Testing and Modeling Dependencies Between a Network and Nodal Attributes

Bailey K. Fosdick[1] and Peter D. Hoff[1,2]

Working Paper no. 131

Center for Statistics and the Social Sciences

University of Washington

Seattle, WA 98195-4322

June 12, 2013

Abstract

Network analysis is often focused on characterizing the dependencies between network relations and node-level attributes. Potential relationships are typically explored by modeling the network as a function of the nodal attributes or by modeling the attributes as a function of the network. These methods require specification of the exact nature of the association between the network and attributes, reduce the network data to a small number of summary statistics, and are unable provide predictions simultaneously for missing attribute and network information. Existing methods that model the attributes and network jointly also assume the data are fully observed. In this article we introduce a unified approach to analysis that addresses these shortcomings. We use a latent variable model to obtain a low dimensional representation of the network in terms of node-specific network factors and use a test of dependence between the network factors and attributes as a surrogate for a test of dependence between the network and attributes. We propose a formal testing procedure to determine if dependencies exists between the network factors and attributes. We also introduce a joint model for the network and attributes, for use if the test rejects, that can capture a variety of dependence patterns and be used to make inference and predictions for missing observations.

Keywords: hypothesis test; joint model; latent variable model; prediction; relational data

1 Introduction

A common goal in the analysis of network data is to characterize the dependence between network relations and a set of node-specific attributes. For example in recent years many studies in the social sciences have examined the relationship between individuals' friendship networks and their health measures, such as happiness (Fowler and Christakis (2008)), smoking and drinking behavior (Kiuru et al. (2010)), and obesity (Christakis and Fowler (2007), de la Haye et al. (2010)). Similarly in the biological sciences, scientists are interested in the relationship between how proteins interact and their biological importance (see Butland et al. (2005) for example). In each of these applications, the data consists of two parts: the network relations $\{y_{i,j} : i, j \in \{1, ..., n\}\}$ representing a measure of the directed relationship between each pair of nodes i and j, and p-variate nodal attributes $\{\boldsymbol{x}_i : i \in \{1, ..., n\}\}$. In the case of a social network, the nodes, network relations, and attributes often represent people, their friendships, and their demographic and behavioral characteristics, respectively.

Traditional approaches to describing the dependence between a network and attributes rely on statistical methods that model either the network conditional on the attributes or the attributes conditional on the network. In the social sciences, this first perspective parallels the theory of "social selection", whereby individuals' attributes influence the formation of their social relations, and the second perspective is motivated by "social influence" theory whereby individuals' relations affect their attributes.

Methods that model the network as a function of the attributes commonly specify a regression framework for the dependence: the probability of the relation $y_{i,j}$ is a function of $\boldsymbol{\beta}^T \boldsymbol{x}_{i,j}$ where $\boldsymbol{x}_{i,j} = f(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the attributes for nodes i and j, and $\boldsymbol{\beta}$ is an unknown parameter vector. The covariate vector $\boldsymbol{x}_{i,j}$ typically includes terms for each attribute of the sender node i and receiver node j, as well as interaction terms measuring the similarity between the sender and receiver attributes. These interaction terms are frequently defined as the absolute difference between the attributes, an indicator of whether an attribute is the same for both the sender and receiver node (in the case of discrete attributes), or the product of the nodes' attributes. Examples of network models that can accommodate such a regression term are exponentially parameterized random graph models (ERGM) (Frank and Strauss (1986), Wasserman and Pattison (1996), Snijders et al. (2006), Hunter and Handcock (2006)) and la-

tent variables models (Hoff et al. (2002), Hoff (2005)). This latter class of models regresses a function of the network on both attribute terms and node-specific latent variables; Austin et al. (2013) proposed a slight modification to this class where the network is expressed as a function of nodal latent variables and the latent variables are regressed on the attributes.

Methods for assessing the impact of the network on nodal attributes often regress each node's attributes on the attributes of other nodes in the network according to the network relations. For example, Christakis and Fowler (2007) use a logistic regression model to estimate the degree to which an individual's obesity status can be explained by the obesity status of individuals in their social network (children, neighbors, spouse, etc.). Other similar models include the auto-regressive network effects models of Erbring and Young (1979) and Marsden and Friedkin (1993) and the p* social influence models of Robins et al. (2001). All of these models are univariate, focusing a single attribute of interest that is possibly subject to social influence.

While modeling the network and attributes as functions of one another is able to provide some insight into their dependence structure, there are two primary drawbacks to utilizing these methods for analysis. First, neither modeling framework allows for simultaneous inference about the dependencies between and among the network relations and attributes. For example, when analyzing data on an adolescent friendship network and individuals' health behaviors there may be interest in whether smoking habits and obesity status are conditionally independent given the network. Addressing this question of dependence between attributes conditional on the network is impossible using either of the conditional modeling frameworks. A second limitation of these methods is that they are unable to accommodate, and provide predictions for, datasets that have both missing network and attribute information. In the conditional modeling frameworks either the network or attributes are assumed to be fully observed.

Fellows and Handcock (2012) proposed a new class of models called exponential-family random network models which is a combination of an ERGM and a Gibbs random field. This joint network and attribute model addresses the first limitation of the conditional models and could potentially (with modification) address the second limitation of imputing missing data. However these models, like ERGMs, are difficult to estimate and can suffer from model degeneracy problems, where networks simulated from the fitted model are unlike that which was observed (Handcock (2003), Schweinberger (2011)). Kim and Leskovec (2011) and Kim and Leskovec (2012) proposed a simple joint attribute and network model for which mathematical analysis
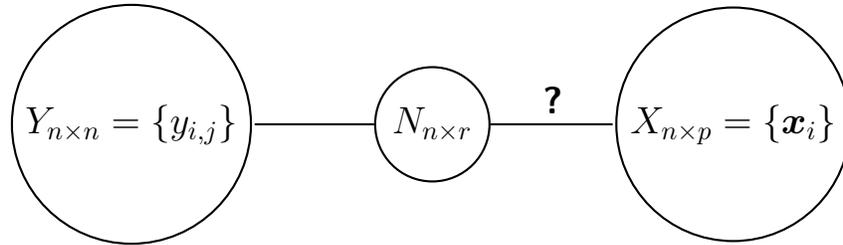
Figure 1: The primary patterns in the network Y are represented by r node-specific factors N. To determine if dependencies exist between the network Y and the p nodal attributes X, we propose testing for a the relationship between the network factors and attributes.

on network connectivity and degree distributions is tractable. However, this model class only accommodates categorical attributes and assumes no missing attribute or network information. Both of these existing joint modeling frameworks lack traditional procedures for testing whether the joint model is appropriate and dependencies even exist between the network and attributes.

