ANALYZING STRUCTURAL RELATIONS IN MULTI-VARIATE DYADIC BINARY DATA

Jorge González
Francis Tuerlinckx
Paul De Boeck
K.U.Leuven, Belgium

ABSTRACT

In social network studies, most often only a single relation (or link) between the actors is investigated. When more than one link has been recorded, the two-way sociomatrix becomes a three-way array with the set of links being the third way. In this paper, we present a model which simultaneously accounts for the three ways in the data. Random effects are used to model the between-actor variability, both on senders and receivers side. In addition, structural relations between the linking variables are investigated. The model is applied to a study of popularity and strength in a class of students. It is shown that popularity can be seen as a linear function of strength on the receivers' side, but not on the senders' side.

INTRODUCTION

Dyadic data arise from the relationships of a pair of objects, called a *dyad*. Examples of objects are people, organizations, or countries, and examples of relationships are friendship, competition, or collaboration. In this paper we will use the term *actors* to refer to the objects, and *links* to refer to relationships. In the field of social networks (e.g., Wasserman & Faust, 1994), the set of actors and the links between them are called a network. Social network analysis aims at the study of the possible structures underlying the links between actors. The data are stored in a rectangular array in which the same group of actors is considered as *senders* in the rows and *receivers* in the columns. The entries of the matrix represent the links between actors. If a group of n actors is considered, then the link between sender i and receiver j is recorded in the variable X_{ij} . In the binary case, we have that $X_{ij} = 1$ if i has a link with j , and $X_{ij} = 0$ otherwise. The resulting $n \times n$ matrix is commonly known as *sociomatrix*. Figure 1 shows what the structure of the sociomatrix looks like.

Note that if $X_{ij} = 1$, this does not necessarily mean that $X_{ji} = 1$. For example, i may like j without j liking i . Note also that the main diagonal of the matrix is commonly not available (NA) because self-ties do not exist or are undefined.

Several characteristics of the data in the sociomatrix can be summarized using simple descriptive statistics called graph-theoretic statistics. For example, one

Figure 1. An example of a sociomatrix. The observation X_{ij} represents the link between sender i and receiver j .

may compute the total number of links $X_{++} = \sum_{i,j} X_{ij}$, or the number of mutual dyads (or reciprocated links) $M = \sum_{i < j} X_{ij} X_{ji}$. Another informative statistic is the number of relational links that actor i originates, called the outdegree $X_{i+} = \sum_{j=1}^n X_{ij}$ or vice versa, the indegree or degree of prestige (receptivity) $X_{+j} = \sum_{i=1}^n X_{ij}$, etc. For a complete description of these statistics and their interpretation see Wasserman and Faust (1994).

Besides these simple descriptive measures, more complex statistical models have been developed for the analysis of sociomatrices. Holland and Leinhardt (1981) made use of log-linear models and introduced the so-called p_I family of models. These models, and all the subsequent extensions of it (e.g., Fienberg & Wasserman, 1981; Wasserman & Galaskiewicz, 1984; Wang & Wong, 1987), assume that the probability distribution of the random variable X_{ij} is a function of both the graph-theoretic statistics and parameters associated to them. However, these models were criticized because they assumed independence between actors' relationships. A more flexible class of models which take into account network dependence is the p^* class of models (e.g., Wasserman & Pattison, 1996; Anderson, Wasserman, & Crouch, 1999). A different approach is the use of random effects models (Van Duijn, Snijders, & Zijlstra, 2004; Hoff, 2003). The latter approach is also taken in this paper.

ANALYZING MULTIVARIATE DYADIC BINARY DATA

In the literature, most attention has been paid to univariate networks in which only a single link between actors is studied. Fienberg, Meyer, and Wasserman (1985); Pattison and Wasserman (1999); Gill and Swartz (2004); Zijlstra, Van Duijn, and Snijders (2006), are exceptions because they propose models for extensions to multivariate networks. However, it is very often the case that more than one link within a network is recorded. When more than one link is recorded, the data are stored in a three-dimensional array of dimension $n \times n \times r$ with r being the number of links recorded. An example is shown in Figure 2. A model accounting for the three-dimensions of the data array is then desirable.

A simple solution to handle the three-way data array would be the aggregation over the senders or the receivers. The resulting two-way data could be fitted with standard models of social networks (see references above). However, aggregation has severe disadvantages. After aggregation over the senders, one can only answer questions with respect to differences and correlations across receivers. Of course, to investigate differences and the correlational structure on the senders side, one can always aggregate over the receivers, but the net result is two unrelated models. For example, if strength and popularity are the two links under study, when aggregating over senders, the questions asked concern whether persons who are judged to be strong are also considered to be popular. When aggregating over receivers, one might ask whether persons who rate others as strong tend to rate them also as popular. The relations between the links (strength and popularity) in the two models can be similar or even identical but they don't need to be, neither on statistical grounds, nor on psychological grounds.

