chain Monte Carlo algorithm for Bayesian model fitting of a hierarchy of AME network models. Section 7 discusses some directions for future research.

2. SOCIAL RELATIONS REGRESSION

2.1 ANOVA and Social Relations Model

Numeric sociomatrices typically exhibit certain statistical features. For example, it is often the case that values of the dyadic variable in a given row of the sociomatrix are correlated with one another, in the sense that high and low values are not equally distributed among the rows, resulting in substantial heterogeneity of the row means of the sociomatrix. Such heterogeneity can be explained by the fact that the relations within a row all share a common "sender," or row index. If sender i_1 is more "sociable" than sender i_2, we would expect the values in row i_1 to be larger than those in row i_2, on average. In this way, heterogeneity of the nodes in terms of their sociability contributes to an across-row variance of the row means of the sociomatrix. Similarly, nodal heterogeneity in "popularity" contributes to the across-column variance of the column means.

A classical approach to evaluating across-row and across-column heterogeneity in a data matrix is the ANOVA decomposition. A statistical model based on the ANOVA decomposition posits that the variability of the $y_{i,j}$'s around some overall mean μ is well represented by additive row and column effects:

$$y_{i,j} = \mu + a_i + b_j + \epsilon_{i,j}. \tag{2.1}$$

In this model, heterogeneity among the a_i's and b_j's gives rise to observed heterogeneity in the row means and column means of the sociomatrix, respectively.

While straightforward to implement, a classical ANOVA analysis ignores a fundamental characteristic of directed dyadic data: Each node appears in the dataset as both a sender and a receiver of relations, or equivalently, the row and column labels of the data matrix refer to the same set of nodes. In the context of the ANOVA model, this means that each node i has two additive effects: a row effect a_i and a column effect b_i. Since each pair of effects (a_i, b_i) shares a node, a correlation between the vectors $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ may be expected. Additionally, each dyad $\{i, j\}$ has two outcomes, $y_{i,j}$ and $y_{j,i}$. As such, the possibility that $\epsilon_{i,j}$ and $\epsilon_{j,i}$ are correlated should be considered.

We illustrate these phenomena empirically with a sociomatrix of export data among $n = 30$ countries. Here, $y_{i,j}$ is the 1990 export volume from country i to country j, in log billions of dollars. For each country $i = 1, \ldots, n$, $\hat{a}_i$ is the ith row mean minus the grand mean $\hat{\mu}$ of the sociomatrix, and $\hat{b}_i$ is the ith column mean minus $\hat{\mu}$. The left panel of Figure 2 shows that these row and column effects are strongly correlated—countries with large export volumes typically have larger than average import volumes as well. A scatterplot of $\hat{\epsilon}_{i,j} = y_{i,j} - (\hat{\mu} + \hat{a}_i + \hat{b}_j)$ versus $\hat{\epsilon}_{j,i}$ in the right panel of the plot indicates a strong dyadic correlation, even after controlling for country-specific heterogeneity in export and import volumes.

The standard ANOVA model of a data matrix quantifies row variation, column variation and residual variation. However, the ANOVA model does not quantify the sender–receiver or dyadic correlations that are apparent from the figure, and that are present in most other dyadic datasets that I have seen. A model that does quantify these

correlations, and therefore provides a more complete description of many sociomatrices, was introduced in the social psychology literature by Warner, Kenny and Stoto (1979). This more complete model, called the social relations model (SRM), is a random effects model given by (2.1) but with the additional assumptions that the a_i's, b_j's and $\epsilon_{i,j}$'s are mean-zero random variables for which

$$\begin{aligned} \operatorname{Var}\left[\begin{pmatrix} a_i \\ b_i \end{pmatrix}\right] &= \Sigma = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \\ \operatorname{Var}\left[\begin{pmatrix} \epsilon_{i,j} \\ \epsilon_{j,i} \end{pmatrix}\right] &= \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \end{aligned} \tag{2.2}$$

with effects otherwise being independent. Straightforward calculations show that under this random effects model, the variance of the relational variable is $\operatorname{Var}[y_{i,j}] = \sigma_a^2 + \sigma_b^2 + \sigma^2$, and the covariances among the relations are

$$\operatorname{Cov}[y_{i,j}, y_{i,k}] = \sigma_a^2$$

(within-row covariance)

$$\operatorname{Cov}[y_{i,j}, y_{k,j}] = \sigma_b^2$$

(within-column covariance)

$$\operatorname{Cov}[y_{i,j}, y_{j,k}] = \sigma_{ab}$$

(row-column covariance)

$$\operatorname{Cov}[y_{i,j}, y_{j,i}] = 2\sigma_{ab} + \rho\sigma^2$$

(row-column covariance plus reciprocity)

with all other covariances between elements of $\mathbf{Y}$ being zero. We refer to this covariance model as the *social relations covariance model*. Unbiased moment-based estimators of μ, Σ, σ^2 and ρ are derived in Warner, Kenny and Stoto (1979), and standard errors for these estimators are obtained in Bond and Lashley (1996). Under the additional assumption that the random effects are jointly normally distributed, Wong (1982) provides an EM algorithm for maximum likelihood estimation, Gill and Swartz (2001) develop a Bayesian method for parameter estimation, and Li and Loken (2002) discuss connections to models in genetics and extensions to repeated-measures dyadic data.

2.2 Social Relations Regression Model

Often we wish to quantify the association between a particular dyadic variable and some other dyadic or nodal variables. Useful for such situations is a type of linear mixed effects model we refer to as the *social relations regression model* (SRRM), which combines a linear regression model with the covariance structure of the SRM as follows:

$$y_{i,j} = \boldsymbol{\beta}^\top \mathbf{x}_{i,j} + a_i + b_j + \epsilon_{i,j}, \tag{2.3}$$

where $\mathbf{x}_{i,j}$ is a p-dimensional vector of regressors and $\boldsymbol{\beta}$ is a vector of regression coefficients to be estimated. The vector $\mathbf{x}_{i,j}$ may contain variables that are specific to nodes or pairs of nodes. For example, we may have $\mathbf{x}_{i,j} = (\mathbf{x}_{r,i}, \mathbf{x}_{c,j}, \mathbf{x}_{d,i,j})$ where $\mathbf{x}_{r,i}$ is a vector of characteristics of node i as a sender or row object, $\mathbf{x}_{c,j}$ is a vector of characteristics of node j as a receiver or column object, and $\mathbf{x}_{d,i,j}$ is a vector of characteristics of the ordered pair (i, j).

We illustrate the use of the SRRM with a more detailed analysis of the international trade dataset described above. This dataset also includes several other variables, such as country-specific measures of gross domestic product (GDP) and polity (a measure of citizen access to government), as well as the geographic distance between pairs of county capitals. Our objective in this example is to quantify the relationship between trade and polity after controlling for the effects of GDP and geographic distance. We first do so with a naive ordinary linear regression model of the form

$$\begin{aligned} y_{i,j} = \beta_0 &+ \beta_{r,1}\text{polity}_i + \beta_{r,2}\text{gdp}_i + \beta_{c,1}\text{polity}_j \\ &+ \beta_{c,2}\text{gdp}_j + \beta_d \text{distance}_{i,j} + \epsilon_{i,j}, \end{aligned}$$

where polity_i is a measure of country i's polity score on a scale from 1 to 10, gdp_i is the log GDP of country i in dollars, $\text{distance}_{i,j}$ is the log distance in miles between capitals of countries i and j, and the $\epsilon_{i,j}$'s are assumed to be i.i.d. mean-zero error terms. This model is a "gravity model" of trade (Isard, 1954, Bergstrand, 1985), where trade flow is analogous to a gravitational force between countries, and GDP plays the role of mass. Gravity models of this type are widely used to empirically evaluate different theories of international trade (Baier and Bergstrand, 2009).

Regression parameter estimates and standard errors assuming an i.i.d. error model are given in the first column of Table 1. Based upon the ratio of parameter estimates to standard errors, we would conclude that the hypothesis of no polity effects is inconsistent with an i.i.d. error model. However, while technically valid, this conclusion is not particularly interesting given that we expect row, column and dyadic dependence for network data such as these, and thus doubt the i.i.d. error model *a priori*. More interesting is an evaluation of whether or not the hypothesis of no polity effects is consistent with a social relations covariance model. The parameter estimates and standard errors for the SRRM in the second column of the table indicate that indeed it is: the parameter estimates of the polity effects are not substantially larger than their standard errors.

3. MULTIPLICATIVE EFFECTS MODELS

While more reasonable than an ordinary regression model, SRRMs applied to many datasets often exhibit

TABLE 1
Parameter estimates and standard errors from the trade data using a normal linear regression model with i.i.d. errors, a SRRM, and an AME model

	IID			SRRM			AME		
Regressor	$\hat{\beta}$	se($\hat{\beta}$)	t-ratio	$\hat{\beta}$	se($\hat{\beta}$)	t-ratio	$\hat{\beta}$	se($\hat{\beta}$)	t-ratio
exporter polity	0.015	0.004	4.166	0.015	0.016	0.939	0.013	0.016	0.786
importer polity	0.022	0.004	6.070	0.022	0.016	1.420	0.018	0.015	1.173
exporter GDP	0.411	0.021	19.623	0.401	0.097	4.117	0.340	0.103	3.306
importer GDP	0.398	0.020	19.504	0.391	0.093	4.189	0.331	0.101	3.266
distance	−0.057	0.004	−13.360	−0.064	0.006	−11.578	−0.041	0.004	−10.724

noticeable lack of fit. In particular, it is often observed that real networks exhibit patterns among triples of nodes such as transitivity, balance and clustering (Wasserman and Faust, 1994). For example, in the context of fitting a regression model, the notion of balance would correspond to there generally being a higher-than expected relationship (i.e., a positive residual) between nodes j and k if that between i and j and i and k were also both higher than expected. Such patterns can be quantified with summary statistics such as $\sum_{i\neq j\neq k\neq i}\hat{e}_{i,j}\hat{e}_{i,k}\hat{e}_{j,k}/(n(n-1)(n-2))$, where $\hat{e}_{i,j}$ is a residual from an ordinary least-squares fit of the $y_{i,j}$'s to the $\mathbf{x}_{i,j}$'s.

One way to evaluate the fit of a model in terms of a particular summary statistic is to simulate sociomatrices from the fitted model, and compare the summary statistics computed from the simulated sociomatrices to the value computed from the observed sociomatrix. Such a comparison is made in the pink histogram in the upper-left plot of Figure 3, which shows the posterior predictive distribution of the above-mentioned triadic summary statistic from a Bayesian fit of the SRRM to the trade data. Specifically, the posterior predictive distribution was constructed by simulating 2000 values of the parameters $\{\boldsymbol{\beta}, a_1,\ldots,a_n, b_1,\ldots,b_n,\sigma^2,\rho\}$ from their posterior distribution (using a Markov chain Monte Carlo approximation described in Section 6), then simulating a sociomatrix from the model given by equation (2.3) for each of these 2000 parameter values, and finally computing the triadic summary statistic for each of these 2000 simulated sociomatrices. The sociomatrices simulated from the fitted SRRM consistently exhibit far less third-order dependence than the observed sociomatrix, indicating a lack-of-fit with respect to this summary statistic. The reason why the SRRM is unable to capture this third-order dependence is that each $\hat{e}_{i,j}$ from a sociomatrix simulated from the SRRM will be approximately given by $\hat{e}_{i,j}\approx\hat{a}_i+\hat{b}_j+\epsilon_{i,j}$, where the $\hat{a}_i$'s and $\hat{b}_j$'s approximately sum to zero across nodes and the $\epsilon_{i,j}$'s are mean zero normal random variables. In this case, averages of the form $\sum_{i\neq j\neq k\neq i}(\hat{a}_i+\hat{b}_j+\epsilon_{i,j})(\hat{a}_i+\hat{b}_k+\epsilon_{i,k})(\hat{a}_j+\hat{b}_k+\epsilon_{j,k})/(n(n-1)(n-2))$ will be close to zero. Perhaps more simply, by thinking of the a_i's and b_j's as random effects it is intuitively clear that the SRRM is unable to represent third-order dependence because it assumes that the differences between the $y_{i,j}$'s and $\boldsymbol{\beta}^\top\mathbf{x}_{i,j}$'s are mean-zero Gaussian random variables, and all third-order moments of mean-zero Gaussian random variables are zero.