In this article, we propose a unified approach to the analysis of network and attribute data. This approach allows for testing for dependencies between the network and attributes, and in the event the test concludes such dependencies exists, jointly modeling the network and attributes to make inference and obtain predictions for missing values. Our proposed methodology can be summarized as follows. Investigating the dependence between network data Y and attribute data X is difficult since network data is often high dimensional, containing relational information on each pair of nodes, and there lacks a one-to-one correspondence between nodes' network relations and their attributes. For these reasons, in Section 2 we propose representing the $(n \times n)$ matrix of network relations Y with a low dimensional structure defined by an $(n \times r)$ matrix N of node-specific network factors $(r \ll n)$. These network factors N are not observed directly and hence are estimated from the observed network Y using a network model.

In Section 3 we propose evaluating whether dependencies exist between the network Y and attributes X by formally testing for correlation between the estimated network factors N and the attributes X. A conceptual representation of this testing framework is shown in Figure 1. If the network is independent of the attributes, then any functions of the network, specifically the network factors, are also independent of the network. Therefore, any test of association between the network factors and attributes will have the correct Type I error rate. A key advantage

of this approach is that the overall relationship between the network and an arbitrary number of attributes can be assessed without needing to construct a complex regression model of the network relations on the attributes or perform variable selection. In Section 4 we investigate the loss in power for the test between the network factors and attributes as a result of not observing the network factors directly.

A joint model for the network and attributes is presented in Section 5 for use when the test of independence between the network factors and attributes rejects. This joint model allows for simultaneous estimation and inference on the dependence between and within the network and attributes, as well as provides methodology for handling and predicting missing network and attribute data. We show that the joint model conditional on the attributes can be viewed as a reduced rank regression of the network relations on attribute interactions. This further motivates the model as a mechanism for parsimoniously characterizing attribute and network dependence. In Section 6 our proposed methodology in used to analyze data from the National Longitudinal Study of Adolescent Health. In a cross validation experiment, we demonstrate that predictions of missing attribute data can be improved by basing imputations on both observed network and attribute information instead of attribute data alone. We conclude with a discussion in Section 7.

2   Calculation of node-specific network factors

The latent space network models in Hoff et al. (2002) and latent variable models in Hoff (2005) and Hoff (2009) provide parsimonious representations of the patterns in a network using node-specific latent factors. These models have been shown to capture a variety of network dependence patterns such as homophily, transitivity, reciprocity, and heterogeneity in node sociability and popularity. We consider an extension of the model presented in Hoff (2009) that contains additive and multiplicative latent effects, as well as structure for within dyad correlation. Ultimately, we use this model to obtain a low-dimensional representation of the network in terms of interpretable node-specific factors. We describe the model for continuous network data, however at the end of this section we briefly discuss how these methods can be extended to model ordinal or binary relations.

Let $y_{i,j}$ represent a continuous measure of the directed relation between nodes i and j and

consider the following model:

$$y_{i,j} = \mu + a_i + b_j + \boldsymbol{u}_i^T \boldsymbol{v}_j + e_{i,j}, \qquad a_i, b_j \in \mathbb{R}, \qquad \boldsymbol{u}_i, \boldsymbol{v}_j \in \mathbb{R}^k. \qquad (1)$$

The overall mean relation is represented by μ and the random error by $e_{i,j}$. The additive sender effect a_i and receiver effect b_j are often interpreted as a measure of node i's sociability (i.e. outgoingness) and node j's popularity respectively. The multiplicative interaction effect $\boldsymbol{u}_i^T \boldsymbol{v}_j$ can capture higher order dependence, such as network transitivity, balance, and clustering (Hoff (2005)). One interpretation of the these effects comes from the concept of an underlying social space (Mcfarland and Brown (1973), Faust (1988)), whereby nodes that are close to one another in the underlying space exhibit similar network patterns. In this context, the node-specific sender factors $\boldsymbol{u}_i$ and receiver factors $\boldsymbol{v}_i$ can be interpreted as k-dimensional representations of the underlying outgoing (sending) and incoming (receiving) behaviors of node i. A similar interpretation was used to motivate the latent position models in Hoff et al. (2002).

The random errors are modeled as Gaussian, independent across dyads, and correlated within a dyad:

$$(e_{i,j}, e_{j,i})^T \overset{\text{iid}}{\sim} \text{normal}_2\left(\boldsymbol{0}, \sigma_e^2\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)\right). \qquad (2)$$

The additive and multiplicative node-specific factors are also modeled as Gaussian and independent across nodes:

$$(a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T)^T \overset{\text{iid}}{\sim} \text{normal}_{2+2k}(\boldsymbol{0}, \Sigma_{abuv}) \qquad \Sigma_{abuv} = \begin{pmatrix} \Sigma_{ab} & \Sigma_{ab,uv} \\ \Sigma_{uv,ab} & \Sigma_{uv} \end{pmatrix}. \qquad (3)$$

The within dyad correlation ρ is interpreted as a measure of network relation reciprocity and together with the additive effects a_i and b_j induces the covariance structure from the social relations model (Warner et al. (1979), Wong (1982)).

A Bayesian estimation procedure for this network model has been implemented in the 'amen' package in the open source computing software program R; however the implementation restricts $\Sigma_{ab,uv} = 0$. Under this restriction the model can capture third-order dependence patterns between relation "cycles", such as $\{y_{i,j}, y_{j,k}, y_{k,i}\}$ or $\{y_{i,j}, y_{j,k}, y_{i,k}\}$ where the edges create a closed loop (ignoring direction), but not between noncyclic relation triples such as $\{y_{i,j}, y_{j,i}, y_{k,i}\}$. By allowing the additive and multiplicative effects to be dependent (i.e. $\Sigma_{ab,uv} \neq 0$) as in (3), the

model is able to capture a larger class of dependencies. Specifically, it can capture correlation among sets of relation triples where each relation in the set shares at least one node with another relation in the set (i.e. dependence between $\{y_{i,j}, y_{j,k}, y_{k,l}\}$, but not between $\{y_{i,j}, y_{j,k}, y_{m,l}\}$). One justification for allowing such dependence is that latent factors that act additively, affecting node popularity and sociability, plausibly also impact the network in a multiplicative manner. A modified version of the 'amen' R package that supports Bayesian parameter estimation for the network model presented here is available at the corresponding author's website.

Motivation via the singular value decomposition

A key strength of the network model in (1) is its ability to capture a variety of common network phenomena, however an alternative motivation for the model stems from its relationship to the singular value decomposition (SVD). The singular value decomposition is a matrix factorization that is commonly used to obtain an approximation of a matrix M by another matrix $\widehat{M}$ which is of reduced rank and contains the main patterns of the original matrix M. The SVD-based approximation $\widehat{M}$ is the optimal matrix approximation of its rank with respect to squared error loss. Here we show that the model in (1) is similar to an SVD-based approximation of the network Y, and hence can be viewed as a low dimensional representation of the network that captures the primary patterns in the relations.