In this paper we follow the approach of random effects models to simultaneously take into account the variability that exist among senders and receivers for each link. Moreover, we would like to study the structure of the relations between the links, both on the sender and on the receiver side. This will lead to a so-called "double structure" model.

The double structure nature of the model is an interesting feature because the structures at the senders' and receivers' side do not necessarily have to be identical. In an example that will be elaborated in detail in the application section, senders rate receivers on two features: strength and popularity. It will be assumed that senders and receivers take separate (but possibly correlated) positions on a latent strength dimension and on a latent popularity dimension. In a next step, the latent popularity score is made a function of the latent strength score. Psychologically, this means that both the sender and the receiver have a contribution in the existence of the link. In other words stated, some senders tend to have more links than others based on their psychological characteristics and social position, and some receivers attract more links than others based on the same or other psychological characteristics and characteristics of their social position. It is hypothesized that among pupils, popularity is related to strength. Pupils who are considered strong would be considered also popular, whereas those who are considered weak are seen as not popular. The correlation between strength and popularity is a natural one over receivers because they are the persons who are rated. A correlation over senders would mean that raters who consider people in general as strong, would consider people in general also as popular. Such a correlation cannot reflect a reality but only a perception or opinion about others.

Figure 2. Three-way array of senders by receivers by links. Each slide of the cube is a sociomatrix. The observation X_{ijk} represents the tie between sender i and receiver j regarding link k .

For example, when the correlation is positive over receivers, it means that the receivers who are considered stronger are also considered more popular. One may assume that it reflects an external reality, possibly a collectively constructed reality in the social group under consideration, but more than just a purely internal conception of the world.

When the correlation is derived over the senders, then it reflects the relation between attributing strength and attributing popularity, independent of the receivers, because the correlated effects are sender effects that contribute independently of the receivers. Any such correlation is therefore probably based on mental conceptions or social attitudes from the part of the senders.

The models presented in this paper can disentangle the two kinds of contributions, separate the two correlations and compare them. If the correlations agree, then the correlation based on the receiver effects are suspicious, in that it may have been induced in an indirect or direct way by an implicit theory and not by an

ANALYZING MULTIVARIATE DYADIC BINARY DATA

objective reality. However, when the two do not agree, this is much likely, and the correlation of the receiver effects may then not be seen as rooted in people's implicit theory.

In the following, our model to analyze three-way dyadic data is introduced, and the estimation of the model following a Bayesian approach is explained. A study of popularity and strength in a class of students is presented. We finalize the paper with some final remarks and discussion.

The Model

The basic model we will apply here to analyze multivariate social network data is a special case of the double-structure structural equation model (2sSEM) as described in González, De Boeck, and Tuerlinckx (2008). González et al. (2008) focused on the analysis of a three-dimensional data array with binary data on persons, situations, and emotional responses, such that a simultaneous study of the individual differences and situational differences structure of the three-mode data set was possible. We will use the model here for the analysis of multivariate network data in which a sample of persons has been asked to rate each other on a set of personality items. The items can be thought of measuring latent variables. In addition, we consider structural relations between the links under study. The structure of individual differences between the raters in terms of the personality items is not necessarily equal to the structure of the individual differences of the rated persons in terms of the same personality items. As explained earlier, one could even argue that the equality of both structures is suspicious since then the former might be a mere reflection of beliefs from the part of the raters. Here we briefly introduce the model and how it can be estimated. For a more complete description the reader is referred to González et al. (2008).

The 2sSEM model is a generalized linear mixed model (GLMM) (e.g., McCulloch & Searle, 2001), with a linear predictor, a logit link function, a distribution from the exponential family and population distributions for the random effects. The probability of sender i having a link with receiver j is modeled as a function of sender's and receiver's effects and a general effect of the link under study. We account for differences in senders and receivers considering their effects as random effects.