To model such higher-order network patterns we must go beyond node-specific effects that combine additively. One solution is to include additional nodal effects into the model that combine multiplicatively. For example, for binary and count dyadic variables, Hoff (2005) proposed modeling $y_{i,j}$ as a function of $\boldsymbol{\beta}^\top\mathbf{x}_{i,j}+\mathbf{u}_i^\top\mathbf{u}_j+a_i+b_j$, where each $\mathbf{u}_i$ is a low-dimensional vector of node-specific parameters, or latent features. The effect $\gamma_{i,j}=$

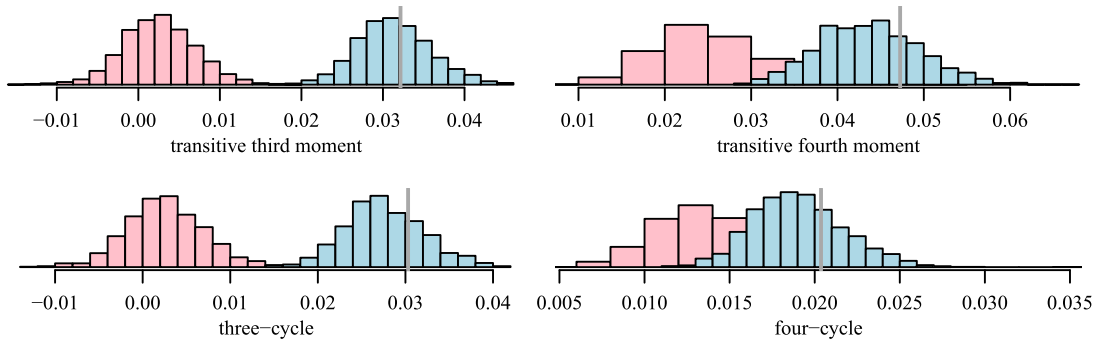

FIG. 3. *Posterior predictive distributions of third- and fourth-order goodness-of-fit statistics. The pink histograms correspond to the SRRM fit, the blue to the AME fit. The observed values of the statistics are given by vertical gray lines.*

$\mathbf{u}_i^\top \mathbf{u}_j$ can capture some forms of triadic dependence because $\gamma_{i,j}\gamma_{i,k}\gamma_{k,j}$ can be nonzero on average across triads, even if the $\mathbf{u}_i$'s sum to zero. To see this, consider the simplest case where the dimension of the vector is one, so the latent feature for each node is just the scalar u_i. In this case $\gamma_{i,j}\gamma_{i,k}\gamma_{j,k} = u_i^2 u_j^2 u_k^2$, which is always nonnegative, and so its across-triad average is strictly positive unless all the u_i's are zero.

A more flexible multiplicative effects model for directed dyadic data is obtained by replacing $\mathbf{u}_i^\top \mathbf{u}_j$ with $\mathbf{u}_i^\top \mathbf{v}_j$, so that $\mathbf{u}_i$ represents latent features of node i as a sender of relations, and $\mathbf{v}_i$ represents features of i as a receiver. Combining such a multiplicative term with the terms in the SRRM yields the following *additive and multiplicative effects* (AME) model for dyadic data:

$$\begin{aligned} &y_{i,j} = \boldsymbol{\beta}^\top \mathbf{x}_{i,j} + \mathbf{u}_i^\top \mathbf{v}_j + a_i + b_j + \epsilon_{i,j} \\ &\{(\epsilon_{i,j}, \epsilon_{j,i}) : i < j\} \sim \text{i.i.d. } N_2\bigl(\mathbf{0}, \sigma^2 \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)\bigr). \end{aligned} \tag{3.1}$$

Specifically, we refer to the model given by (3.1) as a Gaussian AME model, since the observed data are conditionally Gaussian, given $\boldsymbol{\beta}$ and the node-specific additive and multiplicative features. Simpler versions of (3.1) for a variety of types of dyadic variables (including counts, binary variables and undirected relations) have appeared in Hoff (2005, 2008, 2009b).

The matrix form of this model can be expressed as

$$\mathbf{Y} = \mathbf{M} + \mathbf{a}\mathbf{1}^\top + \mathbf{1}\mathbf{b}^\top + \mathbf{U}\mathbf{V}^\top + \mathbf{E},$$

where $m_{i,j} = \boldsymbol{\beta}^\top \mathbf{x}_{i,j}$, $\mathbf{a} = (a_1, \ldots, a_n)$, $\mathbf{b} = (b_1, \ldots, b_n)$ and $\mathbf{U}$ and $\mathbf{V}$ are $n \times r$ matrices with ith rows equal to the r-dimensional latent feature vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ respectively. In matrix form, this model represents the deviations of the sociomatrix $\mathbf{Y}$ from the linear regression model $\mathbf{M}$ as the sum of a rank-1 matrix of row effects, a rank-1 matrix of column effects, a rank-r matrix $\mathbf{U}\mathbf{V}^\top$ and a noise matrix $\mathbf{E}$. Without the regression term and distributional assumptions about the node-specific effects and $\mathbf{E}$, this model is essentially a special case of an *additive main effects, multiplicative interaction* (AMMI) model (Gollob, 1968, Bradu and Gabriel, 1974), a class of matrix models developed in the psychometric and agronomy literature for data arising from two-way layouts. Since sociomatrices have additional structure—the outcomes $y_{i,j}$ and $y_{j,i}$ involves the same pair of nodes—our version of the AMMI model includes the possibility that $\epsilon_{i,j}$ and $\epsilon_{j,i}$ are correlated.

Hoff et al. (2013) proposed a random effects AME model, which in addition to (3.1), specified that

$$\begin{aligned} &(\mathbf{u}_1, \mathbf{v}_1), \ldots, (\mathbf{u}_n, \mathbf{v}_n) \sim \text{i.i.d. } N_{2r}(\mathbf{0}, \Psi), \\ &(a_1, b_1), \ldots, (a_n, b_n) \sim \text{i.i.d. } N_2(\mathbf{0}, \Sigma). \end{aligned} \tag{3.2}$$

One motivation for the random effects model is that it provides a means of shrinking the node-specific effects $\{(a_i, b_i, \mathbf{u}_i, \mathbf{v}_i), i = 1, \ldots, n\}$ to prevent overfitting. Another motivation, discussed in more detail in Section 5.1, is that it provides summaries of certain network dependencies via the global parameters Σ and Ψ.

To illustrate the how the inclusion of multiplicative effects improves model fit, we obtained an approximation to the posterior predictive distribution of the triadic goodness-of-fit statistic $\sum_{i,j,k} \hat{e}_{i,j}\hat{e}_{i,k}\hat{e}_{j,k}/(n(n-1)(n-2))$ under an AME model with two-dimensional multiplicative effects and the same regressors as the SRRM (polity, GDP and geographic distance). The posterior predictive distribution was approximated by simulating values of the parameters $\{\boldsymbol{\beta}, \{a_i, b_i, \mathbf{u}_i, \mathbf{v}_i : i = 1, \ldots, n\}, \sigma^2, \rho\}$ from their posterior distribution, simulating a sociomatrix from (3.1) for each set of simulated parameters, and computing the goodness-of-fit statistic for each simulated sociomatrix.

A histogram of this posterior predictive distribution is given in light blue in the upper-left plot of Figure 3, along with that of the SRRM fit (also included in the figure are posterior predictive comparisons of three other statistics, which will be discussed in Section 5.1). The posterior predictive distribution obtained under the AME fit is roughly centered around the observed value of the statistic indicating that, unlike the SRRM, the AME model with two-dimensional multiplicative effects is able to describe this third-order residual dependency in the trade data. Had this not been the case, we might have considered using a larger dimension for the multiplicative effects. In this manner, goodness-of-fit plots such as these can provide a heuristic for selecting the dimension of the multiplicative effects. Finally, parameter estimates and standard errors for the regression coefficients in this AME model are given in the third column of Table 1. Parameter estimates are slightly smaller than those of the SRRM, but the main conclusions are the same.

As shown with the goodness-of-fit statistics, the multiplicative effects can be viewed as a means to quantify global higher-order dependence patterns in a sociomatrix. However, these effects can also be interpreted as representing omitted regression variables or uncovering group structure among the nodes. This interpretation is based on the observation that the strength or presence of ties between nodes is often related to similarities of node-level attributes. For example, suppose for each node i that x_i is the indicator that i is a member of a particular group or has a particular trait. Then $x_i x_j$ is the indicator that i and j are co-members of this group, and this fact may have some effect on their relationship $y_{i,j}$. A positive association between $x_i x_j$ and $y_{i,j}$ is referred to as homophily, and a negative association as anti-homophily. Quantifying homophily on an observed attribute can be done with a SRRM by creating a dyadic regressor $x_{d,i,j}$ from a nodal regressor x_i through multiplication ($x_{d,i,j} =$

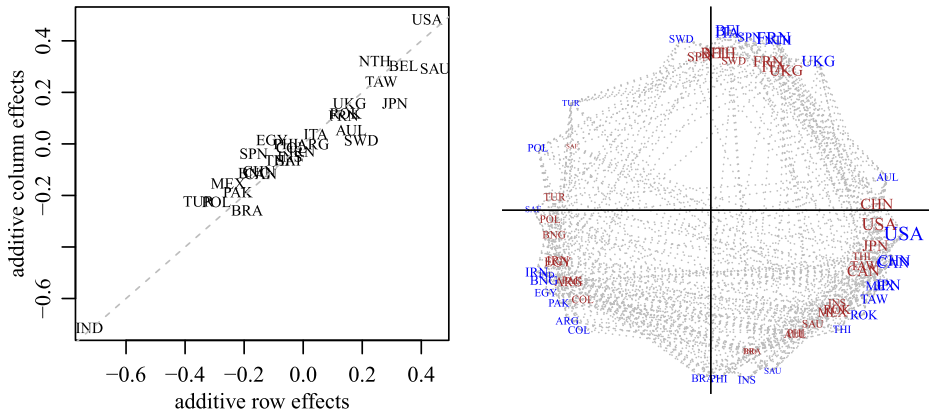

FIG. 4. *Estimates of node-specific effects. The left panel gives additive row effects versus additive column effects. The plot on the right gives estimates of* $\mathbf{u}_i$ *in red and* $\mathbf{v}_i$ *in blue for each country* $i = 1, \ldots, n$. *The country names indicate the direction of these vectors, and the size of the plotting text indicates their magnitude. A dashed line is drawn between an export-import pair if their trade flow is larger than expected based on the other terms in the model.*

$x_i x_j$) or some other operation. However, the possibility that not all relevant nodal attributes are included in a network dataset motivates inclusion of the multiplicative term $\mathbf{u}_i^\top \mathbf{v}_j$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ represent unobserved latent features of node i as a sender and receiver of relations, respectively.

These latent features may be estimated and examined to highlight additional structure in the data beyond that explained by the SRRM. For example, estimates of the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s of the rank-2 AME fit to the trade data are displayed in Figure 4. Recall that this model includes polity, GDP and geographic distance as regressors, in addition to the additive effects and multiplicative latent features. The interpretation of the multiplicative factors is that if $\mathbf{u}_i$ and $\mathbf{v}_j$ are large and in the same direction, then nodes i and j tend to have observed trade flows larger than $\boldsymbol{\beta}^\top \mathbf{x}_{i,j} + a_i + b_j$, that is, larger than what is predicted by the additive effects alone. As can be seen from the figure, the estimates of the latent features from these data highlight some geographically related clustering of nodes, in particular, a cluster of Pacific rim countries and a cluster of mostly European countries. These are patterns that, while related to geographic distance, are not well represented by a single linear relationship between log-trade and log-distance in the regression model.