The network model in (1) can be written in matrix form as $Y = M + E$, where

$$M = \mu \mathbf{1_n}\mathbf{1_n}^T + \boldsymbol{a}\mathbf{1_n}^T + \mathbf{1_n}\boldsymbol{b}^T + UV^T, \tag{4}$$

$\boldsymbol{a}$ and $\boldsymbol{b}$ are $(n \times 1)$ vectors of the additive sender and receiver factors, U and V are $(n \times k)$ matrices of multiplicative factors, and E is an $(n \times n)$ matrix of errors.

The singular value decomposition of an arbitrary $(n \times n)$ matrix Y is written $Y = ACB^T$, where A and B are orthogonal $(n \times n)$ matrices and C is an $(n \times n)$ diagonal matrix with non-negative decreasing diagonal elements. The rank-k matrix that best approximates Y based on squared-error loss is given by $\widehat{M} = \widehat{A}\widehat{C}\widehat{B}^T$ where $\widehat{A} = A[, 1\colon k]$, $\widehat{C} = C[1\colon k, 1\colon k]$ and $\widehat{B} = B[, 1\colon k]$. Absorbing $\widehat{C}$ into $\widehat{A}$ and/or $\widehat{B}$, the best rank-k approximation is written $\widehat{M} = \breve{A}\breve{B}^T$. Letting $\boldsymbol{\mu_A}, \boldsymbol{\mu_B} \in \mathbb{R}^k$ contain the columns means of $\breve{A}$ and $\breve{B}$ respectively, $\widehat{M}$ can be expressed:

$$\widehat{M} = \mu_{AB}\mathbf{1_n}\mathbf{1_n}^T + \tilde{\boldsymbol{a}}\mathbf{1_n}^T + \mathbf{1_n}\tilde{\boldsymbol{b}}^T + \widetilde{A}\widetilde{B}^T, \tag{5}$$

$$\widetilde{A} = (\breve{A} - \mathbf{1_n}\boldsymbol{\mu_A}^T), \qquad \tilde{\boldsymbol{a}} = \widetilde{A}\boldsymbol{\mu_B}, \qquad \mu_{AB} = \boldsymbol{\mu_A}^T\boldsymbol{\mu_B},$$

$$\widetilde{B} = (\breve{B} - \mathbf{1_n}\boldsymbol{\mu_B}^T), \qquad \tilde{\boldsymbol{b}} = \widetilde{B}\boldsymbol{\mu_A},$$

where $\widetilde{A}$ and $\widetilde{B}$ represent multiplicative factors with mean-zero columns, $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ represent mean-zero row and column factors, and μ_{AB} is an overall mean.

Observe that the representation in (5) resembles that in (4) for the network model. This illustrates the additive and multiplicative effects structure in the network model is similar to a rank-k matrix approximation of the network. Note that in the decomposition in (5), there is functional dependence between the overall mean, additive, and multiplicative effects. Since there are no such restrictions in the network model, the latent network effects represent a slightly larger class of approximations than the set of matrices with rank-k.

From the representation in (4), it is evident that the multiplicative network effects individually are nonidentifiable: the probability model for Y is the same with multiplicative latent factors U and V as it is with factors UG^T and VG^{-1} for any nonsingular $(k \times k)$ matrix G. This issue is discussed further in Sections 3 and 5.

Non-continuous network measures

Observed network information is often not continuous. For example it is common for network data to be binary where $y_{i,j}$ is an indicator of whether the relation between nodes i and j exceeds some threshold, or ordinal where $y_{i,j}$ represents, for instance, the relative rank of node j from the perspective of node i. To model non-continuous relations, the network model in (1) can be incorporated into a generalized linear model framework by modeling $y_{i,j} = \ell(z_{i,j})$ where $z_{i,j}$ is a continuous measure of the pairwise relation and ℓ is a link function defining the relationship between $z_{i,j}$ and $y_{i,j}$. The latent continuous network measure $z_{i,j}$ is then modeled using the network model in (1) in place of $y_{i,j}$. In the case of binary data, a probit or logit link function may be appropriate, and in the ordinal case the ordered probit can be considered. Hoff et al. (2012) discusses additional link functions which account for censoring of binary and ordinal relations when nodes are restricted on the number of relations they can send (i.e. the number of non-zero relations in a row of Y). Section 6 illustrates the use of an appropriate link function for fixed rank nomination data from the National Longitudinal Study of Adolescent Health.

3 Testing for dependencies

The goals in an analysis of network and attribute data are often threefold: 1) to determine whether dependencies exist between the network and attributes, 2) to model and estimate these dependencies, and finally 3) to make inference and possibly make predictions for missing data. The first step in any such analysis is to formally test for dependencies between the network and attributes.

A classical approach to determining whether there is an association between the nodal attributes $X_{n \times p}$ and network relations $Y_{n \times n}$ would be to test whether dependencies exist between X and the rows of Y or between X and the columns of Y. This would involve hypothesizing that each attribute is uncorrelated with each node's outgoing relations (H_0: $\text{Cov}(X[,i], Y[j,]) = 0$ for all i,j) or incoming relations (H_0: $\text{Cov}(X[,i], Y[,j]) = 0$ for all i,j) and investigating the evidence against these claims. However, conventional multivariate analysis tests are not applicable to these problems since these tests address relationships between $p + n$ variables based on n observations.

We propose an alternative testing approach using the the estimated latent network factors $N_{n \times (2k+2)} = [\boldsymbol{a}, \boldsymbol{b}, U, V]$ from the network model in (1). The nodal attributes $X = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]^T$ are independent of the network Y if and only if the attributes are independent of any function of the network. As described in Section 2, the network factors N provide a simplified representation of the network. Thus, we propose testing for dependence between the latent network factors N and attributes X on the basis that rejecting such a test would imply dependence between the network Y and attributes X (see Figure 1). However, the latent network factors N are not observed so in practice they must be estimated from the observed network Y. In this section we propose a test for dependence between the estimated network factors and attributes, discuss invariances in the test, and describe an exact likelihood ratio testing procedure. We also discuss alternative interpretations of the test that do not involve distributional assumptions on both the latent factors and attributes.

Suppose the nodal attributes $\{\boldsymbol{x}_i : i \in \{1, ..., n\}\}$ are continuous and mean-zero, and let $\boldsymbol{n}_i = (a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T)^T$ denote the (estimated) latent network factors for node i. We propose testing for linear dependence between the network factors and attributes using a classical multivariate test based on the assumption that the network factors and attributes are samples from a multivariate

normal distribution:

$$\left(\boldsymbol{x}_i^T, a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T\right)^T = \left(\boldsymbol{x}_i^T, \boldsymbol{n}_i^T\right)^T \overset{\text{iid}}{\sim} \text{normal}_{p+2+2k}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \Sigma_{XN} = \begin{pmatrix} \Sigma_X & \Sigma_{X,N} \\ \Sigma_{N,X} & \Sigma_N \end{pmatrix}\right). \quad (6)$$

The null and alternative hypotheses for this test are

$$\text{H}_0 : \Sigma_{X,N} = 0 \qquad \text{vs.} \qquad \text{H}_1 : \Sigma_{X,N} \neq 0 \qquad \text{based on (6).} \quad (7)$$