Model for one link

The observed random variable X_{ij} is used to denote whether sender i has a link with receiver j . We have that $X_{ij} = 1$ if i has a link with j , and $X_{ij} = 0$ otherwise. Let π_{ij} be the probability of sender i having a link with receiver j , i.e., $\pi_{ij} = P(X_{ij} = 1)$. Then,

$$\log \left(\frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \text{logit}(\pi_{ij}) = \eta_{ij} = \theta_i + \beta_j + \tau, \quad (1)$$

where θ_i and β_j represent an effect for sender i and for receiver j ($i, j = 1, \dots, n$), respectively, and τ is a general fixed effect of the link under study. It is assumed that $(\theta_i, \beta_i)^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$, a multivariate normal distribution with vector mean $\mathbf{0}$ and covariance matrix $\Sigma_{\theta\beta}$. Here, the random variable X_{ij} follows a Bernoulli distribution with parameter π_{ij} , η_{ij} is the linear predictor, and the logit is the link function.

Model for multiple links

When more than one link is considered, we add an additional index k ($k = 1, \dots, r$), in Equation (1). The model reads now as

$$\log \left(\frac{\pi_{ijk}}{1 - \pi_{ijk}} \right) = \text{logit}(\pi_{ijk}) = \eta_{ijk} = \theta_{ik} + \beta_{jk} + \tau_k, \quad (2)$$

where θ_{ik} refers to the score of sender i on the k -th link (latent trait), and, likewise, β_{jk} refers to the score of receiver j on the k -th link.

Suppose that in addition we are interested in the study of possible relations between the links under study. For example, one could hypothesize that link 2 is linearly related with link 1 and one would like to know whether this structural relation is the same for both, senders and receivers. The model equations would read as shown in (3)

$$\begin{aligned} \eta_{ij1} &= \theta_{i1} + \beta_{j1} + \tau_1, \\ \eta_{ij2} &= \theta_{i2} + \beta_{j2} + \tau_2, \end{aligned} \quad (3)$$

where $\theta_{i2} = \omega\theta_{i1} + \theta_{i2}^*$ and $\beta_{j2} = \lambda\beta_{j1} + \beta_{j2}^*$. The parameters ω and λ are regression coefficients and θ_{i2}^* and β_{j2}^* are error terms with distribution $(\theta_{i2}^*, \beta_{j2}^*)^T \sim N(\mathbf{0}, \Sigma_{\theta^*\beta^*})$. It is also assumed that $(\theta_i, \beta_i)^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$.

Estimation

We follow a Bayesian approach (e.g., Gelman, Carlin, Stern, & Rubin, 2003) to estimate the model. In the Bayesian approach, the uncertainty about a parameter vector $\mathcal{J}$ is expressed in the prior distribution $p(\mathcal{J})$. Thus, prior distributions are to be assigned for all the parameters of interest in $\mathcal{J}$. Denoting the likelihood of the data $\mathbf{X}$, given $\mathcal{J}$, by $p(\mathbf{X} | \mathcal{J})$, the inference is then based on the joint posterior distribution $p(\mathcal{J} | \mathbf{X})$, which after applying conditional distribution rules can be written as

$$p(\mathcal{J} | \mathbf{X}) \propto p(\mathbf{X} | \mathcal{J})p(\mathcal{J}). \quad (4)$$

It follows that the posterior distribution is proportional to the likelihood times the prior. Note that

$\mathcal{J} = (\theta, \beta, \tau, \sigma_{\theta}^2, \sigma_{\beta}^2, \rho_{\theta\beta})$, for the model in Equation 1, and

$\mathcal{J} = (\theta, \beta, \tau, \omega, \lambda, \sigma_{\theta}^2, \sigma_{\beta}^2, \rho_{\theta\beta}, \sigma_{\theta^*}^2, \sigma_{\beta^*}^2, \rho_{\theta^*\beta^*})$ for the model in Equation 3, and $\mathbf{X}$ are the data. The last elements of $\mathcal{J}$ are the components of the matrices $\Sigma_{\theta\beta}$ and $\Sigma_{\theta^*\beta^*}$, respectively.

As prior distributions for τ , ω , and λ , we use diffuse normal distributions with zero mean and variance 10000. These prior distributions are diffuse in the sense of having a large variance that does not contain any prior information about the location of the parameter. For the covariance matrices, inverse Wishart priors with an identity scale matrix and 2 degrees of freedom were used. These prior distributions were chosen to represent little prior information.

After drawing a sample from the posterior distribution, the posterior sample mean and standard deviations can be used to summarize the distribution of each parameter of interest. Samples from the posterior distribution can be simulated using Markov Chain Monte Carlo (MCMC) (Gelman et al., 2003) techniques. We use WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003), a software program

ANALYZING MULTIVARIATE DYADIC BINARY DATA

for Bayesian analysis of complex statistical models using MCMC techniques. The WinBUGS code used to fit the models in this paper is shown in Appendix A.

