# $p_2$: a random effects model with covariates for directed graphs

Marijtje A.J. van Duijn
Tom A.B. Snijders
Bonne J.H. Zijlstra

Department of Sociology/Statistics and Measurement Theory
ICS/Heijmans Institute
University of Groningen

Running head: The $p_2$ model

This is a preprint of an Aricle accepted for publication in Statistica Neerlandica


The authors would like to thank Emmanuel Lazega for his permission to use the data. Requests for reprints can be sent to Grote Rozenstraat 31, 9712 TG Groningen, the Netherlands, or to m.a.j.van.duijn@ppsw.rug.nl

## Abstract

A random effects model is proposed for the analysis of binary dyadic data that represent a social network or directed graph, using nodal and/or dyadic attributes as covariates. The network structure is reflected by modeling the dependence between the relations to and from the same actor or node. Parameter estimates are proposed that are based on an iterated generalized least squares procedure. An application is presented to a data set on friendship relations between American lawyers.



## Keywords

adjacency matrix; dependent binary data; GLMM; IGLS; logistic regression; $p_1$ model; random effects; social network analysis;

# 1 Introduction

Social network analysis is concerned with relations, i.e., patterns of ties between interdependent entities. The special form and structure of the usually binary network data leads to non-standard statistical analysis of social networks. The complicated dependency structures lead to complexities in the statistical modeling.

Important concepts are the reciprocity of relations and the role or position of the entities. These concepts are usually defined endogenously, based on the relations within the social network, but can be extended in a statistical approach using exogenous or explanatory variables.

The statistical model proposed here takes into account the dependent nature of the data and the relation with explanatory variables. It can be seen as an extension of the well-known $p_1$ model for the analysis of complete network data (Holland and Leinhardt, 1981). This extension to a Generalized Linear Mixed Model (GLMM) allows the inclusion of covariates and models the remaining variability by random effects.

After a succinct introduction to social network analysis in the following subsection, the $p_1$ and related models are described in the next two subsections. The $p_2$ model is proposed in Section 2. IGLS estimation of the $p_1$ and $p_2$ model using first-order Taylor approximations is treated in the third section. An overview of GL(M)Ms and estimation methods for binary dependent data developed in other research areas and their relation to the $p_2$ model is given. All these models can be qualified as modified logistic regression models. Model selection and testing procedures are proposed in Section 3 as well. In Section 4 the $p_2$ model is illustrated by a data set on friendship relations between the associate lawyers of an American law firm. We conclude with a discussion of the $p_2$ model and suggest improvements of its estimation.

## 1.1 Social networks and some notation

Here we will give just a brief introduction to social networks. For an elaborate description of social networks, their development, application and analysis, see Wasserman and Faust (1994).

A social network is a finite set of actors and the relation(s) defined on this set (Wasserman and Faust, 1994, p. 20). The actors are social entities (people, organizations, countries, etc.) whose specific ties (friendship, competition, collaboration, etc.), constitute the network. A social network with one directed relation can be represented by a directed graph, or by a matrix known as a sociomatrix or adjacency matrix. For undirected relations the sociomatrix is symmetric. The close association between social network analysis and graph theory is apparent in the terminology. Actors are sometimes called nodes, ties are called lines or arcs and often these terms are used interchangeably, as is also done here.

Assuming that the network consists of $n$ actors, it can be represented by an $n \times n$ matrix of zeros and ones. In this adjacency matrix, denoted by $Y$, each row and the corresponding column represent an actor in the network, or a node in the graph. $Y_{ij}$ is the tie variable from actor $i$ to actor $j$, or the indicator of a directed line from node $i$ to node $j$, that takes on values 1 (tie present) or 0 (tie absent). The pair of tie variables $(Y_{ij}, Y_{ji})$ is called a dyad, that can be null, $(0,0)$, mutual or reciprocal, $(1,1)$, or asymmetric, $(0,1)$ and $(1,0)$. Usually $Y_{ii}$ is defined to be 0, since in most applications self-ties are undefined or non-existent. Ignoring the diagonal, the $n(n-1)$ tie variables or, equivalently, the $n(n-1)/2$ dyads, are considered the complete network data.

Together with attributes of the actors, the data form a so-called social relational system (Wasserman and Faust, 1994, p. 89). We will deal with a social relational system consisting of directed relations and attributes for all actors. The focus is on complete network data, i.e., data representing all existing ties of a given kind within a given set of $n$ actors, but the methods are applicable also if the relational data are incomplete.

## 1.2 The $p_1$ model

Complete network data can be analyzed by many statistical and non-statistical methods and models. For an overview, see Chapters 13 and 15 of Wasserman and Faust (1994).

The model proposed here starts from the so-called $p_1$ model for complete network data, introduced by Holland and Leinhardt (1981). This defines the probability function of each dyad $(Y_{ij}, Y_{ji})$ representing the directed relations between nodes $i$ and $j$ in a complete network consisting of $n$ nodes as

$$P\{Y_{ij} = y_1, Y_{ji} = y_2\} = \exp\{y_1(\mu + \alpha_i + \beta_j) + y_2(\mu + \alpha_j + \beta_i) + y_1 y_2 \rho\}/k_{ij}, \\ y_1, y_2 = 0, 1; i, j = 1, \ldots, n; i \neq j, \quad (1)$$

where

$$k_{ij} = 1 + \exp(\mu + \alpha_i + \beta_j) + \exp(\mu + \alpha_j + \beta_i) + \exp(2\mu + \alpha_i + \beta_j + \alpha_j + \beta_i + \rho). \quad (2)$$

The dyads are assumed to be independent.

The interpretation of the parameters in Holland and Leinhardt's (1981) terminology is that $\mu$ is a density parameter, constituting an overall mean, $\alpha_i$ is a productivity parameter, characterizing $i$ as a sender, $\beta_j$ is an attractiveness parameter, characterizing $j$ as a receiver, and $\rho$ is a parameter indicating the force of reciprocation. The parameters $\alpha_i$ and $\beta_j$ are collected in $n \times 1$ vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Imposing an identification restriction on both vectors, the number of non-redundant parameters is $2n$, which increases with the number of nodes. Holland and Leinhardt called the probability function (1) $p_1$: "... the first or

simplest family of distributions on digraphs that might be considered for social network data. This is because it expresses the two elementary social tendencies of reciprocation and differential attraction." (Holland and Leinhardt, 1981, footnote p. 36). Our model is called $p_2$ since we regard it as a direct successor to the $p_1$ model: it retains reciprocation and differential attraction but now links these concepts to actor and dyadic attributes. Further, the number of statistical parameters is limited by treating the parts of the productivity and attractiveness parameters that are not explained by the covariates as random effects.

## 1.3 Earlier extensions of the $p_1$ model

Following Holland and Leinhardt's pioneering 1981 article, a considerable amount of research has been devoted to the $p_1$ model. Central was – and still is – the question of how to define adequate models and estimation methods for the analysis of network data. Holland and Leinhardt (1981) obtain Maximum Likelihood estimators for the $p_1$ model using generalized iterative scaling. After showing that the ML estimators can be readily obtained from an Iterative Proportional Fitting (IPF) algorithm by redefining the complete network data as a four-dimensional contingency table (Fienberg and Wasserman, 1981), Wasserman has (co-)authored many papers on interesting extensions of $p_1$. An overview of procedures for model fitting and parameter estimation is given in Wasserman and Weaver (1985).

Extensions of $p_1$ to deal with network data of multiple relations estimated with IPF are described in Fienberg, Meyer, and Wasserman (1985). Analysis of network data on valued (discrete, not just binary) relations in the line of the $p_1$ model is proposed by Wasserman and Galaskiewicz (1984) and by Wasserman and Iacobucci (1986). Kenny and La Voie (1984) proposed the Social Relations Model (SRM) that can be viewed as a random effects model for *continuous* social interaction data (e.g., social networks or so-called round robin experiments) with a complex but accurate variance structure (see also Warner, Kenny and Stoto, 1979). Snijders and Kenny (1999) formulate the SRM as a multilevel model with cross-nested random effects, that can be estimated with standard multilevel software like MLwiN (Rasbash, Browne, Healy, Cameron, and Charlton, 2000).