Network model and test invariances

As mentioned in Section 2, the network model in (1) is invariant under transformations of the multiplicative latent factors. Formally, this nonidentifiability can be expressed as a invariance of the probability model under transformations of network factors $\{\boldsymbol{n}_i : i \in \{1, ..., n\}\}$ by elements of group

$$\mathcal{G}_N = \left\{ \boldsymbol{G}_N = \begin{pmatrix} \text{I}_2 & 0 & 0 \\ 0 & A^T & 0 \\ 0 & 0 & A^{-1} \end{pmatrix} : A_{k \times k} \text{ nonsingular} \right\},$$

which act via multiplication on the left: $\boldsymbol{n}_i = (a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T)^T \to \boldsymbol{G}_N \boldsymbol{n}_i = (a_i, b_i, A^T \boldsymbol{u}_i^T, A^{-1} \boldsymbol{v}_i^T)^T$. It would be undesirable for the test in (7) to depend on which latent factors in the set $\{\{\boldsymbol{G}_N \boldsymbol{n}_i : i \in \{1, ..., n\}\} : \boldsymbol{G}_N \in \mathcal{G}_N\}$ are selected to represent the network. Define $\mathcal{G}$ to be the extension of group $\mathcal{G}_N$ to transformations of $\left(\boldsymbol{x}_i^T, \boldsymbol{n}_i^T\right)^T$:

$$\mathcal{G} = \left\{ \boldsymbol{G} = \begin{pmatrix} I_p & 0 \\ 0 & \boldsymbol{G}_N \end{pmatrix} : \boldsymbol{G}_N \in \mathcal{G}_N \right\},$$

which acts via left multiplication and leaves $\boldsymbol{x}_i$ unchanged. We define $\mathcal{G}$ in order to relate the invariance in the network model parameterization to the test in (7).

The testing problem in (7) is itself invariant under left multiplication of $\left(\boldsymbol{x}_i^T, \boldsymbol{n}_i^T\right)^T$ by elements in the group $\mathcal{F}$, where $\mathcal{F}$ is defined

$$\mathcal{F} = \left\{ \boldsymbol{F} = \begin{pmatrix} B^X & 0 \\ 0 & B^N \end{pmatrix} : B_{p \times p}^X, B_{(2k+2) \times (2k+2)}^N \text{ nonsingular} \right\}.$$

An $\mathcal{F}$-invariant test is a test for (7) that produces the same results for all attributes and network factors that are equivalent under group $\mathcal{F}$. Observe that $\mathcal{G}$ is a subgroup of $\mathcal{F}$. This implies that an $\mathcal{F}$-invariant test will also respect the $\mathcal{G}$-invariances in the network relations probability

9

model. In other words, all attributes and latent network factors that are equivalent under group $\mathcal{G}$ will generate the same test results for (7) under an $\mathcal{F}$-invariant test.

Likelihood ratio test

There is no uniformly most powerful invariant test for (7), however the likelihood ratio test is $\mathcal{F}$-invariant, unbiased (Perlman and Olkin (1980)), and generally performs well. Let $N = [\boldsymbol{a}, \boldsymbol{b}, U, V]$ be the $(n \times (2k + 2))$ matrix of network factors. The likelihood ratio test statistic can be written

$$\Lambda = \frac{\max_{\Sigma} L(\Sigma | N, X)}{\max_{\Sigma_X, \Sigma_N} L_0(\Sigma_X, \Sigma_N | N, X)} = \prod_{i=1}^{p \wedge (2k+2)} (1 - r_i^2)^{-n/2} \tag{8}$$

where L_0 and L refer to the likelihood corresponding to the multivariate normal model in (6) with and without restricting $\Sigma_{N,X} = 0$. The term r_i^2 is the ith eigenvalue of

$$(X^T X)^{-1/2} (X^T N)(N^T N)^{-1}(N^T X)(X^T X)^{-1/2},$$

and its positive square root is commonly referred to as the ith canonical correlation between N and X. This correlation represents the largest correlation obtainable between a linear combination of attributes and a linear combination of the network factors such that the linear combinations are uncorrelated with the respective combinations used to obtain the first $i - 1$ correlations. The test based on (8) rejects the null hypothesis for large values of Λ and was shown to have monotonically increasing power as a function of each population canonical correlation (Anderson and Gupta (1964)).

Under the null hypothesis, $W = \Lambda^{-2/n}$ has a Wilks' Lambda $U(p, 2k+2, n-(2k+2))$ distribution, which is equivalent to the product of independent, Beta distributed random variables (Muirhead (1982)):

$$W \sim U(p, 2k+2, n-(2k+2)) = \prod_{i=1}^{p} \text{Beta}\left(\frac{n - (2k+2) - p + i}{2}, \frac{2k+2}{2}\right).$$

The α-quantiles for this distribution can be obtained via Monte Carlo estimation and used to perform exact level-α tests for (7).

Alternative interpretation of the test

The test in (7) was derived as the likelihood ratio test for a model where both the network

factors and attributes are samples from a normal distribution. However in some cases these assumptions may not be appropriate. Fortunately, alternative interpretations of the test exist that do not rely on such assumptions. The likelihood ratio test in (8) for the test in (7) is the same as the likelihood ratio test to determine whether the coefficients in a linear regression are nonzero, where either the network factors are regressed on the attributes or the attributes are regressed on the network factors. These conditional tests can be expressed

$$H_0 : \boldsymbol{\beta}_{X|N} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\beta}_{X|N} \neq \mathbf{0} \quad \text{based on} \quad \boldsymbol{x}_i | \boldsymbol{n}_i \overset{\text{iid}}{\sim} \text{normal} \left(\beta_{X|N} \boldsymbol{n}_i, \Sigma_{X|N} \right), \quad \text{and} \quad (9)$$

$$H_0 : \boldsymbol{\beta}_{N|X} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\beta}_{N|X} \neq \mathbf{0} \quad \text{based on} \quad \boldsymbol{n}_i | \boldsymbol{x}_i \overset{\text{iid}}{\sim} \text{normal} \left(\beta_{N|X} \boldsymbol{x}_i, \Sigma_{N|X} \right), \quad (10)$$

where $\beta_{X|N}$ and $\beta_{N|X}$ are $(p \times (2 + 2k))$ and $((2 + 2k) \times p)$ matrices, respectively. If the nodal attributes were specified as part of the study design or are binary or ordinal, it may be inappropriate to model them as Gaussian as is done in (7). Instead it may be preferable to test for dependence between the attributes and network factors via the conditional formulation in (10), where no distributional assumptions are placed on X. The likelihood ratio test for the tests in (7), (9) and (10) are identical, so the testing framework presented here is appropriate if the assumption of normality is reasonable for one or both of the network factors and attributes.