Model checking

To assess the goodness of fit of the models, we use posterior predictive checks (PPC) (Gelman, Meng, & Stern, 1996). Using this technique, lack of fit can be assessed by the tail-area probability or p -value of a test quantity, such as a global fit statistic. A reasonable range for the p -value is considered to be the interval 0.05 and 0.95 (Gelman et al., 2003). In this context, a p -value near to 0.5 would indicate a good fit, whereas more extreme values are an indication of a poor fit.

The definition of a global goodness-of-fit test quantity for the model, and the description for computing Bayesian p -values are shown in Appendix B.

Application

Data

The data are from two classes of 13-14-year-old pupils in a secondary school in Belgium. The first class has 22 students whereas the second one has 21 students. Students were presented a questionnaire with 7 questions concerning each of their classmates. The questionnaire sheet consisted of a grid with different letters to identify the classmates in the rows, and with 7 questions in columns. We will focus our study on the first class and the following two questions: 1) How strong do you feel with regard to this classmate?, and 2) How popular do you think is this classmate? For the first question, the following 5-point scale was used: 0=very weak, 1=weak, 2=neutral, 3=strong, and 4=very strong. For the second question, the following 5-point scale was used: 0=not popular at all, 1=a little popular, 2=considerably popular, 3=very popular, and 4=extremely popular. In order to illustrate the binary data approach which is quite common for social networks, the original data were dichotomized recoding the values 0, 1, and 2 as 0, and 3 and 4 as 1 in question 1). In question 2) the values 1, 2, 3, and 4 were recoded as 1, and 0 remains 0 in order to obtain an optimal classification. The dichotomized data are shown in Appendix A. The basic question we ask with respect to these data is whether one's perceived popularity can be explained by one's perceived strength. We wonder whether also on the sender's side, the given ratings of popularity can be considered as a function of the given strength ratings.

For the strength link, the total number of ties (or ones in this case) is 97. The outdegree, i.e., the number of ties sent by each student varies considerably, ranging from 0 to 21 with a mean of 4.41 and standard deviation 6.25. The indegree, i.e., the number of ties received varies between 1 and 12 with a standard deviation 2.46.

For the popularity link, the total number of ties is 312, with outdegree varying between 6 and 21 with mean 14.18 and standard deviation 5.46. The indegree varies between 7 and 19 with standard deviation 3.91.

Models

We fit three different models to the data denoted as Model 1, Model 2, and Model 3. In Model 1 the logit of the probability of sender i having a strength tie with receiver j is modeled as a function of a sender parameter $\theta_{i, str}$ plus a receiver

parameter $\beta_{j, str}$ plus a parameter representing the effect of the strength link τ_{str} . The model equation reads as

$$\eta_{ij, str} = \theta_{i, str} + \beta_{j, str} + \tau_{str}, \quad (5)$$

where $(\theta_{i, str}, \beta_{i, str})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$.

Model 2 is exactly the same model as Model 1 but considers the popularity link. The model equation is written as

$$\eta_{ij, pop} = \theta_{i, pop} + \beta_{j, pop} + \tau_{pop} \quad (6)$$

where $(\theta_{i, pop}, \beta_{i, pop})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$. Note that Models 1 and 2 consider each only one link so that they are two separate models, one for strength and one for popularity.

In Model 3, on the other hand, both the strength and popularity links are modeled together and, moreover, popularity is made a linear function of strength. The model equation can be written as

$$\begin{aligned} \eta_{ij, str} &= \theta_{i, str} + \beta_{j, str} + \tau_{str} \\ \eta_{ij, pop} &= \underbrace{\omega\theta_{i, str} + \theta_{i, pop}^*}_{\theta_{i, pop}} + \underbrace{\lambda\beta_{j, str} + \beta_{j, pop}^*}_{\beta_{j, pop}} + \tau_{pop} \end{aligned} \quad (7)$$

where θ , β , and τ are interpreted as explained before, ω and λ are regression sender and receiver coefficients, respectively, and θ^* and β^* are error terms. It is assumed that $(\theta_{i, str}, \beta_{i, str})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$, and $(\theta_{i, pop}^*, \beta_{i, pop}^*)^T \sim N(\mathbf{0}, \Sigma_{\theta^*\beta^*})$. Model 3 considers both links at the same time assuming that popularity is linearly related with strength at both senders and receivers side.