Holland and Leinhardt (1981, p. 34) mention the importance of using available covariates or nodal attributes in the analysis of network data without further suggestions for implementation. Fienberg and Wasserman (1981) incorporate covariates in the $p_1$ model by defining subgroups on the basis of (categorical) nodal attributes and assuming the expansiveness and actractiveness parameters to be equal for actors in the same subgroup. This idea is further developed in the so-called a priori stochastic blockmodels introduced in Holland, Laskey, and Leinhardt (1983) and in a direct extension of the $p_1$ model by Wang and Wong (1987).

The notion that the assumption of dyad independence is strong and often doubtful can already be found in FIENBERG et al. (1985) and in WONG's (1987) Bayesian version of the $p_1$ model. We cannot assume that actor $i$'s relation with actor $j$ is completely independent of his (her) relation with actor $k$, even after taking into account actor attributes. The models for Markov graphs proposed by FRANK and STRAUSS (1986) do express dependence between dyads involving the same actors and more general network dependence structures. The idea is to find those network characteristics that provide sufficient statistics to represent the dependence. Estimating the resulting models, however, is – in general – not straightforward. Among other things, FRANK and STRAUSS (1986) suggest pseudolikelihood estimation where the likelihood function is approximated by the product over all observations of the conditional probability of one observation given all other observations, which method is elaborated in STRAUSS and IKEDA (1990). Models that can be formulated as so-called logit models are estimated relatively easily with logistic regression estimation techniques such as Iterative Generalized Least Squares. This approach was extended to the $p_1$ model by RENNOLLS (1995) and, more completely, by WASSERMAN and PATTISON (1996), see also ANDERSON, WASSERMAN, and CROUCH (1999). The resulting model is a so-called $p^*$ model which takes network dependence into account by the choice of suitable network statistics (to be specified by the researcher). This approach allows estimation of many different models with standard software for logistic regression. Generalizations of $p^*$ to multivariate relations and to valued relations can be found in PATTISON and WASSERMAN (1999) and in ROBINS, PATTISON, and WASSERMAN (1999), respectively.

Even though the parameter estimates obtained with pseudolikelihood are reported for some models to be consistent with the corresponding Maximum Likelihood estimates (STRAUSS and IKEDA, 1990), their standard errors as produced by standard software packages cannot be used because of the non-independence of observations. Different models can be compared by likelihood ratio statistics, but the null distributions of these statistics are as yet unknown. In an application of an autologistic model for the study of spatial patterns of damaged trees, using a model with less far-ranging dependence structure, PREISLER (1993) proposed a solution by using a parametric bootstrap procedure to estimate standard errors. SNIJDERS (2002) tried to solve the estimation problem using Markov Chain Monte Carlo methods by regarding the $p^*$ model as an exponential family for which analytic calculations are impossible but which can be simulated. The $p^*$ model, however, is shown to have convergence problems; alternative conditional estimation methods are proposed by SNIJDERS and VAN DUIJN (2002).

# 2 The $p_2$ model

The aim of the $p_2$ model is to relate binary network data to covariates while taking into account the specific network structure. This requires a kind of bivariate logistic regression model capable of handling the dependence of network data.

The $p_2$ model is based on formula (1). The density and reciprocity parameters, however, are allowed to vary over dyads, and are therefore denoted $\mu_{ij}$ and $\rho_{ij}$. Moreover, parameters $\alpha_i$, $\beta_j$, $\mu_{ij}$, and $\rho_{ij}$ are further modelled using covariates.

We first consider the node-specific parameters $\alpha_i$ and $\beta_j$. Covariates and random effects are included in a linear regression model for the productivity and attractiveness parameters:

$$\boldsymbol{\alpha} = \boldsymbol{X_1\gamma_1} + \boldsymbol{A}, \tag{3}$$

$$\boldsymbol{\beta} = \boldsymbol{X_2\gamma_2} + \boldsymbol{B}. \tag{4}$$

This formulation expresses the plausible idea that attractiveness (or popularity) and productivity (or sociability) depend on actor attributes (denoted by $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively, where the same or different attributes may be used for attractiveness and productivity) with corresponding weights $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. Naturally, the attributes do not explain all variation in attractiveness and productivity parameters, as is represented by the residual terms $\boldsymbol{A}$ and $\boldsymbol{B}$, $n \times 1$ vectors with components $A_i$ and $B_i$. The residuals are modeled as normally distributed random variables with expectation 0 and variances $\sigma_A^2$ and $\sigma_B^2$, respectively. Parameters $\sigma_A^2$ and $\sigma_B^2$ can be interpreted as unexplained variance, that is, the variance of the $\alpha$'s and $\beta$'s that is left after taking into account the effect of the covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The productivity and attractiveness parameters of the same node are correlated: $\text{cov}(A_i, B_i) = \sigma_{AB}$ for all $i$. Independence is assumed for parameters of different actors by setting $\text{cov}(A_i, A_j)$=$\text{cov}(B_i, B_j)$=$\text{cov}(A_i, B_j) = 0$ for $i \neq j$ (cf. Wong, 1987).

If no external information on actors is available, the terms $\boldsymbol{X_1\gamma_1}$ and $\boldsymbol{X_2\gamma_2}$ vanish and $\sigma_A^2$ and $\sigma_B^2$ denote the variances of the $\alpha$ and $\beta$ parameters, respectively. Then a pure random effects model results with, apart from the density and reciprocity parameters, only two variance parameters and one covariance parameter. Thus, the $p_2$ model without covariates is a more parsimonious model than the $p_1$ model with well-interpretable parameters. Obviously, the fit of $p_1$ will be better than the fit of $p_2$ without covariates.

Second, the reduction of parameters enables relaxation of the assumption that the density and reciprocity and parameters are constant over dyads, made by Holland and Leinhardt (1981) for the $p_1$ model for reasons of model identification, and also proposed by Fienberg and Wasserman (1981) and by Wasserman and Galaskiewicz (1984). In the $p_2$ model these parameters

are linearly related to dyadic attributes, denoted by $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$. The density parameters are modelled as

$$\mu_{ij} \quad = \quad \mu + \boldsymbol{Z_{1ij}}\boldsymbol{\delta}_1. \tag{5}$$

Because of the substantive interpretation of reciprocity $\rho$ is assumed to be constant within dyads: $\rho_{ij} = \rho_{ji}$ (cf. Fienberg and Wasserman, 1981), and modelled as

$$\rho_{ij} \quad = \quad \rho + \boldsymbol{Z_{2ij}}\boldsymbol{\delta}_2, \tag{6}$$

where we require $\boldsymbol{Z_{2ij}} = \boldsymbol{Z_{2ji}}$.

Both parameter equations contain a constant (intercept) and a variable part. As can be seen from (5) and (6), different dyadic attributes may be used to model density and reciprocity. From an interpretational point of view, however, it is recommended to choose the dyadic attributes selected to model reciprocity which appear in $\boldsymbol{Z}_2$ from the variables (and their linear combinations) that are used in $\boldsymbol{Z}_1$, to explain density. Thus, it is possible to distinguish a reciprocity effect from the "overall" density effect. To distinguish between the meaning of effects on density and on reciprocity, it should be kept in mind that $\rho_{ij}$ is the log-odds ratio in the $2 \times 2$ table corresponding to the dyad $(Y_{ij}, Y_{ji})$ while the probability that $Y_{ij} = 1$ is an increasing function of $\mu_{ij}$ but also of $\rho_{ij}$; further, $\mu_{ij}$ is the log-odds of $Y_{ij}$ given that $Y_{ji} = 0$ whereas $\mu_{ij} + \rho_{ij}$ is the log-odds of $Y_{ij}$ given that $Y_{ji} = 1$.