So far the AME framework has been described in the context of directed network relations. An AME framework for undirected relations is simpler, as in this case there is no need for separate sender and receiver effects. The undirected AME model reduces to

$$y_{i,j} = \boldsymbol{\beta}^\top \mathbf{x}_{i,j} + \mathbf{u}_i^\top \Lambda \mathbf{u}_j + a_i + a_j + \epsilon_{i,j}$$

for undirected pairs $\{(i, j) : 1 \le i < j \le n\}$, where $\{\epsilon_{i,j} : 1 \le i < j \le n\} \sim$ i.i.d. $N(0, \sigma_\epsilon^2)$. The multiplicative term $\mathbf{u}_i^\top \Lambda \mathbf{u}_j$ now includes as a parameter a diagonal matrix Λ of "eigenvalues," whose nonzero entries can be either positive or negative. This generalizes the role of the multiplicative effects in the directed AME model in the sense that, in the directed case the matrix $\mathbf{U}\mathbf{V}^\top$ can be any rank-r matrix, whereas in the undirected case the matrix $\mathbf{U}\Lambda\mathbf{U}^\top$ can be any symmetric rank-r matrix. Interpretation of the parameter Λ in terms of describing homophily and anti-homophily on unobserved attributes is described in Hoff (2008).

4. TRANSFORMATION MODELS FOR NON-GAUSSIAN NETWORKS

On their original scale, many dyadic variables are not well represented by a model with a Gaussian error term. In some cases, such as with the trade data, a dyadic variable can be transformed so that the Gaussian AME model is reasonable. In other cases, such as with binary, ordinal, discrete or sparse variables, no such transformation is available. Examples of such data include measures of friendship that are binary (not friends/friends) or ordinal (dislike/neutral/like), discrete counts of conflictual events between countries, or durations of phone calls between pairs of individuals. In this section we describe extensions of the Gaussian AME model to accommodate ordinal dyadic variables, which for our purposes includes variables for which the possible values can be put in some meaningful order. This includes discrete variables (such as binary indicators or counts), ordered qualitative variables (such as low/medium/high), and even continuous variables. The extensions are based on latent variable representations of probit and ordinal probit regression models.

4.1 Binary and Ordinal Network Data

Let $\mathbf{S}$ be the observed sociomatrix for a dyadic variable $s_{i,j}$. The simplest type of ordinal dyadic variable is a binary variable indicating the presence of some type of

relationship between i and j, so that $s_{i,j} = 0$ or 1 depending on whether a social link is absent or present, respectively. One model for quantifying the association between such a binary variable and other variables is probit regression, which models the probability of a link between i and j as $\Phi(\boldsymbol{\beta}^\top \mathbf{x}_{i,j})$, where Φ is the standard normal CDF. As is well known, the probit regression model has a latent variable representation in which $s_{i,j}$ is the binary indicator that some latent normal random variable, say $y_{i,j} \sim N(\boldsymbol{\beta}^\top \mathbf{x}_{i,j}, 1)$, is greater than zero (Albert and Chib, 1993). An ordinary probit regression model corresponds to the $y_{i,j}$'s being independent, which is generally an inappropriate assumption for network data. However, a model for binary data that does capture the types of network dependencies discussed in the previous section, such as row and column covariance, dyadic correlation, and higher-order dependence, is an AME model for the latent $y_{i,j}$'s:

$$\begin{aligned} y_{i,j} &= \boldsymbol{\beta}^\top \mathbf{x}_{i,j} + \mathbf{u}_i^\top \mathbf{v}_j + a_i + b_j + \epsilon_{i,j}, \\ s_{i,j} &= g(y_{i,j}), \end{aligned} \tag{4.1}$$

where the a_i's b_i's and $\epsilon_{i,j}$'s follow the SRM covariance model and $g(y)$ is the binary indicator that $y > 0$. Without the multiplicative term $\mathbf{u}_i^\top \mathbf{v}_j$, this is basically a generalized linear mixed effects model. With the multiplicative term but without the SRM covariance structure, this model is a type of generalized bilinear regression (Gabriel, 1998). Including both the multiplicative term and the SRM covariance structure yields a regression model for binary social network data that can represent many of the types of patterns seen in network data.

This probit AME model for binary data extends in a natural way to accommodate ordinal data with more than two levels. As with binary data, we model the observed sociomatrix $\mathbf{S}$ as being a function of a latent sociomatrix $\mathbf{Y}$ that follows a Gaussian AME distribution. Specifically, the model is the same as in equation (4.1) but with g being a nondecreasing function. Such a model is a type of Gaussian transformation model (Bickel and Ritov, 1997).

One approach to estimation for these models is as follows: For both the probit and ordinal probit models, observation of $\mathbf{S}$ tells us that $\mathbf{Y}$ lies in a certain set, say $\mathbf{Y} \in C(\mathbf{S})$. For the binary probit model, this set is simply given by $C(\mathbf{S}) = \{\mathbf{Y} \in \mathbb{R}^{n\times n} : \text{sign}(y_{i,j}) = \text{sign}(2s_{i,j} - 1)\}$, that is, $s_{i,j} = 1$ implies $y_{i,j} > 0$ and $s_{i,j} = 0$ implies $y_{i,j} < 0$. For the ordinal probit model, since g is nondecreasing we have $C(\mathbf{S}) = \{\mathbf{Y} \in \mathbb{R}^{n\times n} : \max_{i'j'}\{y_{i',j'} : s_{i',j'} < s_{i,j}\} < y_{i,j} < \min_{i'j'}\{y_{i',j'} : s_{i,j} < s_{i'j'}\}\}$. A likelihood based on the knowledge that $\mathbf{Y} \in C(\mathbf{S})$ is given by $L(\boldsymbol{\theta}) = \Pr(\mathbf{Y} \in C(\mathbf{S})|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the parameters in the Gaussian AME model for $\mathbf{Y}$. While a closed form expression for this likelihood is unavailable, a Bayesian approach to estimation and inference is feasible via Gibbs sampling by iteratively simulating $\boldsymbol{\theta}$ from its full conditional distribution given $\mathbf{Y}$, then simulating $\mathbf{Y}$ from its conditional distribution given $\boldsymbol{\theta}$ but constrained to lie in $C(\mathbf{S})$. More details are presented in Section 6.

4.2 Censored and Ranked Nomination Data

Data on human social networks are often obtained by asking participants in a study to name and rank a fixed number of people with whom they are friends. Such a survey method is called a *fixed ranked nomination* (FRN) scheme, and is used in studies of institutions such as schools or businesses. For example, the National Longitudinal Study of Adolescent Health (Harris et al., 2009) asked middle and high school students to nominate and rank up to five members of the same sex as friends, and five members of the opposite sex as friends.

Data obtained from FRN schemes are similar to ordinal data, in that the ranks of a person's friends may be viewed as an ordinal response. However, FRN data are also censored in a complicated way. Consider a study in which people are asked to name and rank up to and including their top five friends. If person i nominates five people but doesn't nominate person j, then $s_{i,j}$ is censored: The data cannot tell us whether j is i's sixth best friend, or whether j is not liked by i at all. On the other hand, if person i nominates four people as friends but could have nominated five, then person i's data are not censored—the absence of a nomination by i of j indicates that i does not consider j a friend.

A likelihood-based approach to modeling FRN data using an AME model was developed in Hoff et al. (2013). Similar to the approach for ordinal dyadic data described above, this methodology treats the observed ranked outcomes $\mathbf{S}$ as a function of an underlying continuous sociomatrix $\mathbf{Y}$ of affinities that is generated from an AME model. Letting m be the maximum number of nominations allowed, and coding $s_{i,j} \in \{m, m-1, \ldots, 1, 0\}$ so that $s_{i,j} = m$ indicates that j is i's most liked friend, the FRN likelihood is derived from the following constraints that the observed ranks $\mathbf{S}$ tell us about the underlying dyadic variables $\mathbf{Y}$:

$$s_{i,j} > 0 \Rightarrow y_{i,j} > 0, \tag{4.2}$$

$$s_{i,j} > s_{i,k} \Rightarrow y_{i,j} > y_{i,k}, \tag{4.3}$$

$$s_{i,j} = 0 \quad \text{and} \quad d_i < m \Rightarrow y_{i,j} \le 0. \tag{4.4}$$

Constraint (4.2) indicates that if i ranks j, then i has a positive relation with j ($y_{i,j} > 0$), and constraint (4.3) indicates that a higher rank corresponds to a more positive relation. Letting $d_i \in \{0, \ldots, m\}$ be the number of people that i ranks, constraint (4.4) indicates that if i could have made additional friendship nominations but chose not to nominate j, they then do not consider j a friend. However, if $s_{i,j} = 0$ but $d_i = m$ then person i's unranked relationships are censored, and so $y_{i,j}$ could be positive even

though $s_{i,j} = 0$. In this case, all that is known about $y_{i,j}$ is that it is less than $y_{i,k}$ for any person k ranked by i. In summary, observation of $\mathbf{S}$ tells us that $\mathbf{Y} \in C(\mathbf{S})$ where $C(\mathbf{S})$ is defined by conditions (4.2)–(4.4). As with the probit and ordinal AME models, Bayesian inference for this transformation model can proceed by iteratively simulating values of the model parameters and the unknown values of $\mathbf{Y}$ from their full conditional distributions.

5. RELATIONSHIPS TO OTHER MODELS

Two popular categories of statistical network models are exponentially parameterized random graph models (ERGMs) and latent variables models. One appealing feature of ERGMs is that models for evaluating specific global network patterns of interest can be constructed, in principle, simply by including an appropriate sufficient statistic in the model specification. In contrast, much of the appeal of latent variable models stems from their description of local, micro-level patterns of relationships among specific nodes. The AME class of models can characterize local patterns via estimates of the node-specific effects $\{(a_i, b_i, \mathbf{u}_i, \mathbf{v}_i) : i = 1, \ldots, n\}$, and global patterns via the parameters $\{\boldsymbol{\beta}, \Sigma, \Psi, \sigma^2, \rho\}$. In the next two subsections we compare the AME class to ERGMs and to some latent variable models in terms of how these types of patterns are represented.

5.1 Comparisons to ERGMs

An ERGM is a probability model for a binary sociomatrix that includes densities of the form $p(\mathbf{Y}) = c(\boldsymbol{\theta}) \exp(\boldsymbol{\theta} \cdot \mathbf{t}(\mathbf{Y}))$, where $\mathbf{t}(\mathbf{Y})$ is a vector of sufficient statistics and $\boldsymbol{\theta}$ is a parameter to be estimated. Early applications of ERGMs used a small number of simple sufficient statistics, often much smaller than the number of nodes, and in this sense were typically used to describe "global" patterns in the data. An exception to this was the not-infrequent inclusion of out- and in-degree statistics that can characterize the differential sociability and popularity of the nodes. For example, one of the first ERGMs to be widely used and studied was the "p_1" model (Holland and Leinhardt, 1981) with density

$$p(\mathbf{Y}) \propto \exp\Bigg(\mu \sum_{i,j} y_{i,j} + \sum_i \Bigg(a_i \sum_j y_{i,j} + b_i \sum_j y_{j,i}\Bigg) + \rho \sum_{i,j} y_{i,j} y_{j,i}\Bigg),$$

which includes as sufficient statistics the total number of ties $\sum_{i,j} y_{i,j}$, the number of reciprocated ties $\sum_{i,j} y_{i,j} y_{j,i}$ and the in- and out-degrees $\{\sum_j y_{i,j}, \sum_j y_{j,i}, i = 1, \ldots, n\}$. The parameters in this model represent roughly the same data features as they do in the SRM: an overall mean of the relations (μ), heterogeneity in row and column means (the a_i's and b_i's) and dyadic correlation (ρ). Similarities are also found between the SRRM and the "p_2" model developed by van Duijn, Snijders and Zijlstra (2004). The p_2 model extends the p_1 model by including regressors (as does the SRRM), and additionally treats the node-level parameters a_i and b_i as potentially correlated random effects (as do the SRM and SRRM).

Holland and Leinhardt (1981) concede that the p_1 model is of limited utility due to its inability to describe more complex forms of dependency such as transitivity or clustering. While inclusion of appropriate regressors, either in a p_2 model or SRRM, can represent some degree of higher-order dependency, often such models still exhibit lack-of-fit, and so more complex models are often desired. The ERGM approach to describing higher-order dependencies is to include additional sufficient statistics, such as the number of triangles observed in the graph, or the number of cycles of various lengths. This approach has great appeal to researchers interested in testing social theories of tie formation, as one can compare the fits of two models, one with and one without a statistic that represents a particular network dependency of interest.