Models 1 and 2 are conceptually inferior to Model 3 in that analyzing multiple links simultaneously provides more general results compared to two analyzes of only one single link at a time. When analyzing multiple links at the same time, it is possible to investigate the relation between the links, and to compare the structure in question between senders and receivers. In despite of this, and only for explorative purposes, we fit a fourth model (Model 4) in which both links are allowed to correlate at sender's and receiver's side but there is no specific hypothesis about linear dependency between the links. Because the latent values can also correlate between senders and receivers this model actually assumes a four-variate normal distribution for the latent values. The model equation read as

$$\begin{aligned} \eta_{ij, str} &= \theta_{i, str} + \beta_{j, str} + \tau_{str}, \\ \eta_{ij, pop} &= \theta_{i, pop} + \beta_{j, pop} + \tau_{pop}, \end{aligned} \quad (8)$$

with $(\theta_{i, str}, \theta_{i, pop}, \beta_{i, str}, \beta_{i, pop})^T \sim N(\mathbf{0}, \Sigma_{\theta\beta})$. The covariance matrix $\Sigma_{\theta\beta}$ is unstructured and all (co)variances should be estimated (in the results section we will report the values for the correlations). The variances are denoted as $\sigma_{\theta_{name}}^2$ or $\sigma_{\beta_{name}}^2$ where *name* can be equal to *str* or *pop*. Notation for correlations is very similar (e.g., the correlation between sender's strength and receiver's popularity is denoted as $\rho_{\theta_{str}, \beta_{pop}}$). Model 4 can be considered an exploratory model that can be used to understand the data better.

Results

The posterior *p*-values for the four fitted models are 0.40, 0.55, 0.44, and 0.42, respectively. The *p*-value stems from the posterior predictive check (PPC) as

ANALYZING MULTIVARIATE DYADIC BINARY DATA

explained earlier (see also Appendix B), and refers to the absolute goodness of fit. The parameters estimates are presented in Table 1.

Table 1. Parameter estimates for the four fitted models. Posterior standard deviations appear between parentheses.

	Model 1	Model 2	Model 3	Model 4
$\sigma_{\theta_{str}}^2$	8.41 (4.63)		7.73* (3.69)	7.44* (3.45)
$\sigma_{\theta_{pop}}^2$		4.84* (2.45)		4.42* (2.11)
$\sigma_{\beta_{str}}^2$	1.50 (0.77)		1.26 (0.68)	1.38 (0.71)
$\sigma_{\beta_{pop}}^2$		1.51* (0.66)		1.50* (0.65)
$\rho_{\theta_{str}, \theta_{pop}}$				0.04 (0.24)
$\rho_{\theta_{str}, \beta_{str}}$	-0.37 (0.27)		-0.41 (0.25)	-0.41 (0.27)
$\rho_{\theta_{str}, \beta_{pop}}$				0.41 (0.23)
$\rho_{\theta_{pop}, \beta_{str}}$				0.03 (0.28)
$\rho_{\theta_{pop}, \beta_{pop}}$		-0.44* (0.22)		-0.41 (0.23)
$\rho_{\beta_{str}, \beta_{pop}}$				-0.66* (0.18)
τ_{str}	-2.57* (0.69)		-2.49* (0.61)	-2.48* (0.60)
τ_{pop}		1.45* (0.45)	1.56* (0.55)	1.42* (0.46)
ω			0.06 (0.22)	
λ			-0.92* (0.36)	
σ_{θ}^{2*}			5.60 (2.94)	
σ_{β}^{2*}			0.84 (0.46)	
$\rho_{\theta_{pop}, \beta_{pop}}^*$			-0.58* (0.22)	

Note. $\sigma_{\theta_l}^2$ = senders' variance for the l -th link, ($l = str, pop$); $\sigma_{\beta_l}^2$ = receivers' variance for the l -th link, ($l = str, pop$); $\rho_{\theta_{str}, \theta_{pop}}$ = correlation between senders' strength and popularity link; $\rho_{\theta_l, \beta_{l'}}$ = correlation between senders' l -th link and receivers l' -th link, ($l, l' = str, pop$); $\rho_{\beta_{str}, \beta_{pop}}$ = correlation between receivers' strength and popularity link; τ_l = effect of link l , ($l = str, pop$); ω = senders' regression coefficient; λ = receivers' regression coefficient; σ_{θ}^{2*} = residual senders' variance; σ_{β}^{2*} = residual receivers' variance. *Note.* Model 1 = separate model for strength link; Model 2 = separate model for popularity link; Model 3 = joint model for strength and popularity links, with causal relations; Model 4 = joint model for strength and popularity links, without causal links but free correlations instead. *Note.* * = significant using the rule (estimate/posterior sd) > |2|.