In our example analyzing the friendship networks between lawyers we study the effect of "similarity" and "superiority" in terms of seniority, where a low number indicates a high status. Similarity (of dyads) is considered an important determinant for the explanation of density and reciprocity, as has been demonstrated extensively in friendship research (see, e.g., Zeggelink, 1993).

We will define similarity as the absolute difference between two nodal attribute values, and thus in fact define dissimilarity: the larger the value the more dissimilar the two nodes are. (Dis)similarity can be defined more generally as a variable measuring the distance between two nodes with respect to some characteristic. A negative dissimilarity effect on density has the interpretation that, other things being equal, a relation is more probable between similar than between dissimilar actors. A negative dissimilarity effect on reciprocity has the interpretation that similar pairs of actors are more likely to have a mutual or null relation than dissimilar pairs. A positive dissimilarity effect on reciprocity can be understood as reducing the accompanying negative density effect that affects the probability of a mutual relation twice (cf. (1)). We can separate the effect of similarity on the occurrence of reciprocal relations from the effect of similarity on the occurrence of any relation when we use similarity to model both $\mu$ and $\rho$.

Superiority is a non-symmetric variable that, in our application, is defined as the difference between the sender's seniority and the receiver's seniority. This

variable is only suitable for modeling density and not for reciprocity because its value depends on the direction. A positive superiority effect on the density indicates that a relation from a junior to a senior lawyer (indicated by a positive difference in seniority) is more likely than a relation in the opposite direction.

Deriving dyadic covariates from nodal attributes may lead to collinear nodal and dyadic attributes, and therefore requires special attention of the researcher in the model selection process. Dyadic attributes can also be obtained from other available network data, for instance a (symmetrized) network of a different relation.

# 3 Estimation of the $p_2$ model

In this section we start with the IGLS estimation procedure for the $p_1$ model defined as a Generalized Linear Model. From this, the IGLS estimation of the $p_2$ model, a Generalized Linear Mixed Model, involving more complex formulas, follows straightforwardly, as will be shown in the second subsection, followed by a subsection on model selection and testing. A comparison of the $p_2$ model with other GL(M)Ms for dependent binary data and their estimation methods concludes the section.

## 3.1 The $p_1$ model formulated as a GLM

The probability function (1) can be rewritten as the product of two probability functions: the unconditional probability of $Y_{ij}$, and the conditional probability of $Y_{ji}$, given the value of $Y_{ij}$,

$$\begin{aligned} &P\{Y_{ij}=y_1, Y_{ji}=y_2\} = P\{Y_{ij}=y_1\}P\{Y_{ji}=y_2|Y_{ij}=y_1\} = \\ &(\exp\{y_1(\mu+\alpha_i+\beta_j)\}+\exp\{y_1(\mu+\alpha_i+\beta_j+\rho)+\mu+\alpha_j+\beta_i\})/k_{1ij}\times \\ &\exp\{y_2(\mu+\alpha_j+\beta_i+y_1\rho)\}/k_{2ji}, \end{aligned} \tag{7}$$

with

$$k_{1ij} = k_{ij} \text{ as in (2)}$$

and

$$k_{2ji} = 1+\exp(\mu+\alpha_j+\beta_i+y_1\rho).$$

The order of $i$ and $j$ is arbitrary. In Bonney's (1987) regressive logistic model for dependent binary observations, each of the conditional probabilities is assumed to be logistic. For the $p_1$ model only the second (conditional) component has a logistic structure.

The dyads can be modeled using the expected values (denoted by $\mathcal{E}$) of their components:

$$\left\{ \begin{array}{lcl} Y_{ij} & = & \mathcal{E}(Y_{ij}) + E_{1ij} \\ Y_{ji} & = & \mathcal{E}(Y_{ji}|Y_{ij}) + E_{2ji}, \end{array} \right. \tag{8}$$

where

$$\begin{array}{rcl} \mathcal{E}(Y_{ij}) = P\{Y_{ij} = 1\} & = & (\exp(\mu + \alpha_i + \beta_j) + \\ & & + \exp(2\mu + \alpha_i + \alpha_j + \beta_i + \beta_j + \rho))/k_{1ij}, \end{array} \tag{9}$$

and

$$\mathcal{E}(Y_{ji}|Y_{ij} = y_1) = P\{Y_{ji} = 1|Y_{ij} = y_1\} = \exp(\mu + \alpha_j + \beta_i + y_1\rho)/k_{2ji}. \tag{10}$$

This implies that $\mathcal{E}(E_{1ij}) = \mathcal{E}(E_{2ji}) = 0$. Note that $\mathrm{var}(E_{1ij}) = \mathrm{var}(Y_{ij})$; $\mathrm{var}(E_{2ji}) = \mathrm{var}(Y_{ji}|Y_{ij})$.

Define $\boldsymbol{\theta}_{1ij} = (\mu, \rho, \alpha_i, \alpha_j, \beta_i, \beta_j)$ and $\boldsymbol{\theta}_{2ji} = (\mu, \rho, \alpha_j, \beta_i)$. Let $F_1(\boldsymbol{\theta}_{1ij})$ denote $\mathcal{E}(Y_{ij})$, and $F_2(\boldsymbol{\theta}_{2ji}, y_1)$ denote $\mathcal{E}(Y_{ji}|Y_{ij} = y_1)$. Because the values of the dyad elements are 0 and 1, their variances can be expressed as

$$\mathrm{var}(Y_{ij}) = F_1(\boldsymbol{\theta}_{1ij})(1 - F_1(\boldsymbol{\theta}_{1ij})) \tag{11}$$

and

$$\mathrm{var}(Y_{ji}|Y_{ij} = y_1) = F_2(\boldsymbol{\theta}_{2ji}, y_1)(1 - F_2(\boldsymbol{\theta}_{2ji}, y_1)). \tag{12}$$

This Generalized Linear Model can be estimated using Iterative Generalized (or Weighted) Least Squares (IGLS), yielding Maximum Likelihood estimators of $\boldsymbol{\theta} = (\mu, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta})$ (see, e.g. McCullagh and Nelder, 1989, Section 2.5). Identification restrictions are needed for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, e.g., $\alpha_1 = \beta_1 = 0$ or $\sum_i \alpha_i = \sum_j \beta_j = 0$ as in Holland and Leinhardt (1981).

The general IGLS estimation algorithm alternately computes first $\boldsymbol{\theta}$ given the current value of the variances (11) and (12), and then these variances given the residuals obtained in the first step. IGLS for non-linear models requires an extra step: updating the linearization of the non-linear link functions $F_1(\boldsymbol{\theta}_{1ij})$ given in (10) and $F_2(\boldsymbol{\theta}_{2ji}, y_1)$ given in (10).

The linearization at each iteration step is based on a first-order Taylor expansion (omitting subscripts of $F$ and $\boldsymbol{\theta}$) around the current point $\boldsymbol{\theta}^0$ :

$$F(\boldsymbol{\theta}) \approx F(\boldsymbol{\theta}^0) + \frac{\partial F}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \, . \tag{13}$$

We introduce the following matrix notation, with $\boldsymbol{D} = \partial F/\partial \boldsymbol{\theta} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$ to explicate the estimation process sketched above, using (8) rewritten as $\boldsymbol{Y} = \boldsymbol{F}(\boldsymbol{\theta}) + \boldsymbol{E}$ and (13):

$$\boldsymbol{Y} - \boldsymbol{F}(\boldsymbol{\theta}^0) + \boldsymbol{D}\boldsymbol{\theta}^0 \approx \boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{E}, \tag{14}$$

where

- The $n(n-1)$-dimensional vector $\boldsymbol{Y}$ contains both $Y_{ij}$ and $Y_{ji}$ for all $n(n-1)/2$ dyads. The order of $i$ and $j$ is arbitrary and does not influence the results.

- $\boldsymbol{F}(\boldsymbol{\theta}^0)$ is the vector consisting of $F_1(\boldsymbol{\theta}_{1ij})$ and $F_2(\boldsymbol{\theta}_{2ji}, y_1)$ corresponding to $\boldsymbol{Y}$, cf. (8).