However, simultaneous inclusion of certain types of statistics, such as the number of triangles in a graph, along with statistics that intuitively represent degree heterogeneity, can lead to model degeneracy (Handcock, 2003). Two approaches to ameliorating this problem include constraining the parameter space away from problematic values, and finding alternative summary statistics to represent transitivity. The former approach has been studied by Hunter and Handcock (2006) and Snijders et al. (2006), who considered curved exponential families models, that is, models with densities of the form $p(\mathbf{Y}) = c(\psi) \exp(\boldsymbol{\theta}(\psi) \cdot \mathbf{t}(\mathbf{Y}))$, where now the value of $\boldsymbol{\theta}$ is constrained to lie on a curve indexed by the parameter ψ. These articles also propose new transitivity statistics for inclusion in ERGMs, and Hunter, Goodreau and Handcock (2008) show empirically on several datasets that models based on these alternative statistics avoid some of the degeneracy problems created by the use of more traditional network statistics. Similarly, Schweinberger and Handcock (2015) replace problematic transitivity statistics with a latent blockmodel structure, resulting in improved model stability. A comprehensive overview of ERGMs with case studies is given by Lusher, Koskinen and Robins (2013) and a recent review of ERGM research is given in Schweinberger et al. (2017).

In contrast, the AME approach to representing complex patterns in a sociomatrix is with the low-rank matrix $\mathbf{UV}^\top$. From a fixed effects perspective, an AME model can fit the observed sociomatrix with an arbitrary degree of precision, if the dimension of the latent features is sufficiently large. This is because an $n \times n$ matrix $\mathbf{Y}$ can be approximated to an arbitrary degree of precision by a product $\mathbf{UV}^\top$ of two $n \times r$ matrices $\mathbf{U}$ and $\mathbf{V}$, by choosing r to be sufficiently large (Eckart and Young, 1936). Of course,

what we hope for is that only a small value of r is necessary to achieve a good fit. It has been my experience that many networks can be well approximated in this sense by low-dimensional latent variable models, such as AME models (sometimes even perfectly—see, e.g., Section 4.2 of Hoff, Raftery and Handcock, 2002). In these cases, an AME model provides a model-based low-dimensional representation of the observed network.

However, the types of global network dependencies that can be represented by the Gaussian random effects model (3.2) for the nodal effects are more limited. Specifically, the $2r \times 2r$ covariance matrix covariance matrix Ψ of $(\mathbf{u}_i, \mathbf{v}_i)$ can encode some types of third-order and higher dependencies, but not others. For example, consider third-order moments of $e_{i,j} = y_{i,j} - \boldsymbol{\beta}^\top \mathbf{x}_{i,j}$. Under the AME model we have $e_{i,j} = \mathbf{u}_i^\top \mathbf{v}_j + a_i + b_j + \epsilon_{i,j}$. Because $a_i + b_j + \epsilon_{i,j}$ is a mean zero Gaussian random variable under this model, and all third-order moments of mean-zero Gaussian random variables are zero, then the third-order moments of the $e_{i,j}$'s will be the same as those of the $\gamma_{i,j}$'s, where $\gamma_{i,j} = \mathbf{u}_i^\top \mathbf{v}_j$. Therefore, under the random effects model (3.2) the third-order moment $\mathrm{E}[e_{i,j} e_{j,k} e_{k,i}]$ corresponding to a three-cycle is given by

$$\begin{aligned}\mathrm{E}[e_{i,j} e_{j,k} e_{k,i}] &= \mathrm{E}[\gamma_{i,j} \gamma_{j,k} \gamma_{k,i}] \\ &= \mathrm{E}\big[\mathbf{u}_i^\top \mathbf{v}_j \mathbf{u}_j^\top \mathbf{v}_k \mathbf{u}_k^\top \mathbf{v}_i\big] \\ &= \mathrm{tr}\big(\mathrm{E}\big[\mathbf{v}_i \mathbf{u}_i^\top \mathbf{v}_j \mathbf{u}_j^\top \mathbf{v}_k \mathbf{u}_k^\top\big]\big) = \mathrm{tr}\big(\Psi_{uv}^3\big),\end{aligned}$$

where $\Psi_{uv} = \mathrm{E}[\mathbf{u}_i \mathbf{v}_i^\top]$. This moment can be positive, negative or zero depending on the values of Ψ_{uv}. A similar calculation shows that the third-order transitive moment $\mathrm{E}[e_{i,j} e_{i,k} e_{j,k}]$ is given by $\mathrm{tr}(\Psi_{uu} \Psi_{vv} \Psi_{uv})$ where $\Psi_{uu} = \mathrm{E}[\mathbf{u}_i \mathbf{u}_i^\top]$ and $\Psi_{vv} = \mathrm{E}[\mathbf{v}_i \mathbf{v}_i^\top]$. However, not all higher-order moments can be represented by the random effects model for the multiplicative effects, and even when they can, the parameter Ψ describing the moments can be hard to interpret. For example, consider a fourth-order moment $\mathrm{E}[\gamma_{i,j} \gamma_{j,k} \gamma_{k,l} \gamma_{l,i}]$ corresponding to the four-cycle $i \to j \to k \to l \to i$. Straightforward calculations show that this moment is

$$\mathrm{E}[\gamma_{i,j} \gamma_{j,k} \gamma_{k,l} \gamma_{l,i}] = \mathrm{tr}\big(\Psi_{uv}^4\big),$$

and that more generally, the kth order moment corresponding to the k-cycle $i_1 \to \cdots \to i_k \to i_1$ is $\mathrm{tr}(\Psi_{uv}^k)$. Depending on the dimension r of the multiplicative effects, it might not be possible for all such kth order moments to be separately estimated. For example, if $r = 1$ then $\mathrm{E}[\gamma_{i,j} \gamma_{j,k} \gamma_{k,i}] = \sigma_{uv}^3$ and $\mathrm{E}[\gamma_{i,j} \gamma_{j,k} \gamma_{k,l} \gamma_{l,i}] = \sigma_{uv}^4$. These moments are not separately estimable since they are both completely determined by the single parameter σ_{uv}. To separately estimate such moments, the dimension r must be increased, but calculation of the value of r that will permit separate estimation of moments corresponding to k-cycles for a range of k-values is very tricky.

Furthermore, some types of higher-order moments are not estimable at all by a Gaussian multiplicative random effects model. For example, consider the fourth-order transitive moment $\mathrm{E}[e_{i,j} e_{i,k} e_{j,k} e_{k,j}]$, which might be of interest in evaluating a theory of how the relations $y_{j,k}, y_{k,j}$ between two nodes j and k might be associated with their relations $y_{i,j}, y_{i,k}$ to a third party, node i. Calculations similar to those done above show that $\mathrm{E}[\gamma_{i,j} \gamma_{i,k} \gamma_{j,k} \gamma_{k,j}]$ is zero, and so the corresponding moment $\mathrm{E}[e_{i,j} e_{i,k} e_{j,k} e_{k,j}]$ will be determined by Σ, and therefore confounded with the covariance of the additive effects.

However, we emphasize that these limitations are a result of the limited ability of the Gaussian random effects model to summarize the heterogeneity of the node specific effects $\{(a_i, b_i, \mathbf{u}_i, \mathbf{v}_i) : i = 1, \ldots, n\}$, and are not limitations of the ability of these nodal effects themselves to represent the sociomatrix. The range of network patterns that can be described by these effects is much broader than the range of patterns that can be descried by the simple random effects model that summarizes the effects. As evidence of this, we refer back to Figure 3, which compares the AME and SRRM models in terms of four goodness-of-fit statistics corresponding to the four third- and fourth-order moments described above. The SRRM exhibits lack of fit in terms of all four of these statistics, whereas the AME model does not exhibit any substantial lack of fit. In particular, the AME model improves the fit in terms of the transitive fourth-order moment, even though this pattern is not representable by the random effects model for the multiplicative effects, as described in the previous paragraph.

To summarize, an estimate of the multiplicative effects matrix $\mathbf{U}\mathbf{V}^\top$ in an AME model provides a reduced-rank representation of the sociomatrix $\mathbf{Y}$. As the dimension of the multiplicative effects increases, so does the accuracy of the representation. Furthermore, an estimate of Ψ provides a summary of the across-node heterogeneity of the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s, and therefore, also a summary of certain network dependencies induced by these effects. But since Ψ is an incomplete summary of the across-node heterogeneity of the multiplicative effects (it only describes the covariance of these effects), it can only provide an limited summary of the potential network dependencies that are present in the sociomatrix. Additionally, the correspondence between Ψ and the higher-order dependencies it describes is somewhat opaque. For this reason, in situations where the primary goal of a data analysis is to evaluate specific types of higher-order network dependencies, an approach using ERGMs will be more straightforward.

5.2 Comparison to Other Latent Variable Models

While estimates of the parameter Ψ in the random effects model (3.2) can represent certain higher-order network dependencies, estimates of the latent features in the

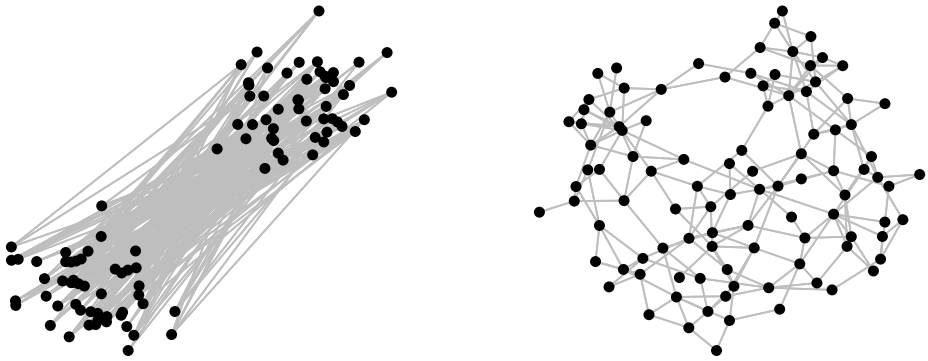

FIG. 5. *Two hypothetical networks. The network on the left can be represented by two groups of stochastically equivalent nodes. The network on the right can be represented by an embedding of the nodes in two-dimensional Euclidean space.*

AME model can also be interpreted locally at the micro-level, in that $\mathbf{u}_i$ and $\mathbf{v}_i$ describe latent features of node i as a sender and receiver of ties. Estimates of the features (such as those displayed in Figure 4) can be used to identify interesting nodes, assist with visualization of network patterns, or be used as an input to other data analysis methods, such as clustering (Rohe, Chatterjee and Yu, 2011). Other popular nonadditive latent variable models for network data include the stochastic blockmodel (Nowicki and Snijders, 2001) and the latent distance model (Hoff, Raftery and Handcock, 2002). The blockmodel assumes each node belongs to an unobserved latent class or "block," and that the relations between two nodes are determined (statistically) by their block memberships. This model is based on the assumption of *stochastic equivalence*, that is, the assumption that the nodes can be divided into groups such that members of the same group have the same distribution of relationships to other nodes. This basic blockmodel has been extended in a number of ways, such as allowing for node-specific degree heterogeneity (Karrer and Newman, 2011, Zhao, Levina and Zhu, 2012), and to allow for nodes to be members of multiple groups (Airoldi et al., 2008). In contrast, the latent distance model assumes each node has some unobserved location in a latent Euclidean "social space," and that the strength of a relation between two nodes is decreasing in the distance between them in this space. This model provides a compact representation of certain patterns seen in social networks such as transitivity and community, that is, the existence subgroups of nodes with strong within-group relations. Like the basic blockmodel, the distance model has been extended to allow for degree heterogeneity (Krivitsky et al., 2009), and has been studied theoretically by Rastelli, Friel and Raftery (2016). Extensions to network embeddings in non-Euclidean metric spaces have been considered by Asta and Shalizi (2015), and embeddings in ultrametric spaces have been proposed by Schweinberger and Snijders (2003).