4   Simulation study

To analyze data with the test outlined in Section 3, the network latent factors N must be estimated from the observed network Y. We expect this to result in a decrease in the power of the test in (7) compared to if the network factors were able to be observed directly. Furthermore, we expect a greater decrease in power when the observed network relations are less informative (i.e. binary rather than continuous). In this section we present a simulation study that quantifies the degree to which power is lost when the network factors are not observed and must be estimated from observed network relations.

Consider the network model in (1) with one multiplicative effect $(k = 1)$, zero mean $(\mu = 0)$, and independent standard normal errors $(\rho = 0, \sigma_e^2 = 1)$:

$$y_{i,j} = a_i + b_j + u_i v_j + e_{i,j}, \qquad a_i, b_j, u_i, v_j \in \mathbb{R}, \qquad e_{i,j} \sim \text{normal}(0, 1). \qquad (11)$$

We consider the case where one nodal attribute is of interest $(p = 1)$ and the attribute and latent network factors have one of the following covariance structures:

A) $\mathrm{Cov}[(x_i, a_i, b_i, u_i, v_i)] = \Sigma_{XN} = \mathrm{I} + \gamma E_{x,a}$,

B) $\mathrm{Cov}[(x_i, a_i, b_i, u_i, v_i)] = \Sigma_{XN} = \mathrm{I} + \gamma E_{x,u}$.

$E_{x,a}$ is the (5×5) matrix of zeros with a one in the entires corresponding to $\mathrm{Cov}[x, a]$ and $\mathrm{Cov}[a, x]$, and $E_{x,u}$ is defined analogously. In scenario A the attribute and each network factor are uncorrelated, except the additive sender factor a_i and the attribute x_i which have correlation γ. Similarly, in scenario B correlation γ exists between the sender multiplicative factor u_i and the attribute x_i.

Monte Carlo estimates of the power based on the level-0.05 likelihood ratio test in (8) for the test in (7) were computed for squared correlation values $\gamma^2 \in \{-0.05, 0, 0.05, 0.1, 0.15, 0.2\}$, network sizes $n \in \{25, 50, 100\}$ and three decreasingly informative observations of the network:

1. $N = [\boldsymbol{a}, \boldsymbol{b}, U, V]$ is observed;

2. N is estimated from a continuous network Y according to (1);

3. N is estimated from a binary network B_d, where B_d is defined as $B_d = \{b_{i,j} : b_{i,j} = 1 \text{ if } y_{i,j} > y_d, \ 0 \text{ otherwise}\}$ and y_d is chosen such that the proportion of network relations greater than y_d (i.e. the network density) is d.

Notice that the binary network B_d is a deterministic function of the continuous network Y. We consider the binary networks with density 0.5 and 0.15. The former case represents a relatively dense binary network with many observed relations, whereas the latter case reflects more common network seen in survey data where information about only a small number of nodes' relations are available. For the continuous network Y and binary networks B_d, the Bayesian estimation procedure was used to obtain estimates of the latent network factors. A probit link function was specified for the binary networks. The additive factors $\boldsymbol{a}$ and $\boldsymbol{b}$ were estimated by their posterior means, and the multiplicative factors U and V were estimated by the first left and right singular vector of the posterior mean of the multiplicative effect UV^T.

Figure 2 shows the power estimates for the two covariance structures A and B and the four network observations (N, Y, $B_{0.5}$, $B_{0.15}$). A single power curve is shown for the latent network factors N since the correlation structures A and B are equivalent with respect to the invariances of the test in (7). Most notably, Figure 2 illustrates there is relatively little power lost when the network factors are estimated from an observed continuous or binary network, even when the network size is small. The power of the test is slightly larger when dependence
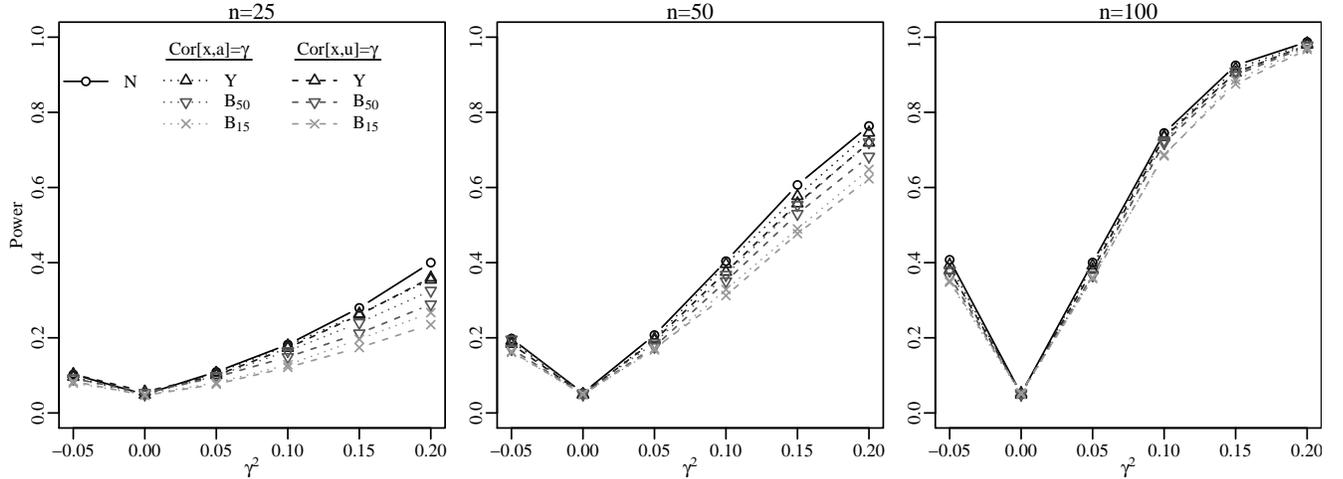
Figure 2: Power when testing for independence between a single attribute x_i and network factors $\{a_i, b_i, u_i, v_i\}$ based on four types of network observations (latent network factors N, continuous network Y, binary network $B_{0.50}$, binary network $B_{0.15}$).

exists between the attribute and an additive factor compared to when it exists between the attribute and a multiplicative factor for continuous and binary network observations. This is likely a consequence of the relative ease with which additive effects are estimated compared to interaction effects. As expected, the power of the test decreases as the observed network information becomes less informative ($N \rightarrow Y \rightarrow B_{0.5} \rightarrow B_{0.15}$), although for even moderate network sizes the power loss is negligible.