- Matrix $\boldsymbol{D}$ contains all partial first derivatives evaluated in $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. For each of the unconditional relations two sender and two receiver effects and for each conditional relation one sender and one receiver effect are included. The partial derivatives are given in the Appendix.

- $\boldsymbol{E}$ is a vector containing the disturbance or error terms $E_{1ij}$ and $E_{2ji}$ in (8) with variances $V_{1ij}$ as in (11), and $V_{2ji}$ as in (12), respectively. The dyad independence implies that all elements of $\boldsymbol{E}$ are uncorrelated. We denote the covariance matrix of $\boldsymbol{E}$ by $\boldsymbol{V}$. $\boldsymbol{V}$ is a square $n(n-1)$ matrix with the appropriate expression $V_{1ij}$ or $V_{2ji}$ on the main diagonal. All off-diagonal elements of $\boldsymbol{V}$ are 0.

It is not necessary to have complete information on all dyads. Missing dyads, or dyads with only one tie variable are allowed. The dimension of $\boldsymbol{Y}$, $\boldsymbol{F}(\boldsymbol{\theta}^0)$, $\boldsymbol{D}$, $\boldsymbol{E}$, and $\boldsymbol{V}$ is then equal to the number of available directed relations.

In each iteration cycle, a GLS regression is performed where the left hand side of (14) is used as the dependent variable which we denote by $\breve{\boldsymbol{Y}}$. The "explanatory" variables are represented by $\boldsymbol{D}$. The iteration steps in the non-linear IGLS algorithm are the following.

1. In each iteration the GLS estimator for $\boldsymbol{\theta}$ is computed, given $\boldsymbol{D}$ and $\boldsymbol{V}$ obtained in the previous iteration. The GLS estimator obtained in the first step of the $(t+1)$th iteration is

$$\hat{\boldsymbol{\theta}}_{t+1} = (\boldsymbol{D}_t' \hat{\boldsymbol{V}}_t^{-1} \boldsymbol{D}_t)^{-1} \boldsymbol{D}_t' \hat{\boldsymbol{V}}_t^{-1} \breve{\boldsymbol{Y}}_t, \tag{15}$$

   where $\hat{\boldsymbol{D}}_t$ and $\hat{\boldsymbol{V}}_t$ are the matrices with parameters according to the previous iteration $\hat{\boldsymbol{\theta}}_t$.

2. In the next step $\hat{\boldsymbol{V}}_{t+1}$ is calculated.

3. In the third and last step the explanatory variables $\boldsymbol{D}$ are first updated and then the dependent variable $\breve{\boldsymbol{Y}}$.

Convergence is reached if the difference between subsequent elements of $\hat{\boldsymbol{\theta}}$ is sufficiently small. The covariance matrix of $\hat{\boldsymbol{\theta}}$ is estimated as $(\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}$, evaluated in the found solution.

## 3.2 IGLS estimation of the $p_2$ model

The $p_2$ model is derived from the $p_1$ model by substituting $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mu$, and $\rho$ by the regression equations (3), (4), (5), and (6), respectively. The presence of random effects ($\boldsymbol{A}$ and $\boldsymbol{B}$) together with the "fixed" regression coefficients ($\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\delta}_1$, and $\boldsymbol{\delta}_2$) renders a Generalized Linear Mixed Model (GLMM) that can be estimated with IGLS, analogous to the IGLS estimation of the $p_1$ model presented in the previous subsection. We use a similar approach as GOLDSTEIN (1991) in his treatment of non-linear multilevel models, but with a more complicated covariance structure (because of – in multilevel terminology – crossed instead of nested random effects).

This procedure is presented after defining the $p_2$ model in accordance with (8):

$$\left\{ \begin{array}{rcl} Y_{ij} & = & F_1(\boldsymbol{\theta}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{1j}, \boldsymbol{X}_{2i}, \boldsymbol{X}_{2j}, \boldsymbol{Z_{1ij}}, \boldsymbol{Z_{2ij}}, A_i, A_j, B_i, B_j) + E_{1ij} \\ Y_{ji} & = & F_2(\boldsymbol{\theta}, \boldsymbol{X}_{1j}, \boldsymbol{X}_{2i}, \boldsymbol{Z_{1ji}}, \boldsymbol{Z_{2ji}}, y_1, A_j, B_i) + E_{2ji} \end{array} \right. , \quad (16)$$

with $\boldsymbol{\theta} = (\mu, \rho, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$. $F_1$ and $F_2$ are the conditional expected values of $Y_{ij}$ and $Y_{ji}$, respectively, given the values of $\boldsymbol{A}$ and $\boldsymbol{B}$. Their exact formulation can be derived from (10) and (10) substituting $\alpha_i, \alpha_j, \beta_i, \beta_j, \mu$ and $\rho$ by (3), (4), (5), and (6). The (now both conditional) variances of $Y_{ij}$ and $Y_{ji}$ are again of the binomial form $F(1-F)$.

Following GOLDSTEIN (1991), the functions $F_1$ and $F_2$ are linearized applying a first-order Taylor expansion, as was done for the $p_1$ model, cf. (13). The expansion is not just around the fixed effects parameter $\boldsymbol{\theta}$ (in an arbitrary point), but also around the random parameters $\boldsymbol{A}$ and $\boldsymbol{B}$ in 0 (their expected value). Define this point as $\boldsymbol{\theta}^0 = (\mu^0, \rho^0, \boldsymbol{\gamma}_1^0, \boldsymbol{\gamma}_2^0, \boldsymbol{\delta}_1^0, \boldsymbol{\delta}_2^0)$, and let $F_1(\boldsymbol{\theta}^0, \boldsymbol{X}, \boldsymbol{Z})$ denote $F_1(\boldsymbol{\theta}^0, \boldsymbol{X}_{1i}, \boldsymbol{X}_{1j}, \boldsymbol{X}_{2i}, \boldsymbol{X}_{2j}, \boldsymbol{Z_{1ij}}, \boldsymbol{Z_{2ij}}, 0, 0, 0, 0)$, while $F_2(\boldsymbol{\theta}^0, \boldsymbol{X}, \boldsymbol{Z})$ denotes $F_2(\boldsymbol{\theta}^0, \boldsymbol{X}_{1j}, \boldsymbol{X}_{2i}, \boldsymbol{Z_{1ji}}, \boldsymbol{Z_{2ji}}, y_1, 0, 0)$. Omitting again all subscripts for notational ease, the approximating equation used for estimation is (cf. (14)):

$$\boldsymbol{Y} - F(\boldsymbol{\theta}^0, \boldsymbol{X}, \boldsymbol{Z}) + \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\theta}^0 \approx \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\theta} + \boldsymbol{C}\boldsymbol{U} + \boldsymbol{E}. \quad (17)$$

$\boldsymbol{Y}$ is defined as before. $\boldsymbol{D}$ is again the covariate matrix with all partial derivatives, now a function of $\boldsymbol{X}$ and $\boldsymbol{Z}$, $\boldsymbol{D(X, Z)} = \boldsymbol{\partial F(\theta, X, Z)/\partial \theta}\,\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$ (see also the Appendix). $\boldsymbol{X}$ and $\boldsymbol{Z}$ denote the matrices containing the sender and receiver covariates, and density and reciprocity covariates, respectively. $\boldsymbol{U}$ is the stacked vector of $\boldsymbol{A}$ and $\boldsymbol{B}$. Its covariance matrix is given by

$$\boldsymbol{\Sigma_U} = \left( \begin{array}{c|c} \sigma_A^2 \boldsymbol{I} & \sigma_{AB} \boldsymbol{I} \\ \hline \sigma_{AB} \boldsymbol{I} & \sigma_B^2 \boldsymbol{I} \end{array} \right),$$

where $\boldsymbol{I}$ is an $n \times n$ identity matrix. $\boldsymbol{C}$ is a design matrix for $\boldsymbol{U}$ providing the appropriate random terms for $\boldsymbol{Y}$, $\boldsymbol{C} = \boldsymbol{\partial F(\theta, X, Z, U)/\partial U}\,\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$, where

$F(\theta, X, Z, U)$ represents the two elements of (16). $\boldsymbol{E}$ is defined as before, now with covariance matrix $\boldsymbol{\Sigma_E}$.