Figure 5 displays two hypothetical symmetric networks, each one of which can be well represented by one of these two latent variable models. The network on the left can be well represented by a two-group stochastic blockmodel in which the within-group density of ties is lower than the between-group density. Such a network is not representable by a latent distance model because in such a model, stochastic equivalence of two nodes is confounded with the expected strength of their relationship: In a latent distance model, two nodes are stochastically equivalent if they are in the same location in the social space. However, if they are in the same location, then the distance between them is zero and so their expected relationship is strong. As such, networks where stochastically equivalent nodes have weak ties will not be well represented by a latent distance model. Conversely, the network displayed on the right side of Figure 5 is very well represented by a two-dimensional latent distance model in which the probability of a tie between two nodes is decreasing in the distance between them. However, representation of this network by a blockmodel would require a large number of blocks (e.g., one block in each subregion of the space), none of which would be particularly cohesive or distinguishable from neighboring blocks.

In contrast to these two extreme networks, real networks exhibit combinations of stochastic equivalence and transitivity in varying amounts. Inference based on either a blockmodel or a distance model will often provide only an incomplete description of the heterogeneity across nodes in terms of how they form ties to others. However, as shown in Hoff (2008), latent variable models based on multiplicative effects (such as AME models) can represent both of these types of network patterns, and therefore provide a generalization of both the stochastic blockmodel and the latent distance model. To explain this generalization, we consider the simple case of an undirected dyadic variable so that the sociomatrix is symmetric. Each of the three types of latent variable models may be written abstractly as $y_{i,j} \sim m_{i,j} + \alpha(\mathbf{u}_i, \mathbf{u}_j)$ where α is some function of the node-specific latent variables $\mathbf{u}_1, \ldots, \mathbf{u}_n$, $m_{i,j}$ consists of any other terms in the model (such as a regression term or additive effects), and "$y \sim x$" means

that the distribution of y is stochastically increasing in x. The three latent variable models correspond to the following three specifications of the function α:

Stochastic blockmodel: $\alpha(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^\top \Theta \mathbf{u}_j$, where $\mathbf{u}_i \in \mathbb{R}^r$ is a standard basis vector indicating block membership, and Θ is $r \times r$ symmetric.

Latent distance model: $\alpha(\mathbf{u}_i, \mathbf{u}_j) = -|\mathbf{u}_i - \mathbf{u}_j|$, where $\mathbf{u}_i \in \mathbb{R}^r$.

Multiplicative effects model: $\alpha(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^\top \Lambda \mathbf{u}_j$, where $\mathbf{u}_i \in \mathbb{R}^r$ and Λ is an $r \times r$ diagonal matrix.

Hoff (2008) referred to the symmetric multiplicative effects model as an "eigenmodel," as the matrix $\mathbf{U}\Lambda\mathbf{U}^\top$ resembles an eigendecomposition of a rank-r matrix. Note that as the $\mathbf{u}_i$'s range over r-dimensional Euclidean space, and Λ ranges over all $r \times r$ diagonal matrices, the matrix $\mathbf{U}\Lambda\mathbf{U}^\top$ ranges over the space of all symmetric rank-r matrices. Similarly, for the asymmetric AME models discussed elsewhere in this article, as the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s range over r-dimensional space, the multiplicative term $\mathbf{U}\mathbf{V}^\top$ ranges over the space of all $n \times n$ rank-r matrices.

To compare these models, we compare the sets of matrices that are representable by their latent variables. Let $\mathcal{S}_n$ be the set of $n \times n$ symmetric matrices, and let

$$\mathcal{B}_r = \{\mathbf{S} \in \mathcal{S}_n : s_{i,j} = \mathbf{u}_i^\top \Theta \mathbf{u}_j,\ \mathbf{u}_i \text{ a standard basis vector}, \Theta \in \mathbb{R}^{r\times r} \text{ symmetric}\};$$

$$\mathcal{D}_r = \{\mathbf{S} \in \mathcal{S}_n : s_{i,j} = -|\mathbf{u}_i - \mathbf{u}_j|, \mathbf{u}_i \in \mathbb{R}^r\};$$

$$\mathcal{E}_r = \{\mathbf{S} \in \mathcal{S}_n : s_{i,j} = \mathbf{u}_i^T \Lambda \mathbf{u}_j, \mathbf{u}_i \in \mathbb{R}^r, \Lambda \text{ a } r \times r \text{ diagonal matrix}\}.$$

In other words, $\mathcal{B}_r$ is the set of matrices expressible as a r-dimensional blockmodel, and $\mathcal{D}_r$ and $\mathcal{E}_r$ are defined similarly for the latent distance and multiplicative effects models, respectively. Hoff (2008) showed the following:

1. $\mathcal{E}_r$ generalizes $\mathcal{B}_r$;
2. $\mathcal{E}_{r+1}$ weakly generalizes $\mathcal{D}_r$;
3. $\mathcal{D}_r$ does not weakly generalize $\mathcal{E}_1$.

Result 1 means that $\mathcal{B}_r$ is a proper subset of $\mathcal{E}_r$ unless $r \geq n$. This is because the matrix $\mathbf{S}$ corresponding to an r-group blockmodel is of rank r or less, and $\mathcal{E}_r$ includes all such matrices. Result 2 means that for any $\mathbf{S} \in \mathcal{D}_r$, there exists an $\tilde{\mathbf{S}} \in \mathcal{E}_{r+1}$ whose elements are a monotonic transformation of those of $\mathbf{S}$, that is, have a numerical order that matches that of the elements of $\mathbf{S}$. From a practical point of view, this means that if an r-dimensional latent distance model fits the data reasonably well, then there exists a link function and a set of $(r+1)$-dimensional multiplicative effects that represents the data equally well. Finally, result 3 says that there exist rank-1 matrices $\mathbf{S}$, expressible via one-dimensional multiplicative effects, that cannot be order-matched by a distance model of *any* dimension. Taken together, these results imply that multiplicative effects models can represent both the types of network patterns representable by stochastic blockmodels and those representable by latent distance models, and so is a more general and flexible class of models than either of these two other latent variable models. See Hoff (2008) for more details and numerical examples.

6. INFERENCE VIA POSTERIOR APPROXIMATION

While maximum likelihood estimation for a Gaussian AME model is feasible, it is quite challenging for binary, ordinal and other AME transformation models because the likelihoods involve intractable integrals arising from the combination of the transformation and the dependencies induced by the SRM. However, reasonably standard Gibbs sampling algorithms can be constructed to provide Bayesian inference for a wide variety of AME network models. We first construct a Gibbs sampler for Gaussian SRRMs, then extend the sampler to accommodate Gaussian AME models, and finally extend the algorithm to fit AME transformation models. These algorithms are implemented in the R package `amen` (Hoff et al., 2012), which includes an R vignette (Hoff, 2015) with several example analyses of binary, ordinal and continuous network data.

6.1 Gibbs Sampling for the SRRM

The unknown quantities in the Gaussian SRRM include the parameters $\boldsymbol{\beta}$, Σ, σ^2, and ρ, and the random effects $\mathbf{a}$ and $\mathbf{b}$. Approximation of the posterior distribution of these quantities is facilitated by using a $N_p(\boldsymbol{\beta}_0, \mathbf{Q}_0^{-1})$ prior distribution for $\boldsymbol{\beta}$, a gamma$(\nu_0/2, \nu_0\sigma_0^2/2)$ prior distribution for $1/\sigma^2$ and a Wishart$(\Sigma_0^{-1}/\eta_0, \eta_0)$ prior distribution for Σ^{-1}. A Gibbs sampler proceeds by iteratively simulating the values of the unknown quantities from their conditional distributions, thereby generating a Markov chain having a stationary distribution equal to the target posterior distribution. Values simulated from this Markov chain can be used to approximate a variety of posterior quantities of interest. Given starting values of the unknown quantities, the algorithm proceeds by iterating the following steps:

1. Simulate $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}\}$ given $\mathbf{Y}$, Σ, σ^2, ρ;
2. Simulate σ^2 given $\mathbf{Y}$, $\boldsymbol{\beta}$, $\mathbf{a}$, $\mathbf{b}$, ρ;
3. Simulate ρ given $\mathbf{Y}$, $\boldsymbol{\beta}$, $\mathbf{a}$, $\mathbf{b}$, σ^2;
4. Simulate Σ given $\mathbf{a}$, $\mathbf{b}$;
5. Simulate missing values of $\mathbf{Y}$ given $\boldsymbol{\beta}$, $\mathbf{a}$, $\mathbf{b}$, σ^2, ρ and observed values of $\mathbf{Y}$.

We include the last step because, while sociomatrices typically have undefined diagonals, the calculations below make use of matrix operations that are only defined on matrices with no missing values. By treating the diagonal values as missing at random, the fact that they are

undefined will not affect the posterior distribution. Additionally, this step permits imputation of other dyadic outcomes that are missing at random.

Steps 2 through 5 are relatively standard. We discuss implementation of these steps before deriving the full conditional distribution of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}\}$. To implement steps 2 and 3, consider the stochastic representation of the SRRM as

$$\mathbf{Y} = \mathbf{M}(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{a}\mathbf{1}^\top + \mathbf{1}\mathbf{b}^\top + \mathbf{E}, \tag{6.1}$$

where $\mathbf{E} = c\mathbf{Z} + d\mathbf{Z}^\top$, with $\mathbf{Z} \sim N_{n\times n}(\mathbf{0}, \mathbf{I})$, $c = \sigma\{(1+\rho)^{1/2} + (1-\rho)^{1/2}\}/2$ and $d = \sigma\{(1+\rho)^{1/2} - (1-\rho)^{1/2}\}/2$. Then $\mathbf{E}$ is a mean-zero Gaussian matrix with $\mathrm{Var}[\binom{e_{i,j}}{e_{j,i}}] = \sigma^2\left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right) \equiv \Sigma_e$ and $\mathrm{Var}[e_{i,i}] = \sigma^2(1+\rho)$, with the elements of $\mathbf{E}$ being otherwise independent. Now given $\boldsymbol{\beta}$, $\mathbf{a}$ and $\mathbf{b}$, construct $\mathbf{E} = \mathbf{Y} - (\mathbf{M}(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{a}\mathbf{1}^\top + \mathbf{1}\mathbf{b}^\top)$. As a function of σ^2 and ρ, the density of $\mathbf{E}$ is proportional to

$$\begin{aligned}&(\sigma^2)^{-n^2/2}(1-\rho^2)^{-\binom{n}{2}/2}(1+\rho)^{-n/2}\\ &\quad\times \exp\{-(SS_1 + SS_2)/[2\sigma^2]\},\end{aligned}$$

where $SS_1 = \sum_{i<j}\binom{e_{i,j}}{e_{j,i}}^\top\left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)^{-1}\binom{e_{i,j}}{e_{j,i}}\}$ and $SS_2 = \sum_{i=1}^n e_{i,i}^2/(1+\rho)$. The full conditional distribution of $1/\sigma^2$ is therefore $\mathrm{gamma}([\nu_0 + n^2]/2, [\nu_0\sigma_0^2 + SS_1 + SS_2]/2)$. As for ρ, we do not know of a standard semi-conjugate prior distribution. However, ρ is just a scalar parameter bounded between -1 and $+1$, and so approximate simulation of ρ from its full conditional distribution (given an arbitrary prior distribution) could be achieved by computing the unnormalized posterior density on a grid of values, or by slice sampling, or instead using a Metropolis–Hastings updating procedure.

To update Σ in step 4, let $\mathbf{f}_i = (\mathbf{a}_i, \mathbf{b}_i)$ and recall that the random effects model for the $\mathbf{f}_i$'s is that $\mathbf{f}_1, \ldots, \mathbf{f}_n \sim$ i.i.d. $N_2(0, \Sigma)$. Given a Wishart prior distribution for Σ^{-1}, the conditional distribution of Σ^{-1} given $\mathbf{f}_1, \ldots, \mathbf{f}_n$ is $\mathrm{Wishart}([\eta_0\Sigma_0 + \mathbf{F}^\top\mathbf{F}]^{-1}, \eta_0 + n)$, where $\mathbf{F}$ is the $n \times 2$ matrix with ith row equal to $\mathbf{f}_i$.