From these results several conclusions can be drawn. First, the single link models each have a good fit to the data. The variability at the sender side is larger than the variability at the receiver side, both for strength ($\sigma_{\theta_{str}}^2 = 8.41 > \sigma_{\beta_{str}}^2 = 1.50$) and popularity ($\sigma_{\theta_{pop}}^2 = 4.84 > \sigma_{\beta_{pop}}^2 = 1.51$). For Model 1 the value $\tau_{str} = -2.57$ can be interpreted as an indication of the probability of having a strength-tie when all the other effects are set to zero (on the logistic scale). In other words, for an average sender and an average receiver, the probability of having a strength-tie is $\exp(-2.57)/(1 + \exp(-2.57)) = 0.07$. The low value of this probability reflects the small number of ties presented in this link (97 out of 462) and means that most pupils don't feel stronger than their classmates. For Model 2, the probability of having a popularity-tie is $\exp(1.45)/(1 + \exp(1.45)) = 0.81$ and this higher value reflects the larger number of ties exhibited by this link (312 out of 462) meaning that most pupils find their classmates to be popular. Second, the multiple link models have both a good fit to the data, as may be derived from the PPC p -value. Note that the estimates of the parameters Model 3 has in common with Model 4 are similar for the two models; the results of common part are fairly stable regardless of the specific structure of the model. The most important parameters in Model 3 are the regression coefficients ω and λ . It is clear that the relation between strength and popularity is different for receivers and senders. The high and negative regression weight $\lambda = -0.92$ corresponds with a correlation $\rho_{\beta_{str}, \beta_{pop}} = -0.66$ between $\beta_{j, str}$ and $\beta_{j, pop}$ (Model 4), and may be considered a reflection of a social reality, although the causal process is unclear. The regression weight for the senders is much smaller $\omega = 0.06$ and not significant, and the corresponding correlation is $\rho_{\theta_{str}, \theta_{pop}} = 0.04$ (Model 4). This is a clear example of how the structure of two modes of a three-mode data array can be quite different. In this particular case, these two modes are the same persons, in their roles of sender and receiver. This makes the finding all the more interesting. A similar correlation for the two would mean that there are reasons to believe that the receiver based correlation is influenced by biases in the senders which are reflected in the sender effects and their correlations. However, for the data set under consideration, the correlations are not similar at all.

Two other interesting findings relate to the correlation between sender effects and receiver effects. The strength one attributes to others as a sender is negatively correlated $\rho_{\theta_{str}, \beta_{str}} = -0.41$ to the strength one is attributed as a receiver by the senders. This means that those who rate others as strong are being rated as not so strong themselves, but note that the correlation is not significant (the posterior standard deviation is 0.25). Further, when looking at the part of popularity which is not explained by strength, it can be noted that those who consider classmates as popular tend to be rated themselves as less popular $\rho_{\theta_{pop}^*, \beta_{pop}^*} = -0.58$. This correlation is significant indeed (the posterior standard deviation is 0.22). It is an intriguing finding that seeing others as popular is negatively correlated with being seen by these others as popular. It actually confirms the mimetic desire theory of Girard (1961). The theory implies that being "desired" (being popular) increases one's value in the eyes of others, and that "desiring" (considering others as popular) diminishes one's value in the eyes of others, including those of the "desired" persons. In other words stated, "desiring" makes one less "desired".

ANALYZING MULTIVARIATE DYADIC BINARY DATA

Conclusion and discussion

In this paper, the analysis of multivariate dyadic binary data has been presented. We have shown that the 2sSEM (González et al., 2008) can be used to analyze three-way data in the form of senders by receivers by links.

Following a random effects approach, the 2sSEM simultaneously accounts for the variability among two of the three ways (senders and receivers) of the data, for each link and also permits studying the structure of the relations between the links, both on the sender and on the receiver side. We believe that this latter characteristic of the model is an extension of existing models for social networks because it allows for functional relations between the links.

In a study of strength and popularity, we have found evidence for a clear difference in the structure at the sender and receiver side which supports our theory that the structures at the senders' and receivers' side do not necessarily have to be identical. Of course with a data set containing more links, more elaborate models with complex structural relations could be fitted.

Finally, we focused on the case of binary dyadic data, but the model can be easily extended in several ways. First, ordered-category such as rating scale data, and continuous data may be modeled as well, and, second, the sets of senders and receivers don't need to be the identical. The fitting of models to study the evolution of links over time is also a possible extension.