The dependent variable $\breve{\boldsymbol{Y}}$ is the left hand side of (17). Its variance can now be expressed as

$$\boldsymbol{V} = \boldsymbol{C}\boldsymbol{\Sigma_U}\boldsymbol{C}' + \boldsymbol{\Sigma_E}.$$

Note that the order of nodes $i$ and $j$ does make a difference in view of the asymmetric definition of $F_1$ and $F_2$ in the random terms $\boldsymbol{A}$ and $\boldsymbol{B}$. Therefore, it seems sensible to order the pairs of actors such that actors are (almost) as often node $i$ as they are node $j$.

The IGLS iteration process consists again of three steps. In the first step the "fixed" effects parameter vector is estimated (cf. (15))

$$\hat{\boldsymbol{\theta}}_{t+1} = (\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})_t' \hat{\boldsymbol{V}}_t^{-1} \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})_t)^{-1} \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})_t' \hat{\boldsymbol{V}}_t^{-1} \breve{\boldsymbol{Y}}_t.$$

In the second step $\boldsymbol{V}$ is estimated in the way proposed by Goldstein (1986) which is similar to a modified Hildreth-Houck model known in the econometric literature (see, e.g., Maddala (1977), Ch. 17). A regression model is formulated for the parameters representing the dependence structure, that is the elements of $\boldsymbol{\Sigma_E}$ and $\boldsymbol{\Sigma_U}$:

$$\boldsymbol{Y}^* = \boldsymbol{X}^* \boldsymbol{v}^* + \boldsymbol{e}^*,$$

where $\boldsymbol{Y}^* = vec((\breve{\boldsymbol{Y}} - \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\theta})(\breve{\boldsymbol{Y}} - \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\theta})')$ consists of the residuals of the first step, and $\boldsymbol{v}^* = (\sigma_A^2, \sigma_B^2, \sigma_{AB})$ contains the parameters to be estimated. $\boldsymbol{X}^*$ is a design matrix corresponding to the parameters in $\boldsymbol{v}^*$, whose elements are given in the Appendix.

Then, the GLS estimator for the parameters of the random part is

$$\boldsymbol{v}^* = (\boldsymbol{X}^{*'} \boldsymbol{V}^{*-1} \boldsymbol{X}^*)^{-1} \boldsymbol{X}^{*'} \boldsymbol{V}^{*-1} \boldsymbol{Y}^*, \tag{18}$$

where $\boldsymbol{V}^*$, a square $n^2(n-1)^2$ matrix, is the covariance matrix of $\boldsymbol{e}^*$. In the normal case, $\boldsymbol{V}^* = \boldsymbol{V} \otimes \boldsymbol{V}$, and then the GLS estimator is equal to the ML estimator (Goldstein, 1986).

The third iteration step consists of updating $\breve{\boldsymbol{Y}}$, $\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$, and $\boldsymbol{V}$.

The estimation process is much more computationally demanding for $p_2$ than for $p_1$, because of the second step in the iteration process which involves large matrices and the time consuming computation of the inverse of the symmetric matrix $\boldsymbol{V}$. The $n(n-1)$ square matrix $\boldsymbol{V}$ can be broken down using the decomposition formula (cf. e.g. Goldstein, 1986, App. 1)

$$\begin{aligned} \boldsymbol{V}^{-1} &= (\boldsymbol{C}\boldsymbol{\Sigma_U}\boldsymbol{C}' + \boldsymbol{\Sigma_E})^{-1} \\ &= \boldsymbol{\Sigma_E}^{-1} - \boldsymbol{\Sigma_E}^{-1}\boldsymbol{C}\boldsymbol{\Sigma_U}\boldsymbol{G}^{-1}\boldsymbol{C}'\boldsymbol{\Sigma_E}^{-1}, \end{aligned} \tag{19}$$

where $\boldsymbol{G} = \boldsymbol{I} + \boldsymbol{C}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{C}\boldsymbol{\Sigma_U}$ is a $2n \times 2n$ matrix (instead of $n(n-1) \times n(n-1)$), reducing the computational burden considerably. More information is given in the Appendix.

In the open software system StOCNET (Boer, Huisman, Snijders, and Zeggelink, 2003), free at the web (http://stat.gamma.rug.nl/stocnet), a $p_2$ module (Zijlstra and Van Duijn, 2003) is available that performs the IGLS estimation of the $p_2$ model sketched above.

## 3.3 Model selection and testing

From the model estimation process several statistics are obtained that may be used for model selection and parameter testing, such as the value of the likelihood function (or deviance) and the standard errors (or covariance matrix) of the parameter estimates. The problem with these measures is that they are based on the Taylor-expansion of the likelihood function. The quality of this approximation is unknown and may vary from model to model (see also Rodríguez and Goldman, 1995), in particular for models with many parameters, making the usual forward or backward testing procedures also more difficult to perform. The likelihood is exact if the variances of the random effects, $\sigma_A^2$ and $\sigma_B^2$, are zero.

Since Likelihood Ratio tests are not very reliable with these approximated likelihoods (Goldstein, 1995, p. 103), our proposal is to use Wald tests for the testing of parameters and to use a careful model selection procedure identifying important (possibly significant) covariates to be tested simultaneously in backward selection steps.

The Wald test statistic $\boldsymbol{W}$ testing the hypothesis that $\boldsymbol{\theta} = \mathbf{0}$ is given by (see, e.g., Serfling, 1980, p. 157)

$$\boldsymbol{W} = \hat{\boldsymbol{\theta}}' \hat{\boldsymbol{V}}^{-1} \hat{\boldsymbol{\theta}},$$

where $\hat{\boldsymbol{V}}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$ which is computed according to the approximation in the preceding subsection. $\boldsymbol{W}$ has an approximate $\chi^2$ distribution with the dimension of $\boldsymbol{\theta}$ as the number of degrees of freedom. $\boldsymbol{W}$ reduces to the well-known $t$-statistic for one-dimensional $\boldsymbol{\theta}$.

## 3.4 Relation of the $p_2$ model to other GL(M)Ms

Defining the $p_1$ and $p_2$ models as GLM and GLMM respectively, puts them in a long and rich tradition of models for binary data in which the assumption of independent pairs is inappropriate. In this section we will give a short overview of related Generalized Linear (Mixed) Models, most of which have been developed for medical applications. The applicability of most models discussed is not limited to pairs of observations but can be extended to longitudinal data.

Rosner (1982) showed that ignoring the interdependence of paired data (following a normal or binomial distribution) leads to invalid results. He developed a polytomous logistic regression model that may be viewed as a reparametrization of the $p_1$ model assuming equal sender and receiver effects for every actor

(Rosner, 1984). The model can be extended with covariates specific for the pair and its elements and is estimated with an iterative Newton-Raphson procedure.

The regressive logistic models proposed by Bonney (1987) also include the $p_1$ model with equal sender and receiver effects as a special case. These models are explicitly designed for sequentially ordered binary data as the key of this approach is the conditioning of one binary outcome on the other (preceding) binary outcomes. By clever parametrization a logistic regression model can be formulated for the conditional observations. This model can be estimated with standard computer programs for logistic regression.

Connolly and Liang (1988) generalized Rosner's (1984) model to conditional logistic regression models for clusters of binary data, that is a logistic model for one observation conditional on the other observation. The model includes covariates (for each observation) and the parameters of its pseudolikelihood function are estimated using estimating functions.