5   Joint model for the network and nodal attributes

If the test in Section 3 rejects the null hypothesis of independence between the attributes and network factors, there is often interest in estimating and making inference on the dependencies, as well as predicting missing network and attribute information. Addressing such inference objectives requires joint modeling of the network Y and attributes X. We propose jointly modeling the network Y and attributes X via a model composed of the network relations model in (1) and (2), and the latent factor and attribute model in (6). For completeness, we include all components of the joint model below.

$$y_{i,j} = \mu + a_i + b_j + \boldsymbol{u}_i^T \boldsymbol{v}_j + e_{i,j} \tag{12}$$

$$(e_{i,j}, e_{j,i})^T \stackrel{\text{iid}}{\sim} \text{normal}_2\left(\boldsymbol{0}, \sigma_e^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{13}$$

$$\left(\boldsymbol{x}_i^T, a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T\right)^T = \left(\boldsymbol{x}_i^T, \boldsymbol{n}_i^T\right)^T \stackrel{\text{iid}}{\sim} \text{normal}_{p+2+2k}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \Sigma_{XN} = \begin{pmatrix} \Sigma_X & \Sigma_{X,N} \\ \Sigma_{N,X} & \Sigma_N \end{pmatrix}\right) \tag{14}$$

Inference for the dependence and conditional dependencies between the attributes and network is based on the covariance matrix Σ_{XN}.

Simplified parameterization

The nonidentifiability of the latent factors discussed in Sections 2 and 3 translates to nonidentifiability of portions of the covariance matrix Σ_{XN}. However, by restricting the covariance matrix to have specific structure, the $\mathcal{G}$-invariance of the network model due to the multiplicative latent factors can be removed.

We propose reparameterizing the model for the latent factors and attributes in (14) by

$$\left(\boldsymbol{x}_i^T, a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T\right) \stackrel{\text{iid}}{\sim} \text{normal}_{p+2+2k}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \Sigma_{XN} = \begin{pmatrix} \Sigma_{Xab} & \Sigma_{Xab,U} & \Sigma_{Xab,V} \\ \Sigma_{U,Xab} & D & \Sigma_{U,V} \\ \Sigma_{V,Xab} & \Sigma_{V,U} & D \end{pmatrix}\right), \tag{15}$$

where D is a diagonal matrix with decreasing elements along the diagonal. This joint model defined by (12), (13), and (15) is not invariant to transformations of the network factors and attributes by elements in the group $\mathcal{G}$, however it continues to posses non-identifiabilty with respect to signs of the entries in U and V. Specifically, the probability of the observed network Y and attributes X is the same with parameters $\{U, V, \Sigma_{XN}\}$ as it is with parameters

$$\left\{US, VS, \begin{pmatrix} I_{p+2} & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & S \end{pmatrix}\Sigma_{XN}\begin{pmatrix} I_{p+2} & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & S \end{pmatrix}\right\},$$

where $S_{k \times k}$ is any diagonal matrix with ± 1's along the diagonal.

Relation to reduced rank regression

The expectation of the network relations conditional on the attributes based on (12) resembles that of a reduced rank regression on (multiplicative) attribute interaction effects. This is noteworthy as the motivations underlying reduced rank regression parallel many of the arguments supporting this network modeling framework.

The expectation of the network factors conditional on the attributes can be written

$$\mathrm{E}[(a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T)^T] = (\beta_{a|X}\boldsymbol{x}_i, \beta_{b|X}\boldsymbol{x}_i, (\beta_{U|X}\boldsymbol{x}_i)^T, (\beta_{V|X}\boldsymbol{x}_i)^T)^T,$$

where $\beta_{a|X}$, $\beta_{b|X}$ are $(p \times 1)$ vectors and $\beta_{U|X}$ and $\beta_{V|X}$ are $((2+2k) \times p)$ matrices of coefficients based on Σ_{XN}. Since the latent factors for different nodes are modeled as independent, the expectation of the network relations in (12) conditional on the attributes is

$$\mathrm{E}[y_{i,j}|\boldsymbol{x}_i, \boldsymbol{x}_j] = \mu + \beta_{a|X}\boldsymbol{x}_i + \beta_{b|X}\boldsymbol{x}_j + \boldsymbol{x}_i^T \beta_{U|X}^T \beta_{V|X}\boldsymbol{x}_j.$$

The interaction term $\boldsymbol{x}_i^T \beta_{U|X}^T \beta_{V|X}\boldsymbol{x}_j$ represents a linear combination of all possible pairwise products between the p sender and p receiver attributes, resulting in p^2 linear effects. The coefficients on these linear effects are given by the $(k \times k)$ matrix $\beta_{U|X}^T \beta_{V|X}$, whose rank is at most equal to the minimum of p and k. Therefore, if the number of attributes is greater than the number of multiplicative network factors ($p \geq k$), linear constraints will exist among the p^2 effect coefficients. In reduced rank regression the coefficient matrix corresponding to the regression of a multivariate outcome on a multivariate predictor is restricted to be reduced rank (Anderson (1951), see Reinsel and Velu (1998) for a comprehensive review). This approach to parameter dimension reduction is motivated by improvement in parameter estimation and interpretation. A similar goal exists in network modeling and is achieved here using the latent network factors. Modeling dependencies between the latent network factors N and attributes X instead of between the network relations Y and attributes X directly allows us to parsimoniously estimate and characterize complex (multiplicative) dependencies without defining a complicated regression model for the network relations. This approach is especially advantageous when the number of attributes is large and/or it is likely at most a small number of attribute pairs are related to the network.

Estimation

Estimation of the parameters in the joint network and attribute model is straightforward in a Bayesian context, where inference is based on the joint posterior distribution of the network factors $\{\boldsymbol{a}, \boldsymbol{b}, U, V\}$ and parameters $\{\sigma_e^2, \rho, \Sigma_{XN}\}$ given the data $\{X, Y\}$. Since an analytic expression of the posterior distribution is not available, it is approximated by samples generated from a Markov chain Monte Carlo (MCMC) algorithm. The MCMC procedure implemented in

the `R` package 'amen' for the model described in Section 2, where the additive and multiplicative factors are uncorrelated, was adapted for the joint model presented here. Details regarding the families of prior distributions considered and the corresponding MCMC algorithm are included in the appendix. Code is provided at the corresponding author's website.

6   Analysis of AddHealth data

We consider data from a survey of 389 high-school students from the National Longitudinal Study of Adolescent Health (AddHealth) (Harris et al. (2009)) and investigate whether evidence exists that student friendships are related to student health behaviors and grade point average (GPA). The data we use includes same-sex friendship nomination data, whereby students identified the top five friends of their sex, in addition to demographic and behavioral information. The data considered here can be described as follows:

- **network information -** $R = \{r_{i,j}\}$: $r_{i,j}$ is the rank of student j in student i's listing of friends (5 = highest, 1 = lowest) or 0 if student i did not list student j;

- **nodal attributes -** $X = [\boldsymbol{x}^{\mathrm{exercise}}, \boldsymbol{x}^{\mathrm{drink}}, \boldsymbol{x}^{\mathrm{gpa}}]$: standardized measures of exercise frequency, drinking frequency, and grade point average;

- **nodal covariate -** $W = [\boldsymbol{w}^{\mathrm{grade}}]$: student grade level (9, 10, 11, or 12).

Students in the same grade and adjacent grades are more likely to be friends than students many grades apart. For this reason, we refine our question of interest to be whether students' attributes (exercise, drinking, and GPA) are associated with their network relations' while controlling for their grade.