APPLIED MULTIVARIATE RESEARCH

Appendix A

WinBUGS code used to fit Model 1.

```
model;
{
  for(i in 1:22){
    for(j in 1:22) {
      logit(pstr[i,j]) <- theta[i,1] + theta[j,2] + tau
      xstr[i,j] ~ dbern(pstr[i,j])
    }
    theta[i,1:2] ~ dmnorm(mu.theta[,R[,i]])
  }
  mu.theta[1]<-0
  mu.theta[2]<-0
  R[1:2,1:2]~dwish(O[,],2)
  O[1,1]<-1
  O[2,1]<-0
  O[1,2]<-0
  O[2,2]<-1
  tau~dnorm(0,0.0001)
  SigmaT[1:2,1:2]<-inverse(R[1:2,1:2])
  CorrT<-SigmaT[1,2]/sqrt(SigmaT[1,1]*SigmaT[2,2])
}
```

WinBUGS code used to fit Model 3.

```
model;
{
  for(i in 1:22){
    for(j in 1:22) {
      logit(pstr[i,j]) <- theta[i,1] + theta[j,2] + tau[1]
      xstr[i,j] ~ dbern(pstr[i,j])
      logit(ppop[i,j]) <- w*theta[i,1]+err[i,1] + la*theta[j,2] + err[j,2] + tau[2]
      xpop[i,j] ~ dbern(ppop[i,j])
    }
    theta[i,1:2] ~ dmnorm(mu.theta[,R[,i]])
    err[i,1:2]~dmnorm(mu.err[,S[,i]])
  }
  mu.theta[1]<-0
  mu.theta[2]<-0
  R[1:2,1:2]~dwish(O[,],2)
  O[1,1]<-1
  O[2,1]<-0
```

ANALYZING MULTIVARIATE DYADIC BINARY DATA

```

O[1,2]<-0
O[2,2]<-1
mu.err[1]<-0
mu.err[2]<-0
S[1:2,1:2]~dwish(P[,],2)
P[1,1]<-1
P[2,1]<-0
P[1,2]<-0
P[2,2]<-1
for(k in 1:2){
tau[k]~ dnorm(0,0.0001)
}
w~ dnorm(0,0.0001)
la~ dnorm(0,0.0001)
SigmaT[1:2,1:2]<-inverse(R[1:2,1:2])
CorrT<-SigmaT[1,2]/sqrt(SigmaT[1,1]*SigmaT[2,2])
SigmaerrT[1:2,1:2]<-inverse(S[1:2,1:2])
CorrerrT<-SigmaerrT[1,2]/sqrt(SigmaerrT[1,1]*SigmaerrT[2,2])
}

data1
xstr[,1] xstr[,2] xstr[,3] xstr[,4] xstr[,5] xstr[,6] xstr[,7] xstr[,8] xstr[,9] xstr[,10]
xstr[,11] xstr[,12] xstr[,13] xstr[,14] xstr[,15] xstr[,16] xstr[,17] xstr[,18]
xstr[,19] xstr[,20] xstr[,21] xstr[,22]
NA 1 0 1 1 1 0 1 1 1 0 1 1 1 1 1 0 1 0 1 1 1
0 NA 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 NA 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 NA 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 NA 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0
0 1 1 1 1 NA 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1
0 0 0 0 0 0 NA 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 NA 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 NA 0 0 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 NA 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 1 0 0 NA 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1 0 0 0 0 NA 0 0 0 0 0 0 0 0
0 0 0 1 0 0 1 1 0 0 0 0 0 NA 0 0 0 0 0 0 0
0 0 0 1 0 0 0 1 0 0 1 1 0 0 NA 0 0 0 0 0 0
0 1 0 0 0 0 1 1 0 0 1 0 0 0 0 NA 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NA 0 0 0 0

```

APPLIED MULTIVARIATE RESEARCH

```
00000000000000000000NA000
00010000000000000000NA00
01111011100011000101NA0
010110011001000000000NA
END
```

data2

```
xpop[,1] xpop[,2] xpop[,3] xpop[,4] xpop[,5] xpop[,6] xpop[,7] xpop[,8]
xpop[,9] xpop[,10] xpop[,11] xpop[,12] xpop[,13] xpop[,14] xpop[,15]
xpop[,16] xpop[,17] xpop[,18] xpop[,19] xpop[,20] xpop[,21] xpop[,22]
NA0100100000000100110010
0NA00100111111011110111
11NA111111111111111111
011NA111001101111101111
1010NA11011101111101110
10101NA1001101100101010
000001NA100100100101000
00111111NA11111111101011
11101110NA1101111101111
011111011NA100111011011
0000011000NA01100101010
11111111111NA1111111111
001011101000NA111000000
1010110000000NA01101010
11101110111011NA1101111
1111111011111111NA111111
1010001100100101NA01010
111111111111111111NA1111
111111111111111111NA1111
1010111000101101101NA10
10101001010010011001NA0
1111111111111111111111NA
END
```

ANALYZING MULTIVARIATE DYADIC BINARY DATA

Appendix B

Posterior predictive p-values

To measure the discrepancy between some features of the model and data, a test quantity $T(\mathbf{x}, \mathcal{G})$ which is calculated on the data and may depend on the model parameters, is used. Then, lack of fit can be assessed by the tail-area probability or p -value of the test quantity. Note that the PPC procedure allows one to choose any test quantity which can be used to check a particular relevant aspect of the model.