Prentice (1988) points out that the conditional models are most appropriate for exploring the dependence between the observations within a pair or cluster, a point elaborated by Neuhaus and Jewell (1990). Alternatively, Prentice (1988) derives a joint model (with covariates for each observation) that also allows for regression on the marginal expectations. It can be estimated with generalized estimating equations. He also mentions the mixture model to accommodate pair or cluster specific covariates. This approach is taken by Stiratelli, Laird and Ware (1984) where estimation is feasible with the EM algorithm (see also Anderson and Aitkin, 1985). The resulting random effects model accommodating the dependence through random regression coefficients is related to the models of Goldstein (1991) and Longford (1994) that are estimated with Iterative Generalized Least Squares (IGLS) and Fisher scoring, respectively.

Many improvements of and alternatives to the earlier mentioned estimation methods have been proposed: Restricted Maximum Likelihood (REML) by Schall (1991); MCMC estimation using the Gibbs sampler by Zeger and Karim (1991), see also Browne and Draper (2003); Penalized Quasi-Likelihood (PQL) by Breslow and Clayton (1993), based on an approximation of the unconditional likelihood; second order approximations by (Goldstein (1994) and Goldstein and Rasbash (1996); numerical integration by Hedeker and Gibbons (1997), see also Rabe-Hesketh, Pickles and Skondral (2001); stochastic EM by McCulloch (1997); nonparametric (latent class) EM by Aitkin (1999); and Laplace approximation by Raudenbush, Yang, and Yosef (2000).

# 4 Data and analysis

To illustrate the $p_2$ model we use data collected by Lazega (1992, 1995, 2001) in a Northeastern US law firm on informal relationships between 71 lawyers,

divided in 36 partners and their 35 associates. Complete networks are available on advice, collaboration, and friendship relations that are defined as binary directed relational variables. Using the p2 module version 2.0.0.7 (Zijlstra and Van Duijn, 2003) available in StOCNET version 1.4 (Boer et al., 2003), we analyze the friendship network between the 35 associates of the law firm, consisting of 595 dyads. The overall density of the friendship network of associates, defined as the percentage of present directed friendship relations, is 0.153. Of the potential 1190 relations, 182 are present, in 62 mutual and 58 asymmetric dyads. Analyses with the $p_2$ model of the advice relationships between partners, between associates, and between all lawyers, can be found in Lazega and Van Duijn (1997).

Several actor covariates are available expressing the formal structure of the organization: location (the firm has three offices in three different cities, 26 of them are in the main office), practice specialty (21 are litigators; 14 corporate lawyers), and seniority (the associates are subdivided into 5 groups corresponding to time of entrance in the firm). Other available actor characteristics are gender (20 men and 15 women), age (ranging from 26 to 53 years with mean 35), and lawschool attended (17 lawyers had their training at the University of Connecticut). From these covariates, similarity variables are derived to explain density and reciprocity. From seniority a "superiority" variable is derived that can be used to model the density parameter (see Section 2). Gender, age, and seniority are used as sender and receiver effects. The networks on advice and collaboration are also available as covariates for the density parameter, although we have to be cautious to use the advice and collaboration data as explanatory variables for the friendship relation because these networks are outcome variables themselves.

- - - insert TABLE 1 here - - -

First an "empty" model is estimated giving overall estimates of $\mu$ and $\rho$ and estimates of the variance parameters, presented as Model 0 in Table 1. No $p$-values are reported for the parameters of the empty model (nor of these parameters in the two subsequent models), since they cannot be left out of the $p_2$ model. In Model 0, the variance of "sending" friendship seems to be somewhat larger than the "receiving" friendship variance. A small negative covariance is observed. The density parameter $\mu$ is negative, indicating a network density smaller than 0.5; the reciprocity parameter $\rho$ is positive, as expected.

Next, all covariates are added to the model one by one, leading to a series of models with one or two (in case of reciprocity which is accompanied by the corresponding density) parameters, not shown here. Location and gender turn out to be important density variables (i.e. as dissimilarity variables). For location also a weak reciprocity effect is found. Seniority is important both as a density variable as well as a sending variable (young associates tend to send more friendship relations). A positive density effect of "superiority" is found. Finally, a positive receiver effect for age is found.

Then a model is estimated that contains all variables found to be important in the previous step. In a backward selection procedure discarding all non-significant effects, a first model is obtained (Model 1). Table 1 shows the parameter estimates of Model 1, their standard errors, and a two-sided $p$-value based on the Wald statistic. In this model, the superiority effect and the dissimilarity density effects of location, seniority, gender and specialty are retained. The density of friendships is smallest among associates of different seniority and gender who do not work in the same office and who have a different practice. A friendship relation is more likely from a junior associate to a senior associate than vice versa. Having one of the above characteristics in common increases the density. A reciprocity effect of similarity of location is included in the model, mainly for illustrational purposes. The positive parameter estimate implies that the negative dissimilarity effect of location is reduced considerably for mutual dyads.

Finally, the advice and collaboration networks are added to Model 1 as explanatory variables. Again, a backward selection procedure is applied, resulting in Model 2 presented in Table 1. Similarity with respect to office, seniority and gender are still quite important, but advice is now the variable with the largest parameter estimate. The similarity effect of location is reduced because its corresponding reciprocity effect is not included in the model. The superiority effect and the dissimilarity effect of specialty are no longer included in the model. These effects become insignificant when advice is included in the model. Advice takes over the role of superiority and specialty similarity. This can be explained by the fact that both both variables contribute to the "explanation" of advice relationships between the associates (cf. Lazega and Van Duijn, 1997).

In both Models 1 and 2 no explanatory variables for the sender and receiver effects are included. It is observed that the sender and the receiver variances are still of the same order of size while the covariance is almost zero in Model 1 and slightly positive in Model 2, implying considerable differences between the friendship sending and receiving behavior of individual associates. Apparently, this heterogeneity could not be explained by the available covariates characterizing mostly the formal structure.

# 5 Summary and discussion

The $p_2$ model was presented as an extension of the $p_1$ model with nodal and dyadic attributes and random effects. Both models were formulated as Generalized Linear Models and an algorithm to perform IGLS estimation with a first-order Taylor approximation was proposed.

In line with the $p_1$ model, the $p_2$ model assumes complete network data, although this is not required for the estimation of the model which can handle dyads of which one or both relations are missing, assuming missingness at random. Missing one or more attributes of a certain actor implies deletion of all dyadic

relations involving that actor; likewise, missing dyadic attributes implies deletion of the corresponding dyadic relation.

The most important difference of the $p_2$ model with other models for dependent binary data is its cross-nested random structure coming from the network in which relations to and from the same actor are assumed to be related. It is exactly this assumption of "dyad dependence" that makes the $p_2$ model so interesting, but at the same time difficult to estimate. Although the IGLS estimation of the first-order Taylor approximation of the non-linear $p_2$ model can be viewed as relatively straightforward, it is easy to criticize. It is a so-called first-order Marginal Quasi Likelihood (MQL) estimation method, which Breslow and Clayton (1993) and Rodríguez and Goldman (1995, 2001) demonstrated to underestimate the variance parameters and the fixed effects, especially if the variance components are large. A second-order approximation together with PQL improves the estimates considerably (Goldstein, 1994; Goldstein and Rasbash, 1996). Further research about higher order approximations for the $p_2$ model would be relevant.

Aitkin (1999) proposed an elegant, nonparametric approximation method which, however, cannot be adapted for the $p_2$ model because of the cross-nested random effects. Probably the best way to improve the estimation of the $p_2$ model will be via exact, although computationally intensive, MCMC methods, following Browne and Draper (2003). Such an approach is expected to also eliminate the sensitivity of the current IGLS method to the order of the dyad elements.

Wright (1997) notices the problem of extra-binomial variation for binary data with a hierarchical structure, especially if the data are sparse, that is with relatively few observations per cluster. The $p_2$ model does not allow for extra-binomial variation, but this could be easily incorporated by adding a dispersion parameter to the variance equations (11) and (12).

Similarly, extensions to binary data with a different dependence structure are feasible. Especially extensions to more complex variance structures are considered important. These may arise from the availability of multiple relations, e.g., in the simultaneous analysis of the friendship, advice, and collaboration networks, or from the presence of multiple networks, e.g. the friendship networks in various companies .