We use the fixed rank nomination likelihood introduced in Hoff et al. (2012) to model the observed network ranks and restriction that at most five friends could be listed on the survey. This likelihood assumes each observed network relation $r_{i,j}$ is the function of an underlying (latent) continuous measure $z_{i,j}$ such that the following relation consistencies are satisfied:

$$r_{i,j} > 0 \Rightarrow z_{i,j} > 0,$$

$$r_{i,j} > r_{i,k} \Rightarrow z_{i,j} > z_{i,k}, \tag{16}$$

$$r_{i,j} = 0 \text{ and student } i \text{ listed} < 5 \text{ friends} \Rightarrow z_{i,j} \le 0.$$

The first association is the link function used in probit regression which assumes that if a friendship is reported, the latent friendship value must exceed a given threshold (in this case 0). The second relation assures consistency of the ranks with the latent friendship measures. Finally, the last association posits that friendships between a given student and all students he/she did not list as a friend must be below the friendship threshold if the nominating student listed fewer than five friends.

The network model for the latent relations $z_{i,j}$ is that given in (1) with additional regression terms for whether students are in the same grade $w_{i,j}^s$ and whether they are in adjacent grades $w_{i,j}^a$:

$$z_{i,j} = \mu + \beta_s w_{i,j}^s + \beta_a w_{i,j}^a + a_i + b_j + \boldsymbol{u}_i^T \boldsymbol{v}_j + e_{i,j}, \qquad a_i, b_i \in \mathbb{R}, \qquad \boldsymbol{u}_i, \boldsymbol{v}_i \in \mathbb{R}^k. \quad (17)$$

Selection of factor dimension k

The multiplicative factor dimensions k for the male and female networks were determined using a method analogous to the scree plot method which is commonly used in factor analysis and principal components analysis. The network model in (16) and (17) was fit to each gender network with $k = 8$. Let M denote the posterior mean estimate of the multiplicative network effect UV^T, and $\widehat{M}$ represent the rank eight matrix approximation of M based on the singular value decomposition. The total variation in $\widehat{M}$ is equal to the sum of the squared singular values: $||\widehat{M}||_F^2 = \sum_{\ell=1}^{8} \lambda_\ell^2$, where $|| \cdot ||_F$ denotes the Frobenius norm and λ_i is the ith singular value. Figure 3 shows the proportion of the total variation in $\widehat{M}$ attributed to each multiplicative effect (i.e. $\lambda_i^2 / \sum_{\ell=1}^{8} \lambda_\ell^2$). For both the male and female network the large majority of the variation in the network relations explained by the eight multiplicative effects is associated with the first three effects. Thus, the multiplicative effect dimension was selected to be three for both networks.

Testing for dependence

As discussed in the Introduction, a traditional approach to modeling dependence between the network relations and nodal attributes would be to include regression terms in the form of sender, receiver, and interaction effects for the three attributes. Including all such effects using this approach would require 15 regression terms. However, by performing the test of
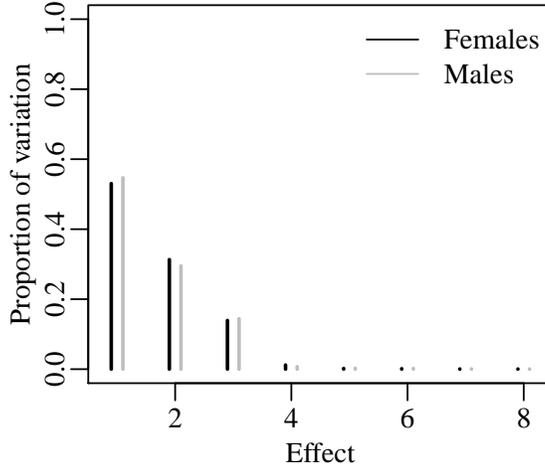
Figure 3: Proportion of variation in the posterior mean eight factor multiplicative effect $\widehat{M}$ that is explained by each multiplicative effect.

independence proposed in Section 3 based on the latent network factors, we are able to assess the evidence for any relationship between the attributes and network without creating a potentially unnecessarily complex network model or performing any model selection.

The latent network factors for the model in (16) and (17) with $k = 3$ were estimated for the male and female networks. The additive factors $\boldsymbol{a}$ and $\boldsymbol{b}$ were estimated by their posterior means, and the multiplicative factors U and V were estimated by the first three left and right singular vectors of the posterior mean of the multiplicative effect UV^T. The test of independence between the network factors and the three nodal attributes for the female and male network resulted in p-values < 0.001. Therefore, based on a 0.05 level test, we reject the null hypothesis of independence between the student attributes and their network relations after accounting for grade structure.

Jointly modeling the network and attributes

The rejections of the tests of independence between the attributes and network suggests the network factors are informative for nodal attribute data. To investigate this claim we performed a 20-fold cross validation on each sex dataset in which 5% of data for each attribute was treated as missing in each experiment. We compared predictions for the missing attributes based on

18

the observed attributes alone to predictions based on both the network and observed attributes. The predictions based solely on the attributes were the fitted values from a regression of each attribute on all other attributes. The predictions based on the network and attributes were the posterior mean estimates from the Bayesian estimation procedure for the joint network and attribute model introduced in Section 5. For each sex dataset, a Markov chain was run for 500 iterations of burn-in followed by an additional 500,000 iterations and samples were thinned to every 25th iteration, resulting in 20,000 simulated values for each missing element. The average effective sample size was 2,607 for the male network and 734 for the female network.

Table 1: Mean squared error for predictions from 20-fold cross validation.

Method	Males			Females		
	Exercise	Drinking	GPA	Exercise	Drinking	GPA
Regression (attributes only)	1.89	3.24	2.38	1.67	2.38	2.29
Joint model (attributes & network)	1.75	2.69	2.18	1.61	2.17	1.93
% improvement	7.4	17.0	8.4	3.6	8.8	15.7

Table 1 shows the mean squared error over the 20 cross validations for each attribute and each sex dataset. The predictions based on the network and attributes improved upon the predictions based on the attributes alone for both sexes and all attributes. The improvement was greatest for male drinking frequency and female GPA where prediction mean squared error was reduced by about 15%. This illustrates that when dependence exists between the network and attributes, improvements in the predictions of missing values can be obtained by using both the network and attribute information.

7 Discussion

In this article we introduced an approach for testing whether dependencies exist between a network and attribute data that relies on a simplified representation of the network in terms of latent node-specific factors. The proposed method tests for dependencies between the network latent factors and attributes as a surrogate for testing for dependencies between the network and attributes. This test was shown to have the correct level under the null hypothesis of independence and have only a slight loss in power due to the fact that the network factors are

not directly observed. Methodology for jointly modeling the network and attributes was also introduced, and in a cross validation experiment, we illustrated that predictions for missing attributes can be improved by basing predictions on both observed network and attribute information rather than on attribute information alone.