The missing entries of $\mathbf{Y}$ may be updated by simulating from their full conditional distributions. The full conditional distribution of diagonal entry $y_{i,i}$ is $N(m_{i,j} + a_i + b_j, \sigma^2(1+\rho))$. If a dyadic pair of outcomes $(y_{i,j}, y_{j,i})$ is missing, then its full conditional distribution is bivariate normal with mean vector $(m_{i,j} + a_i + b_j, m_{j,i} + a_j + b_i)$ and covariance matrix $\sigma^2\left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)$. However, if $y_{i,j}$ is observed and $y_{j,i}$ is not, then the full conditional distribution of $y_{j,i}$ is normal with mean $\rho \times (y_{i,j} - m_{i,j} - a_i - b_j) + m_{j,i} + a_j + b_i$ and variance $\sigma^2(1-\rho^2)$.

Step 1 of the Gibbs sampler requires simulation of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}\}$ from its joint distribution given $\mathbf{Y}$, Σ, σ^2, and ρ. This is challenging because of the dyadic correlation. However, calculations are simplified by transforming $\mathbf{Y}$ so that the dyadic correlation is zero: Given values of σ^2 and ρ, we may construct $\tilde{\mathbf{Y}} = \tilde{c}\mathbf{Y} + \tilde{d}\mathbf{Y}^\top$, where $\tilde{c} = \{(1+\rho)^{-1/2} + (1-\rho)^{-1/2}\}/(2\sigma)$ and $\tilde{d} = \{(1+\rho)^{-1/2} - (1-\rho)^{-1/2}\}/(2\sigma)$. It follows that

$$\tilde{\mathbf{Y}} \overset{d}{=} \mathbf{M}(\tilde{\mathbf{X}}, \boldsymbol{\beta}) + \tilde{\mathbf{a}}\mathbf{1}^\top + \mathbf{1}\tilde{\mathbf{b}}^\top + \mathbf{Z}, \tag{6.2}$$

where $\mathbf{Z} \sim N_{n\times n}(\mathbf{0}, \mathbf{I})$, $\tilde{\mathbf{x}}_{i,j} = \tilde{c}\mathbf{x}_{i,j} + \tilde{d}\mathbf{x}_{j,i}$, $(\tilde{a}_1, \tilde{b}_1), \ldots, (\tilde{a}_n, \tilde{b}_n) \sim$ i.i.d. $N_2(\mathbf{0}, \tilde{\Sigma})$ with $\tilde{\Sigma} = \Sigma_e^{-1/2}\Sigma\Sigma_e^{-1/2}$. Therefore, simulation of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}\}$ from its conditional distribution given $\mathbf{Y}, \Sigma, \sigma^2$ and ρ may be accomplished as follows:

1.a Compute $\tilde{\mathbf{Y}}, \tilde{\mathbf{X}}$ and $\tilde{\Sigma} = \Sigma_e^{-1/2}\Sigma\Sigma_e^{-1/2}$;

1.b Simulate $\{\boldsymbol{\beta}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}\}$ from its conditional distribution based on (6.2);

1.c Set $\binom{a_i}{b_i} = \Sigma_e^{1/2}\binom{\tilde{a}_i}{\tilde{b}_i}$ for $i = 1, \ldots, n$.

Step 1.b may be implemented by simulating $\boldsymbol{\beta}$ conditional on $\{\tilde{\mathbf{Y}}, \tilde{\mathbf{X}}, \tilde{\Sigma}\}$ and then simulating $\{\tilde{\mathbf{a}}, \tilde{\mathbf{b}}\}$ conditional on $\boldsymbol{\beta}$ and $\{\tilde{\mathbf{Y}}, \tilde{\mathbf{X}}, \tilde{\Sigma}\}$. We first derive the latter distribution, as it facilitates the derivation of the former. For notational simplicity, we drop the tildes on the symbols.

Let $\mathbf{Y} = \mathbf{M} + \mathbf{a}\mathbf{1}^\top + \mathbf{1}\mathbf{b}^\top + \mathbf{Z}$ where the elements of $\mathbf{Z}$ are i.i.d. standard normal random variables, and let $\mathbf{f} = (\mathbf{a}, \mathbf{b})$ be the concatenation of $\mathbf{a}$ and $\mathbf{b}$ so that $\mathbf{f} \sim N_{2n}(\mathbf{0}, \Sigma \otimes \mathbf{I})$, where "$\otimes$" denotes the Kronecker product. Vectorizing the formula for $\mathbf{Y}$ gives $\mathbf{y} = \mathbf{m} + \mathbf{W}\mathbf{f} + \mathbf{z}$, where $\mathbf{W}$ is the $n^2 \times 2n$ matrix obtained by column-binding the matrices $\mathbf{1}\otimes\mathbf{I}$ and $\mathbf{I}\otimes\mathbf{1}$, with $\mathbf{I}$ being the $n \times n$ identity matrix and $\mathbf{1}$ the $n \times 1$ vector of ones. Letting $\mathbf{r} = \mathbf{y} - \mathbf{m}$, standard calculations for Bayesian linear regression (e.g., Section 9.2 of Hoff (2009a)) show that the conditional density of $\mathbf{f}$ given $\mathbf{r}$ and Σ is given by

$$\begin{aligned}p(\mathbf{f}|\mathbf{r}, \Sigma) &\propto \exp\bigl(-(\mathbf{r} - \mathbf{W}\mathbf{f})^\top(\mathbf{r} - \mathbf{W}\mathbf{f})/2\bigr)\\ &\quad\times \exp\bigl(-\mathbf{f}^\top(\Sigma^{-1}\otimes\mathbf{I})\mathbf{f}/2\bigr)\\ &\propto \exp\bigl(-\mathbf{f}^\top[\mathbf{W}^\top\mathbf{W} + \Sigma^{-1}\otimes\mathbf{I}]\mathbf{f}/2 + \mathbf{f}^\top\mathbf{W}^\top\mathbf{r}\bigr).\end{aligned}$$

This is the kernel of a multivariate normal distribution with variance $\mathrm{Var}[\mathbf{f}|\mathbf{r}] = (\mathbf{W}^\top\mathbf{W} + \Sigma^{-1}\otimes\mathbf{I})^{-1}$ and expectation $\mathrm{E}[\mathbf{f}|\mathbf{r}] = (\mathbf{W}^\top\mathbf{W} + \Sigma^{-1}\otimes\mathbf{I})^{-1}\mathbf{W}^\top\mathbf{r}$. Some matrix manipulations yield $\mathrm{Var}[\mathbf{f}|\mathbf{r}] = \mathbf{G}\otimes\mathbf{I} - \mathbf{H}\otimes\mathbf{1}\mathbf{1}^\top$, where:

- $\mathbf{G} = (\Sigma^{-1} + n\mathbf{I})^{-1}$;
- $\mathbf{H} = (\Sigma^{-1} + n\mathbf{1}\mathbf{1}^\top)^{-1}\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)\mathbf{G}$.

Now let $\mathbf{s} = \mathbf{W}^\top\mathbf{r} = (\mathbf{1}^\top\mathbf{R}^\top, \mathbf{1}^\top\mathbf{R})$, the concatenation of the row sums and column sums of $\mathbf{R} = \mathbf{Y} - \mathbf{M}$. We then have $\mathrm{E}[\mathbf{f}|\mathbf{r}] = (\mathbf{G}\otimes\mathbf{I})\mathbf{s} - (\mathbf{H}\otimes\mathbf{1}\mathbf{1}^\top)\mathbf{s}$. Writing this in terms of the $n \times 2$ matrix $\mathbf{F}$ whose vectorization is $\mathbf{f}$, we have $\mathrm{E}[\mathbf{F}|\mathbf{R}] = \mathbf{S}\mathbf{G} - t\mathbf{1}\mathbf{1}^\top\mathbf{H}$, where $\mathbf{S}$ is the $n \times 2$ matrix whose first and second columns are the row and column sums of $\mathbf{R}$, respectively, and $t = \mathbf{1}^\top\mathbf{R}\mathbf{1}$, the sum total of the entries of $\mathbf{R}$. Therefore, to simulate $\mathbf{F}$ (and hence $\mathbf{a}$ and $\mathbf{b}$) from its full conditional distribution, we set $\mathbf{F}$ equal to

$$\mathbf{F} = (\mathbf{S}\mathbf{G} - t\mathbf{1}\mathbf{1}^\top\mathbf{H}) + \mathbf{E},$$

where $\mathbf{E}$ is a simulated $n \times 2$ normal matrix with mean zero and variance $\mathbf{G} \otimes \mathbf{I} - \mathbf{H} \otimes \mathbf{1}\mathbf{1}^\top$. To simulate this normal matrix, rewrite $\mathrm{Var}[\mathbf{f}|\mathbf{r}]$ as $\mathrm{Var}[\mathbf{f}|\mathbf{r}] = [\mathbf{G} - n\mathbf{H}] \otimes \mathbf{I} + n\mathbf{H} \otimes [\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n]$, and recognize this as the covariance matrix of

$$\mathbf{Z}_1(\mathbf{G} - n\mathbf{H})^{1/2} + (\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{Z}_2(\sqrt{n}\mathbf{H})^{1/2},$$

where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are both $n \times 2$ matrices of standard normal entries. To summarize, to simulate $\mathbf{F}$ from its full conditional distribution:

1. Simulate two $n \times 2$ matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ with i.i.d. standard normal entries;
2. Compute $\mathbf{E} = \mathbf{Z}_1(\mathbf{G} - n\mathbf{H})^{1/2} + (\mathbf{I} - \mathbf{1}\mathbf{1}^\top/n)\mathbf{Z}_2 \times (\sqrt{n}\mathbf{H})^{1/2}$;
3. Set $\mathbf{F} = (\mathbf{S}\mathbf{G} - t\mathbf{1}\mathbf{1}^\top\mathbf{H}) + \mathbf{E}$.

The conditional mean and variance of $\mathbf{f} = (\mathbf{a}, \mathbf{b})$ can be used to obtain the distribution of $\boldsymbol{\beta}$ given $\mathbf{y}$ and Σ, but marginal over $\mathbf{a}$ and $\mathbf{b}$. First, we find the density of $\mathbf{y}$ conditional on $\boldsymbol{\beta}$ and Σ. This density is given by $p(\mathbf{y}|\boldsymbol{\beta}, \Sigma) = \int p(\mathbf{y}|\mathbf{f}, \boldsymbol{\beta})p(\boldsymbol{f}|\Sigma)\,d\mathbf{f}$. As described above, the conditional distribution of $\mathbf{y}$ given $\mathbf{f}$ and $\boldsymbol{\beta}$ is $N_{n^2}(\mathbf{m} + \mathbf{W}\mathbf{f}, \mathbf{I})$ so the first term in the integrand is $(2\pi)^{-n^2/2}\exp(-\mathbf{r}^\top\mathbf{r}/2 + \mathbf{f}^\top\mathbf{W}^\top\mathbf{r} - \mathbf{f}^\top\mathbf{W}^\top\mathbf{W}\mathbf{f}/2)$. Combining this with the $N_{2n}(0, \Sigma \otimes \mathbf{I})$ prior distribution for $\mathbf{f}$ gives

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\beta}, \Sigma) &= (2\pi)^{-n^2/2}e^{-\|\mathbf{r}\|^2/2}|2\pi\,\Sigma|^{-n/2} \\ &\quad \times \int e^{\mathbf{f}^\top\mathbf{W}^\top\mathbf{r} - \mathbf{f}^\top(\mathbf{W}^\top\mathbf{W} + \Sigma^{-1}\otimes\mathbf{I})\mathbf{f}/2}\,d\mathbf{f} \\ &= (2\pi)^{-n^2/2}e^{-\|\mathbf{r}\|^2/2}|2\pi\,\Sigma|^{-n/2} \\ &\quad \times \int e^{\mathbf{f}^\top\mathrm{Var}[\mathbf{f}|\mathbf{r}]^{-1}\mathrm{E}[\mathbf{f}|\mathbf{r}] - \mathbf{f}^\top\mathrm{Var}[\mathbf{f}|\mathbf{r}]^{-1}\mathbf{f}/2}\,d\mathbf{f}. \end{aligned}$$