The approximate nature of the estimation method is also apparent in the model selection process, which has to be performed with even more care than usual. We cannot use likelihood ratio or deviance tests, but have to resort to Wald tests. With improvement of the estimation method, the model testing may be improved as well.

Even in view of these limitations, we consider the model to be a suitable and interesting extension of the $p_1$ model, as is illustrated in its application to the analysis of informal networks in a law firm.

# Appendix

## First derivatives of $F_1$ and $F_2$

First some additional notation is introduced to simplify the expressions. Define $a = \exp(\mu + \alpha_i + \beta_j)$, $b = \exp(\mu + \alpha_j + \beta_i)$, and $c = \exp(\rho)$ , then (10) can be rewritten as

$$F_1(\boldsymbol{\theta}_{1ij}) = \frac{a(1+bc)}{1+a+b+abc}.$$

The first partial derivatives can now be expressed as

$$\begin{aligned}
\frac{\partial F_1}{\partial \mu} &= \frac{2abc + a + ab^2c}{(1+a+b+abc)^2} \\
\frac{\partial F_1}{\partial \alpha_i} = \frac{\partial F_1}{\partial \beta_j} &= \frac{a(1+bc)(1+b)}{(1+a+b+abc)^2} \\
\frac{\partial F_1}{\partial \alpha_j} = \frac{\partial F_1}{\partial \beta_i} &= \frac{ab(c-1)}{(1+a+b+abc)^2} \\
\frac{\partial F_1}{\partial \rho} &= \frac{abc(1+b)}{(1+a+b+abc)^2}.
\end{aligned}$$

Further, defining $d$ as $\exp(y_1\rho)$, (10) can be written as

$$F_2(\boldsymbol{\theta}_{2ji}, y_1) = \frac{bd}{1+bd}.$$

The first partial derivatives are then

$$\begin{aligned}
\frac{\partial F_2}{\partial \mu} &= \frac{\partial F_2}{\partial \alpha_j} = \frac{\partial F_2}{\partial \beta_i} = \frac{bd}{(1+bd)^2} \\
\frac{\partial F_2}{\partial \rho} &= \frac{y_1 bd}{(1+bd)^2}.
\end{aligned}$$

These are all first partial derivatives needed for the $p_1$ model. From these expressions the partial derivates for the $p_2$ model are derived easily (leaving out the subscripts):

$$\begin{aligned}
\frac{\partial F}{\partial \boldsymbol{\gamma}_1} &= \boldsymbol{X}_1 \frac{\partial F}{\partial \boldsymbol{\alpha}} \\
\frac{\partial F}{\partial \boldsymbol{\gamma}_2} &= \boldsymbol{X}_2 \frac{\partial F}{\partial \boldsymbol{\beta}} \\
\frac{\partial F}{\partial \boldsymbol{\delta}_1} &= \boldsymbol{Z}_1 \frac{\partial F}{\partial \mu} \\
\frac{\partial F}{\partial \boldsymbol{\delta}_2} &= \boldsymbol{Z}_2 \frac{\partial F}{\partial \rho}.
\end{aligned}$$

## Building blocks of $\boldsymbol{v}^*$

The elements of (18) can be obtained relatively easily, using the following matrix properties, (see, e.g. Goldstein, 1986, Goldstein and Rasbash, 1992, and Searle, 1982), and assuming the matrices $M, N, O, P$ have the correct dimensions:

$$(M \otimes N)^{-1} = M^{-1} \otimes N^{-1};$$

$$(vec(P))'(M^{-1} \otimes N^{-1})vec(Q) = tr(P'M^{-1}QN^{-1});$$

$$tr(PQ) = tr(QP),$$

where $tr$ is the trace operator (summing the main diagonal elements).

Let $\boldsymbol{C_A}$ contain the first $n$ columns of $\boldsymbol{C}$ (corresponding to $\boldsymbol{A}$) and $\boldsymbol{C_B}$ its last $n$ columns $\boldsymbol{C}$ (corresponding to $\boldsymbol{B}$). Then $\boldsymbol{X}^*$ can be rewritten as

$$\boldsymbol{X}^* = (vec(\boldsymbol{C_A C_A}'), vec(\boldsymbol{C_B C_B}'), vec(2\boldsymbol{C_A C_B}')).$$

Let

$$\boldsymbol{R} = \breve{\boldsymbol{Y}} - \boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\theta},$$

then

$$\boldsymbol{Y}^* = vec(\boldsymbol{R}\boldsymbol{R}').$$

The evaluation of the right hand side of (18) is broken down in two steps. First $\boldsymbol{X}^{*'}\boldsymbol{V}^{*-1}\boldsymbol{Y}^*$ is computed. Using the symbols introduced before, we get

$$\boldsymbol{X}^{*'}\boldsymbol{V}^{*-1}\boldsymbol{Y}^* = \begin{pmatrix} (vec(\boldsymbol{C_A C_A}'))'(\boldsymbol{V} \otimes \boldsymbol{V})^{-1}vec(\boldsymbol{R}\boldsymbol{R}') \\ (vec(\boldsymbol{C_B C_A}'))'(\boldsymbol{V} \otimes \boldsymbol{V})^{-1}vec(\boldsymbol{R}\boldsymbol{R}') \\ 2(vec(\boldsymbol{C_A C_B}'))'(\boldsymbol{V} \otimes \boldsymbol{V})^{-1}vec(\boldsymbol{R}\boldsymbol{R}') \end{pmatrix}.$$

This expression can be rewritten, using $\boldsymbol{V}^{*-1} = \boldsymbol{V}^{-1} \otimes \boldsymbol{V}^{-1}$ and applying the matrix properties stated above, as

$$\begin{pmatrix} (\boldsymbol{C_A}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{R} - \boldsymbol{C_A}'\boldsymbol{H}\boldsymbol{R})'(\boldsymbol{C_A}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{R} - \boldsymbol{C_A}'\boldsymbol{H}\boldsymbol{R}) \\ (\boldsymbol{C_B}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{R} - \boldsymbol{C_B}'\boldsymbol{H}\boldsymbol{R})'(\boldsymbol{C_B}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{R} - \boldsymbol{C_B}'\boldsymbol{H}\boldsymbol{R}) \\ 2(\boldsymbol{C_A}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{R} - \boldsymbol{C_A}'\boldsymbol{H}\boldsymbol{R})'(\boldsymbol{C_B}'\boldsymbol{\Sigma_E}^{-1}\boldsymbol{R} - \boldsymbol{C_B}'\boldsymbol{H}\boldsymbol{R}) \end{pmatrix},$$

where

$$\boldsymbol{H} = \boldsymbol{\Sigma_E}^{-1}\boldsymbol{C}\boldsymbol{\Sigma_U}\boldsymbol{G}^{-1}\boldsymbol{C}'\boldsymbol{\Sigma_E}^{-1},$$

with

$$\boldsymbol{G} = \boldsymbol{I} + \boldsymbol{C}'{\boldsymbol{\Sigma}_{\boldsymbol{E}}}^{-1}\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{U}}.$$

For the actual computation the $2n$-dimensional vector

$$\boldsymbol{C}'{\boldsymbol{\Sigma}_{\boldsymbol{E}}}^{-1}\boldsymbol{R} - \boldsymbol{C}'\boldsymbol{H}\boldsymbol{R}$$

is evaluated. The first element of $\boldsymbol{X}^{*'}\boldsymbol{V}^{*-1}\boldsymbol{Y}^{*}$ is then the sum of the squared first $n$ elements. Its second element the sum of the squared second $n$ elements. Its third element twice the sum of the crossproducts of the first $n$ with the last $n$ elements.