A historically difficult problem not addressed here is how to select the number of multiplicative factors for the network model. In Section 6 we illustrated a procedure similar to the scree plot method used frequently to choose the number of factors in factor analysis and the number of eigenvectors in principal component analysis. An alternative approach would be to incorporate the dimension selection into the model by placing a prior on the number of factors similar to that proposed in Hoff (2007) for the singular value decomposition. However this would greatly increase the complexity of the model and computation time of estimation.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, pages 327–351.

Anderson, T. W. and Gupta, S. D. (1964). A monotonicity property of the power functions of some tests of the equality of two covariance matrices. *Annals of Mathematical Statistics*, 35(3):1059–1063.

Austin, A., Linkletter, C., and Wu, Z. (2013). Covariate-defined latent space random effects model. *Social Networks*.

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*, 433(7025):531–537.

Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379.

de la Haye, K., Robins, G., Mohr, P., and Wilson, C. (2010). Obesity-related behaviors in adolescent friendship networks. *Social Networks*, 32(3):161–167.

Erbring, L. and Young, A. A. (1979). Individuals and social structure: Contextual effects as endogenous feedback. *Sociological Methods & Research*, 7(4):396–430.

Faust, K. (1988). Comparison of methods for positional analysis: Structural and general equivalence. *Social Networks*, 10:313–341.

Fellows, I. and Handcock, M. S. (2012). Exponential-family Random Network Models. *arXiv preprint:1208.0121*.

Fowler, J. H. and Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal*, 337.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.

Handcock, M. S. (2003). *Statistical models for social networks: Inference and degeneracy*, volume 126, pages 302–322. Committee on Human Factors, National Research Council, The National Academies Press.

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. *Journal of the American Statistical Association*, 100(469):286–295.

Hoff, P. D. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association*, 102(478):674–685.

Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational & Mathematical Organization Theory*, 15:261–272.

Hoff, P. D., Fosdick, B. K., Volfovsky, A., and Stovel, K. (2012). Likelihoods for fixed rank nomination networks. *Technical Report 608, Department of Statistics, University of Washington*.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583.

Kim, M. and Leskovec, J. (2011). Modeling social networks with node attributes using the multiplicative attribute graph model. *arXiv preprint:1106.5053*.

Kim, M. and Leskovec, J. (2012). Multiplicative attribute graph model of real-world networks. *Internet Mathematics*, 8(1-2):113–160.

Kiuru, N., Burk, W. J., Laursen, B., Salmela-Aro, K., and Nurmi, J.-E. (2010). Pressure to drink but not to smoke: disentangling selection and socialization in adolescent peer networks and peer groups. *Journal of adolescence*, 33(6):801–812.

Marsden, P. V. and Friedkin, N. E. (1993). Network studies of social influence. *Sociological Methods & Research*, 22(1):127–151.

Mcfarland, D. and Brown, D. (1973). Social distance as a metric: a systematic introduction to smallest space analysis. In *Bonds of Pluralism: The Form and Substance of Urban Social Networks*, pages 213–253. ed. E. Laumann, New York: Wiley.

Muirhead, R. (1982). *Aspects of multivariate statistical theory*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley.

Perlman, M. D. and Olkin, I. (1980). Unbiasedness of invariant tests for manova and other multivariate problems. *Annals of Statistics*, 8(6):1326–1341.

Reinsel, G. and Velu, R. (1998). *Multivariate Reduced-Rank Regression: Theory and Applications*. Lecture Notes in Statistics. Springer.

Robins, G., Pattison, P., and Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, 66:161–189.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496):1361–1370.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.

Warner, R., Kenny, D., and Stoto, M. (1979). A new round-robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology*, 37:1742–1757.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p. *Psychometrika*, 61:401–425.

Wong, G. Y. (1982). Round-robin analysis of variance via maximum likelihood. *Journal of the American Statistical Association*, 77:714–724.

A   Bayesian estimation procedure

In this section we outline the Bayesian estimation procedure used to obtain parameter estimates for the joint attribute and network model in (14). This procedure is extremely similar to that implemented in the 'amen' package in the statistical computing program R. We present the simple case here where the observed network Y is continuous, there are no regression terms in the network model, and there is no missing data. For details on accommodating non-continuous network data see Hoff et al. (2012) and for including regression terms see Hoff (2005).

Model -

$$y_{i,j} = \mu + a_i + b_j + \boldsymbol{u}_i^T \boldsymbol{v}_j + e_{i,j},$$

$$(e_{i,j}, e_{j,i})^T \overset{\text{iid}}{\sim} \text{normal}_2 \left(\boldsymbol{0}, \sigma_e^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

$$\left(\boldsymbol{x}_i^T, a_i, b_i, \boldsymbol{u}_i^T, \boldsymbol{v}_i^T \right)^T \overset{\text{iid}}{\sim} \text{normal}_{p+2+2k} \left(\boldsymbol{0}, \Sigma_{XN} \right)$$

Prior distributions -

$$\sigma_e^{-2} \sim \text{gamma}(1/2, 1/2)$$

$$\rho \sim \text{uniform}(-1, 1)$$

$$\Sigma_{XN}^{-1} \sim \text{Wishart} \left(p + 2 + 2k + 1, \begin{pmatrix} \Sigma_{X0}^{-1} & 0 \\ 0 & \text{I}_{2+2k} \end{pmatrix} \right)$$

Markov chain Monte Carlo algorithm -

Given initial values of all latent variables $\{\boldsymbol{a}, \boldsymbol{b}, U, V\}$ and parameters $\{\Sigma_{XN}, \rho, \sigma_e^2\}$, the algorithm proceeds as follows:

1. Sample $\boldsymbol{a}, \boldsymbol{b} | Y, X, U, V, \Sigma_{XN}, \rho, \sigma_e^2$ (normal).

2. Sample $\Sigma_{XN} | Y, X, \boldsymbol{a}, \boldsymbol{b}, U, V, \rho, \sigma_e^2$ (inverse-Wishart).

3. Update ρ using a Metropolis-Hastings step with proposal $\rho^* | \rho \sim$ truncated normal$_{[-1,1]}(\rho, \sigma_\rho^2)$;

4. Sample $\sigma_e^2 | Y, X, \boldsymbol{a}, \boldsymbol{b}, U, V, \rho, \Sigma_{XN}$ (inverse-gamma).

5. For each latent factor i:

   - Sample $U[, i] | Y, X, \boldsymbol{a}, \boldsymbol{b}, U[, -i], V, \rho, \sigma_e^2, \Sigma_{XN}$ (normal);

   - Sample $V[, i] | Y, X, \boldsymbol{a}, \boldsymbol{b}, U, V[, -i], \rho, \sigma_e^2, \Sigma_{XN}$ (normal).

Although the estimation algorithm is not constructed based on the unique parameterization of the model, each sample of network factors from the posterior distribution can be transformed using the covariance matrix Σ_{XN} sample to represent a sample from (15). Inference for the relative likeliness of parameter values is based on the posterior distribution over the parameter equivalence classes associated with representations congruent with (15).