Completing the square and integrating over $\mathbf{f}$ gives the marginal density,

$$\begin{aligned} &p(\mathbf{y}|\boldsymbol{\beta}, \Sigma) \\ &\quad = (2\pi)^{-n^2/2}e^{-\|\mathbf{r}\|^2/2}|\Sigma|^{-n/2}\big|\mathrm{Var}[\mathbf{f}|\mathbf{r}]\big|^{1/2} \\ &\qquad \times \exp\big(\mathrm{E}[\mathbf{f}|\mathbf{r}]^\top\mathrm{Var}[\mathbf{f}|\mathbf{r}]^{-1}\mathrm{E}[\mathbf{f}|\mathbf{r}]/2\big). \end{aligned}$$

Plugging in the values of $\mathrm{Var}[\mathbf{f}|\mathbf{r}]$ and $\mathrm{E}[\mathbf{f}|\mathbf{r}]$ and simplifying gives the following expression for the uncorrelated SRRM likelihood:

$$\begin{aligned} &p(\mathbf{y}|\boldsymbol{\beta}, \Sigma) \\ &\quad = (2\pi)^{-n^2/2}|\mathbf{I} + n\,\Sigma|^{-(n-1)/2}\big|\mathbf{I} + n\,\Sigma\mathbf{1}\mathbf{1}^\top\big|^{-1/2} \\ &\qquad \times \exp\big\{-\big(\mathbf{r}^\top\mathbf{r} + t^2\mathbf{1}^\top\mathbf{H}\mathbf{1} - \mathrm{tr}\big(\mathbf{S}^\top\mathbf{S}\mathbf{G}\big)\big)/2\big\}. \end{aligned}$$

This is quadratic in $\mathbf{r}$, and hence also quadratic in $\boldsymbol{\beta}$. Some algebra gives

$$\begin{aligned} &p(\mathbf{y}|\boldsymbol{\beta}, \Sigma) \\ &\quad \propto \exp\big\{-\boldsymbol{\beta}^\top(\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3)\boldsymbol{\beta}/2 \\ &\qquad + \boldsymbol{\beta}^\top(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3)\big\}, \end{aligned}$$

where $\mathbf{Q}_1 = \mathbf{X}^\top\mathbf{X}$ and $\boldsymbol{\ell}_1 = \mathbf{X}^\top\mathbf{y}$, with $\mathbf{X}$ being the $n^2 \times p$ matrix of the $\mathbf{x}_{i,j}$'s; $\mathbf{Q}_2 = n^4h\bar{\mathbf{x}}\bar{\mathbf{x}}^\top$ and $\boldsymbol{\ell}_2 = n^4h\bar{\mathbf{x}}\bar{y}$ with $h = \mathbf{1}^\top\mathbf{H}\mathbf{1}$, $\bar{\mathbf{x}}$ being the average of the $\mathbf{x}_{i,j}$'s and $\bar{y}$ being the average of the $y_{i,j}$'s, and

$$\begin{aligned} \mathbf{Q}_3 &= -n^2\big(g_{11}\bar{\mathbf{X}}_r^\top\bar{\mathbf{X}}_r + g_{12}\big(\bar{\mathbf{X}}_r^\top\bar{\mathbf{X}}_c + \bar{\mathbf{X}}_c^\top\bar{\mathbf{X}}_r\big) + g_{22}\bar{\mathbf{X}}_c^\top\bar{\mathbf{X}}_c\big), \\ \boldsymbol{\ell}_3 &= -n^2\big(g_{11}\bar{\mathbf{X}}_r^\top\bar{\mathbf{y}}_r + g_{12}\big(\mathbf{X}_r^\top\bar{\mathbf{y}}_c + \bar{\mathbf{X}}_c^\top\bar{\mathbf{y}}_r\big) + g_{22}\bar{\mathbf{X}}_c^\top\bar{\mathbf{y}}_c\big), \end{aligned}$$

where $\bar{\mathbf{y}}_r$ is the $n \times 1$ vector of row means of $\mathbf{Y}$, $\bar{\mathbf{X}}_r$ is the $n \times p$ matrix whose ith row is the average of $\mathbf{x}_{i,j}$ over $j = 1, \ldots, n$, and $\bar{\mathbf{y}}_c$ and $\bar{\mathbf{X}}_c$ are analogously defined as column means. Now the prior density for $\boldsymbol{\beta}$ is proportional to $\exp\{-\boldsymbol{\beta}^\top\mathbf{Q}_0\boldsymbol{\beta}/2 + \boldsymbol{\beta}^\top\mathbf{Q}_0\boldsymbol{\beta}_0\}$, and so the conditional density is given by

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{y}, \Sigma) &\propto p(\mathbf{y}|\boldsymbol{\beta}, \Sigma) \times \pi(\boldsymbol{\beta}) \\ &\propto \exp\big\{-\boldsymbol{\beta}^\top(\mathbf{Q}_0 + \mathbf{Q})\boldsymbol{\beta}/2 + \boldsymbol{\beta}^\top(\mathbf{Q}_0\boldsymbol{\beta}_0 + \boldsymbol{\ell})\big\}, \end{aligned}$$

where $\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3$ and $\boldsymbol{\ell} = \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3$. This is a multivariate normal density, with variance $(\mathbf{Q}_0 + \mathbf{Q})^{-1}$ and mean $(\mathbf{Q}_0 + \mathbf{Q})^{-1}(\mathbf{Q}_0\boldsymbol{\beta}_0 + \boldsymbol{\ell})$.

6.2 Gibbs Sampling for the AME

Now suppose that $\mathbf{Y}$ follows a Gaussian AME model, so that $\mathbf{Y} = \mathbf{M}(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{U}\mathbf{V}^\top + \mathbf{a}\mathbf{1}^\top + \mathbf{1}\mathbf{b}^\top + \mathbf{E}$ where the distribution of $\{\mathbf{a}, \mathbf{b}, \mathbf{E}\}$ follows the social relations covariance model with parameters $\{\Sigma, \sigma^2, \rho\}$. Let $(\mathbf{u}_i, \mathbf{v}_i) \sim N_{2r}(\mathbf{0}, \Psi)$ independently across nodes, and let $\Psi^{-1} \sim \mathrm{Wishart}(\Psi_0^{-1}/\kappa_0, \kappa_0)$ *a priori*. The joint posterior distribution of the unknown parameters may be approximated by a Gibbs sampler that iterates the following steps:

1. Update $(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \sigma^2, \rho, \Sigma)$ and the missing values of $\mathbf{Y}$ using the algorithm described in Section 6.1, but with $\mathbf{Y}$ replaced by $\mathbf{Y} - \mathbf{U}\mathbf{V}^\top$;
2. Simulate $\Psi^{-1} \sim \mathrm{Wishart}((\Psi_0\kappa_0 + [\mathbf{U}\mathbf{V}]^\top[\mathbf{U}\mathbf{V}])^{-1}, \kappa_0 + n)$, where $[\mathbf{U}\mathbf{V}]$ is the $n \times 2r$ matrix equal to the column-wise concatenation of $\mathbf{U}$ and $\mathbf{V}$;
3. For each $k = 1, \ldots, r$, simulate the rth columns of $\mathbf{U}$ and $\mathbf{V}$ from their full conditional distributions.

To perform step 3, first consider the full conditional distribution of $\mathbf{u}_1$, the first column of $\mathbf{U}$. Let $\mathbf{R} = \mathbf{Y} - (\mathbf{M}(\mathbf{X}, \boldsymbol{\beta}) + \sum_{k=2}^r \mathbf{u}_k\mathbf{v}_k^\top + \mathbf{a}\mathbf{1}^\top + \mathbf{1}\mathbf{b}^\top)$. Then we have $\mathbf{R} = \mathbf{u}_1\mathbf{v}_1^\top + \mathbf{E}$. Decorrelating gives $\tilde{\mathbf{R}} = \tilde{c}\mathbf{R} + \tilde{d}\mathbf{R} = \tilde{c}\mathbf{u}_1\mathbf{v}_1^\top + \tilde{d}\mathbf{v}_1\mathbf{u}_1^\top + \mathbf{Z}$, and vectorizing gives $\tilde{\mathbf{r}} = [\tilde{c}(\mathbf{v}_1 \otimes \mathbf{I}) + \tilde{d}(\mathbf{I} \otimes \mathbf{v}_1)]\mathbf{u}_1 + \mathbf{z}$. Given $\mathbf{v}_1$, this is a linear regression model with outcome vector $\tilde{\mathbf{r}}$, design matrix $\mathbf{W} = [\tilde{c}(\mathbf{v}_1 \otimes \mathbf{I}) + \tilde{d}(\mathbf{I} \otimes \mathbf{v}_1)]$, regression parameters $\mathbf{u}_1$, and i.i.d. standard normal errors. Let $\boldsymbol{\mu}_{u|v}$ and $\Sigma_{u|v}$ be the conditional mean and variance of $\mathbf{u}_1$ given $\mathbf{v}_1$. Then the conditional distribution of $\mathbf{u}_1$ given $\mathbf{v}_1$ and $\tilde{\mathbf{R}}$ is normal with mean and variance given by

$$\begin{aligned} \mathrm{Var}[\mathbf{u}_1|\tilde{\mathbf{R}}, \mathbf{v}_1] &= \big(\Sigma_{u|v}^{-1} + \mathbf{W}^\top\mathbf{W}\big)^{-1}, \\ \mathrm{E}[\mathbf{u}_1|\tilde{\mathbf{R}}, \mathbf{v}_1] &= \big(\Sigma_{u|v}^{-1} + \mathbf{W}^\top\mathbf{W}\big)^{-1}\big(\Sigma_{u|v}^{-1}\boldsymbol{\mu}_{u|v} + \mathbf{W}^\top\tilde{\mathbf{r}}\big). \end{aligned}$$

Some calculations show that $\mathbf{W}^\top\mathbf{W} = (\tilde{c}^2 + \tilde{d}^2)\|\mathbf{v}_1\|^2\mathbf{I} + 2\tilde{c}\tilde{d}\mathbf{v}_1\mathbf{v}_1^\top$ and $\mathbf{W}^\top\tilde{\mathbf{r}} = (\tilde{c}\tilde{\mathbf{R}} + \tilde{d}\tilde{\mathbf{R}}^\top)\mathbf{v}_1$. The full conditional distribution of $\mathbf{v}_1$, and the other columns of $\mathbf{U}$ and $\mathbf{V}$, may be obtained similarly.

6.3 Gibbs Sampling for Transformation Models

A transformation model assumes that the sociomatrix $\mathbf{S}$ is a function of a latent sociomatrix $\mathbf{Y}$ that follows a Gaussian AME model with parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \rho, \Sigma, \Psi\}$. This collection of parameters does not include σ^2, because for probit models in general and for the other transformation models described in this article, the overall scale of the $y_{i,j}$'s is not identifiable, and so we fix $\sigma^2 = 1$. For the transformation models discussed in Section 4, observation of $\mathbf{S}$ implies that $\mathbf{Y} \in C(\mathbf{S})$. Given starting values of $\mathbf{Y}$ and $\boldsymbol{\theta}$, a Gibbs sampler for approximating the joint posterior distribution of $\mathbf{Y}$ and $\boldsymbol{\theta}$ conditional on $\mathbf{S}$ proceeds by iterating the following steps:

1. Update $\boldsymbol{\theta}$ conditional on $\mathbf{Y}$ with the algorithm described in Section 6.2;
2. Update $\mathbf{Y}$ conditional on $\boldsymbol{\theta}$ and $\mathbf{Y} \in C(\mathbf{S})$.