We use the χ^2 discrepancy measure which assesses the model fit in a global way (Gelman et al., 2003). In the context of the application, the Pearson χ^2 measure is defined as

$$\chi^2(x, \mathcal{G}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{(x_{ijk} - E(x_{ijk} | \mathcal{G}))^2}{\text{Var}(x_{ijk} | \mathcal{G})} \quad (9)$$

where $\mathcal{G}$ contains all the parameters of interest. In practice, given a discrepancy measure $T(\mathbf{x}, \mathcal{G})$, the posterior predictive check p -value is calculated as follows:

1. Draw a vector $\mathcal{G}^{(k)}$ from the posterior distribution.
2. Simulate a replicated data set $\mathbf{x}_{rep}^{(k)}$ from $f(\mathbf{x} | \mathcal{G}^{(k)})$.
3. Calculate $T(\mathbf{x}, \mathcal{G}^{(k)})$ and $T(\mathbf{x}_{rep}^{(k)}, \mathcal{G}^{(k)})$.
4. Repeat 1 to 3 K times.
5. Count the proportion of replicated data sets $\mathbf{x}_{rep}^{(k)}$ for which $T(\mathbf{x}_{rep}^{(k)}, \mathcal{G}^{(k)}) \geq T(\mathbf{x}, \mathcal{G}^{(k)})$.

The proportion obtained in 5 is the posterior p -value.

REFERENCES

- Anderson, C., Wasserman, S., & Crouch, B. (1999). A p^* primer: logit models for social networks. *Social Networks*, 21, 37-66.
- Fienberg, S., Meyer, M., & Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80, 51-67.
- Fienberg, S., & Wasserman, S. (1981). Categorical data analysis of single sociometric relations. *Sociological methodology*, 12, 156-192.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). Bayesian data analysis (2nd ed.). London: Chapman and Hall.
- Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 4, 733-807.
- Gill, P., & Swartz, T. (2004). Bayesian analysis of directed graphs data with applications to social networks. *Journal of the Royal Statistical Society, Series C*, 53, 249-260.
- Girard, R. (1961). *Mensogne romantique et vérité romantique*. Paris: Bernard Grasset.
- González, J., De Boeck, P., & Tuerlinckx, F. (2008). A double-structure structural equation model for three-mode data. *Psychological Methods*, 13, 337-353.

APPLIED MULTIVARIATE RESEARCH

- Hoff, P. (2003). Random effects models for network data. In R. Breiger, K. Carley, & P. Pattison (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers* (pp. 303-312). Washington, DC, The National Academies Press.
- Holland, P., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 77, 33-50.
- McCulloch, C., & Searle, S. (2001). *Generalized, linear, and mixed models*. New York: John Wiley and Sons.
- Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52, 169-193.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS version 1.4 user manual [Computer software manual]. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.
- Van Duijn, M., Snijders, T., & Zijlstra, B. (2004). p_2 : a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58, 234-254.
- Wang, Y., & Wong, G. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82, 8-19.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.
- Wasserman, S., & Galaskiewicz, J. (1984). Some generalizations of p_1 : External constraints, interactions, and non-binary relations. *Social Networks*, 6, 177-192.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov random graphs and p^* . *Psychometrika*, 60, 401-426.
- Zijlstra, B., Van Duijn, M., & Snijders, T. (2006). The multilevel p_2 model: A random effects model for the analysis of multiple social networks. *Methodology*, 2, 42-47.

Author Note

Jorge González, Francis Tuerlinckx, Paul De Boeck, Department of Psychology, K.U.Leuven, Belgium. The authors have been partly supported by the Fund for Scientific Research—Flanders (F.W.O.) Grant no. G.0148.04 and by the K.U. Leuven Research Council Grant no. GOA/2005/04.

We thank Els Brusten and Iven Van Mechelen from the Research Group of Quantitative Psychology and Individual Differences, K.U.Leuven, who kindly provided us the data.

Correspondence should be addressed to Jorge González, who is now at the Centro de Medición MIDE UC, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile. Tel. +56 2 3541730, e-mail: jagonzal@uc.cl