To perform step 2 of this algorithm, first consider the simple probit transformation model where the observed outcome $s_{i,j}$ is the binary indicator that the latent Gaussian variable $y_{i,j}$ is greater than zero. Let $\mu_{i,j} = \boldsymbol{\beta}^\top\mathbf{x}_{i,j} + \mathbf{u}_i^\top\mathbf{v}_j + a_i + b_j$. Then unconditional on $\mathbf{S}$ but given the other parameters, we have that

$$\begin{pmatrix} y_{i,j} \\ y_{j,i} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_{i,j} \\ \mu_{j,i} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

independently across dyads, and that $y_{i,i} \sim N(\mu_{i,i}, 1+\rho)$ independently across diagonal entries. Since the diagonal entries of $\mathbf{S}$ are undefined and the diagonal entries of $\mathbf{Y}$ are uncorrelated with the off-diagonal entries, each $y_{i,i}$ value may be updated from its $N(\mu_{i,j}, 1+\rho)$ distribution. The off-diagonal entries may be updated in two steps: first updating the elements of $\mathbf{Y}$ below the diagonal, and then updating those above. To do so, note that $y_{i,j}|y_{j,i} \sim N(\mu_{i,j} + \rho \times (y_{j,i} - \mu_{j,i}), 1-\rho^2)$. Now in the case of a probit AME model where $s_{i,j}$ is the indicator that $y_{i,j}$ is greater than zero, the full conditional distribution of $y_{i,j}$ is $N(\mu_{i,j} + \rho(y_{j,i} - \mu_{j,i}), 1-\rho^2)$ but constrained to be above zero if $y_{i,j} = 1$ and below zero otherwise. The full conditional distributions under other types of transformation models are also constrained normal distributions, where the constraint depends on the type of transformation. Univariate constrained normal distributions may be easily simulated from using the inverse-CDF method.

7. DISCUSSION

The AME framework is a modular approach for network data analysis based on three statistical models: the social relations covariance model, low-rank matrix representations via multiplicative factors, and Gaussian transformation models. Separately, each of these should be familiar to an applied statistician or data analyst: The first is a type of linear random effects model, the second is analogous to a model-based singular value decomposition, and the third forms the basis of many binary and ordinal regression models. Together, they provide a flexible model-based framework for inference that accounts for many statistical patterns often found in network data, and accommodates a variety of types of dyadic and nodal variables.

Current and future work in this area includes generalizing this framework to analyze datasets from more modern network studies that include multiple sociomatrices on one or more nodesets, such as comparison studies across multiple populations, multiple time points, multiple dyadic variables, or combinations of these. For example, Durante, Dunson and Vogelstein (2017) and Wang, Zhang and Dunson (2019) employ multiplicative network models to describe heterogeneity across a population of brain networks. Some other steps in this direction have been taken by representing a set of sociomatrices as a tensor (Hoff, 2011, 2016). However, these methods are not yet general enough to encompass the wide variety of multivariate, multilevel and longitudinal network datasets that are becoming more prevalent. What is needed is a broad framework like that which is available for generalized linear mixed models from the `nlme` or `lme4` software (Pinheiro and Bates, 2000, Walker et al., 2015), whereby a data analyst may separately select the type of data being analyzed (continuous, binary, count, etc.) and build a complicated model of dependence relationships between subsets of the data. One challenge to developing such a framework for network data is computational—the Gibbs samplers described in this article and implemented in the R package `amen` become cumbersome when the number of nodes is above a few thousand, and other integral approximation methods (such as Laplace approximations) for AME transformation models are infeasible because of the complicated dependence induced by the SRM. Fast, stable parameter estimation for large network datasets may require abandoning estimation based on a full likelihood, in favor of composite likelihood estimation (Lindsay, 1988) or modern method-of-moments approaches (Perry, 2017). Similar approaches have been proposed for other latent variable models, including pseudolikelihood (Amini et al., 2013) and method of moments estimation (Bickel, Chen and Levina, 2011) for blockmodels, and a case-control likelihood approximation for

the latent distance model (Raftery et al., 2012). Particularly promising may be variational approximations, which have been used to estimate block memberships and latent positions in the blockmodel (Celisse, Daudin and Pierre, 2012) and latent distance model (Salter-Townshend and Murphy, 2013).

ACKNOWLEDGMENTS

This research was partially supported by NSF Grant DMS-1505136.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859 https://doi.org/10.1214/13-AOS1138

ASTA, D. M. and SHALIZI, C. R. (2015). Geometric network comparisons. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, *UAI'*15 102–110. AUAI Press, Arlington, VA.

BAIER, S. L. and BERGSTRAND, J. H. (2009). Bonus vetus ols: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics* **77** 77–85.

BERGSTRAND, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Rev. Econ. Stat.* **67** 474–481.

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

BICKEL, P. J. and RITOV, Y. (1997). Local asymptotic normality of ranks and covariates in transformation models. In *Festschrift for Lucien Le Cam* 43–54. Springer, New York. MR1462938

BOND, C. F. and LASHLEY, B. R. (1996). Round-robin analysis of social interaction: Exact and estimated standard errors. *Psychometrika* **61** 303–311.

BRADU, D. and GABRIEL, K. R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. *J. Amer. Statist. Assoc.* **69** 428–436. MR0359192

CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467 https://doi.org/10.1214/12-EJS729

DORFF, C., GALLOP, M. and MINHAS, S. (2019). Networks of violence: Predicting conflict in Nigeria. *The Journal of Politics*. forthcoming.

DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J. Amer. Statist. Assoc.* **112** 1516–1530. MR3750873 https://doi.org/10.1080/01621459.2016.1219260

ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1** 211–218.

GABRIEL, K. R. (1998). Generalised bilinear regression. *Biometrika* **85** 689–700. MR1665834 https://doi.org/10.1093/biomet/85.3.689

GADE, E. K., HAFEZ, M. M. and GABBAY, M. (2019). Fratricide in rebel movements: A network analysis of syrian militant infighting. *Journal of Peace Research* **56** 321–335.

GENOVART, M., GIMENEZ, O., BERTOLERO, A., CHOQUET, R., ORO, D. and PRADEL, R. (2019). Changes in social cohesion in a long-lived species under a perturbation regime. bioRxiv.

GILL, P. S. and SWARTZ, T. B. (2001). Statistical analyses for round robin interaction data. *Canad. J. Statist.* **29** 321–331. MR1840712 https://doi.org/10.2307/3316080

GOLLOB, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **33** 73–115. MR0221658 https://doi.org/10.1007/BF02289676

HANDCOCK, M. S. (2003). Assessing degeneracy in statistical models of social networks Technical Report 39, Center for Statistics and the Social Sciences, Univ. Washington, Seattle, WA.

HARRIS, K., HALPERN, C., WHITSEL, E., HUSSEY, J., TABOR, J., ENTZEL, P. and UDRY, J. (2009). The national longitudinal study of adolescent health: Research design.

HOFF, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J. Amer. Statist. Assoc.* **100** 286–295. MR2156838 https://doi.org/10.1198/016214504000001015

HOFF, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer and S. Roweis, eds.) **20** 657–664. MIT Press, Cambridge, MA.

HOFF, P. D. (2009a). *A First Course in Bayesian Statistical Methods*. *Springer Texts in Statistics*. Springer, New York. MR2648134 https://doi.org/10.1007/978-0-387-92407-6

HOFF, P. D. (2009b). Multiplicative latent factor models for description and prediction of social networks. *Comput. Math. Organ. Theory* **15** 261–272.

HOFF, P. D. (2011). Hierarchical multilinear models for multiway data. *Comput. Statist. Data Anal.* **55** 530–543. MR2736574 https://doi.org/10.1016/j.csda.2010.05.020

HOFF, P. (2015). Dyadic data analysis with amen Technical Report 638, Dept. Statistics, Univ. Washington, Seattle, WA.

HOFF, P. D. (2016). Equivariant and scale-free Tucker decomposition models. *Bayesian Anal.* **11** 627–648. MR3498040 https://doi.org/10.1214/14-BA934

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

HOFF, P., FOSDICK, B., VOLFOVSKY, A. and HE, Y. (2012). amen: Additive and multiplicative effects modeling of networks and relational data.

HOFF, P., FOSDICK, B., VOLFOVSKY, A. and STOVEL, K. (2013). Likelihoods for fixed rank nomination networks. *Netw. Sci.* **1** 253–277.

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

HUNTER, D. R., GOODREAU, S. M. and HANDCOCK, M. S. (2008). Goodness of fit of social network models. *J. Amer. Statist. Assoc.* **103** 248–258. MR2394635 https://doi.org/10.1198/016214507000000446

HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264 https://doi.org/10.1198/106186006X133069

ISARD, W. (1954). Location theory and trade theory: Short-run analysis. *Q. J. Econ.* **68** 305–320.

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

KOSTER, J. (2018). Family ties: The multilevel effects of households and kinship on the networks of individuals. *R. Soc. Open Sci.* **5** 172159. https://doi.org/10.1098/rsos.172159

KRIVITSKY, P. N., HANDCOCK, M. S., RAFTERY, A. E. and HOFF, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Soc. Netw.* **31** 204–213.

LI, H. and LOKEN, E. (2002). A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statist. Sinica* **12** 519–535. MR1902723

LINDSAY, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (*Ithaca*, *NY*, 1987). *Contemp. Math.* **80** 221–239. Amer. Math. Soc., Providence, RI. MR0999014 https://doi.org/10.1090/conm/080/999014

LUSHER, D., KOSKINEN, J. and ROBINS, G. (2013). *Exponential Random Graph Models for Social Networks*: *Theory*, *Methods*, *and Applications*. Cambridge University Press, Cambridge.

NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735

PERRY, P. O. (2017). Fast moment-based estimation for hierarchical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 267–291. MR3597973 https://doi.org/10.1111/rssb.12165

PINHEIRO, J. C. and BATES, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, Berlin.

PRICE, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology* **27** 292–306.

RAFTERY, A. E., NIU, X., HOFF, P. D. and YEUNG, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *J. Comput. Graph. Statist.* **21** 901–919. MR3005803 https://doi.org/10.1080/10618600.2012.679240

RASTELLI, R., FRIEL, N. and RAFTERY, A. E. (2016). Properties of latent variable network models. *Netw. Sci.* **4** 407–432.

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

SALTER-TOWNSHEND, M. and MURPHY, T. B. (2013). Variational Bayesian inference for the latent position cluster model for network data. *Comput. Statist. Data Anal.* **57** 661–671. MR2981116 https://doi.org/10.1016/j.csda.2012.08.004

SCHWEINBERGER, M. and HANDCOCK, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 647–676. MR3351449 https://doi.org/10.1111/rssb.12081

SCHWEINBERGER, M. and SNIJDERS, T. A. B. (2003). Settings in social networks: A measurement model. *Sociol. Method.* **33** 307–341.

SCHWEINBERGER, M., KRIVITSKY, P. N., BUTTS, C. T. and STEWART, J. (2017). Exponential-family models of random graphs: Inference in finite-, super-, and infinite population scenarios. Preprint.

SNIJDERS, T. A., PATTISON, P. E., ROBINS, G. L. and HANDCOCK, M. S. (2006). New specifications for exponential random graph models. *Sociol. Method.* **36** 99–153.

VAN DUIJN, M. A. J., SNIJDERS, T. A. B. and ZIJLSTRA, B. J. H. (2004). p_2: A random effects model with covariates for directed graphs. *Stat. Neerl.* **58** 234–254. MR2064846 https://doi.org/10.1046/j.0039-0402.2003.00258.x

WALKER, S. C., BOLKER, B. M., MÄCHLER, M. and BATES, D. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48.

WANG, L., ZHANG, Z. and DUNSON, D. (2019). Common and individual structure of brain networks. *Ann. Appl. Stat.* **13** 85–112. MR3937422 https://doi.org/10.1214/18-AOAS1193

WARNER, R., KENNY, D. A. and STOTO, M. (1979). A new round robin analysis of variance for social interaction data. *J. Pers. Soc. Psychol.* **37** 1742–1757.

WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis*: *Methods and Applications*. Cambridge Univ. 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* **61** 401–425. MR1424909 https://doi.org/10.1007/BF02294547

WONG, G. Y. (1982). Round robin analysis of variance via maximum likelihood. *J. Amer. Statist. Assoc.* **77** 714–724. MR0686400

YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. MR2504912 https://doi.org/10.1007/978-3-540-77004-6_11

ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083 https://doi.org/10.1214/12-AOS1036