For the evaluation of the symmetric $3 \times 3$ matrix $(\boldsymbol{X}^{*'}\boldsymbol{V}^{*-1}\boldsymbol{X}^{*})^{-1}$ a similar kind of computational scheme is performed. This leads to the evaluation of

$$\boldsymbol{C}'{\boldsymbol{\Sigma}_{\boldsymbol{E}}}^{-1}\boldsymbol{C} - \boldsymbol{C}'\boldsymbol{H}\boldsymbol{C},$$

a symmetric $2n \times 2n$ matrix that consists of four $n \times n$ matrices:

$$\begin{pmatrix} \boldsymbol{C}_{AA} & \boldsymbol{C}_{AB} \\ \boldsymbol{C}_{BA} & \boldsymbol{C}_{BB} \end{pmatrix},$$

with $\boldsymbol{C}_{AB} = \boldsymbol{C}'_{BA}$. The six distinct elements of $(\boldsymbol{X}^{*'}\boldsymbol{V}^{*-1}\boldsymbol{X}^{*})^{-1}$ can then be computed as:

$$\begin{aligned}
(1,1) &: tr(\boldsymbol{C}_{AA}\boldsymbol{C}_{AA}); \\
(1,2) &: tr(\boldsymbol{C}_{AB}\boldsymbol{C}_{BA}); \\
(1,3) &: 2tr(\boldsymbol{C}_{AA}\boldsymbol{C}_{BA}); \\
(2,2) &: tr(\boldsymbol{C}_{BB}\boldsymbol{C}_{BB}); \\
(2,3) &: 2tr(\boldsymbol{C}_{BA}\boldsymbol{C}_{BB}); \\
(3,3) &: tr(\boldsymbol{C}_{AB}\boldsymbol{C}_{AB}) + tr(\boldsymbol{C}_{BA}\boldsymbol{C}_{BA}) + 2tr(\boldsymbol{C}_{AA}\boldsymbol{C}_{BB}).
\end{aligned}$$

# References


Aitkin, M. (1999), A general maximum likelihood analysis of variance components in Generalized Linear Models, *Biometrics* **55**, 117–128.

Anderson, D., and M. Aitkin (1985), Variance component models with binary response: interviewer variability, *Journal of the Royal Statistical Society, Series B* **47**, 203–210.

Anderson, C., S. Wasserman, and B. Crouch (1999), A p* primer: Logit models for social networks, *Social Networks* **21**, 37–66.

Boer, P., M. Huisman, T.A.B. Snijders, and E.P.H. Zeggelink (2003), *StOCNET, an open software system for the advanced statistical analysis of social networks user's manual, version 1.4*, iec ProGAMMA/University of Groningen,

Groningen. http://stat.gamma.rug.nl/stocnet/
Bonney, G.E. (1987), Logistic regression for dependent binary observations, *Biometrics* **43**, 951–973.
Breslow, N.E., and D.G. Clayton (1993), Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**, 9–25.
Browne, W.J., and D. Draper (2003), A comparison of Bayesian and likelihood-based methods for fitting multilevel models, under review.
Connolly, M.A., and K.-Y. Liang (1988), Conditional logistic regression models for correlated binary data, *Biometrika* **75**, 501–506.
Fienberg, S.E., M.M. Meyer, and S.S. Wasserman (1985), Statistical analysis of multiple sociometric relations, *Journal of the American Statistical Association* **80**, 51–67.
Fienberg, S.E. and S. Wasserman (1981), Categorical data analysis of single sociometric relations, in: S. Leinhardt (ed.), *Sociological Methodology 1981*, Jossey-Bass, San Francisco, 156–192.
Frank, O., and D. Strauss (1986), Markov graphs, *Journal of the American Statistical Association* **81**, 832–842.
Goldstein, H. (1986), Multilevel mixed linear model analysis using iterative generalised least squares, *Biometrika* **73**, 43–56.
Goldstein, H. (1991), Nonlinear multilevel models, with an application to discrete response data, *Biometrika* **78**, 45–51.
Goldstein, H. (1994), Improved estimation for logit and loglinear multilevel models, *Multilevel Modelling Newsletter* **6**(1), 2.
Goldstein, H. (1995), *Applied Multilevel Analysis* (2nd ed.), Edward Arnold, London.
Goldstein, H. and J. Rasbash (1992), Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares, *Computational Statistics and Data Analysis* **13**, 63–71.
Goldstein, H. and J. Rasbash (1996), Improved approximations for multilevel models with binary responses, *Journal of the Royal Statistical Society, A* **159**, 505–513.
Hedeker, D., and R.D. Gibbons (1997), Random effects probit and logistic regression for three-level data, *Biometrics* **53**, 1527–1537.
Holland, P.W., K.B. Laskey, and S. Leinhardt (1983), Stochastic blockmodels: first steps, *Social Networks* **5**, 109–137.
Holland, P.W. and S. Leinhardt (1981), An exponential family of probability distributions for directed graphs, *Journal of the American Statistical Association* **77**, 33–50.
Kenny, D.A., and L. La Voie (1984), The social relations model, in: L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 18, Academic Press, New York, 141–182.
Lazega, E. (1992), Analyse de réseaux d'une organisation collégiale: les avocats

d'affaires, *Revue française de sociologie* **23**, 559–589.
Lazega, E. (1995), Les échanges d'idées entre collègues: Concurrence, coopération et flux de conseils dans un cabinet américain d'avocats d'affaires, *Revue suisse de sociologie* **21**, 61–84.
Lazega, E. (2001), *The collegial phenomenon. The social mechanisms of cooperation among peers in a corporate law partnership.* Oxford University Press, Oxford.
Lazega, E. and M.A.J. van Duijn (1997), Formal structure and exchanges of advice in a law firm: a random effects model, *Social Networks* **19**, 375–397.
Longford, N.T. (1994), Logistic regression with random coefficients, *Computational Statistics and Data Analysis* **17**, 1–15.
Maddala, G.S. (1977), *Econometrics*, McGraw Hill, Inc., Singapore.
McCullagh, P. and J.A. Nelder (1989), *Generalized Linear Models.* Chapman and Hall, London.
McCulloch, C.E. (1997), Maximum Likelihood algorithms for generalized linear mixed models, *Journal of the American Statistical Association* **92**, 162–170.
Neuhaus, J.M., and N.P. Jewell (1990), Some comments on Rosner's Multiple Logistic Model for Clustered Data, *Biometrics* **46**, 523–534.
Pattison, P., and S. Wasserman (1999), Logit models and logistic regressions for social networks: II. Multivariate relations, *British Journal of Mathematical and Statistical Psychology* **52**, 169–193.
Preisler, H.K. (1993), Modelling spatial patterns of trees attacked by barkbeetles, *Applied Statistics* **42**, 501–514.
Prentice, R.L. (1988), Correlated binary regression with covariates specific to each binary observation, *Biometrics* **44**, 1033–1048.
Rabe-Hesketh, S., A. Pickles, and S. Skondral (2001), GLLAMM: A general class of multilevel models and a Stata program, *Multilevel Modelling Newsletter* **13**(1), 17–23.
Rasbash, J., W. Browne, M. Healy, B. Cameron, and C. Charlton (2000), *MLwiN. Version 1.10*, Multilevel Models Project, Institute of Education, London.
Raudenbush, S.W., M.-L. Yang, and M. Yosef (2000), Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation, *Journal of Computational and Graphical Statistics* **9**, 141–157.
Rennolls, K. (1995), $p_{\frac{1}{2}}$, in: M.G. Everett and K. Rennolls (eds.), *International conference on social networks proceedings, vol. 1: methodology*, Greenwich University Press, London, 151–160.
Robins, G., P. Pattison, and S. Wasserman (1999), Logit models and logistic regressions for social networks, III. Valued relations, *Psychometrika* **64**, 371–394.
Rodríguez, G., and N. Goldman (1995), An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society, A* **158**, 73–89.

Rodríguez, G., and N. Goldman (2001), Improved estimation procedures for multilevel models with binary response: a case-study, *Journal of the Royal Statistical Society, Series A* **164**, 339–355.
Rosner, B. (1982), Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes, *Biometrics* **38**, 105–114.
Rosner, B. (1984), Multivariate methods in ophthalmology with application to other paired-data situations, *Biometrics* **40**, 1025–1035.
Schall, R. (1991), Estimation in generalized linear models with random effects, *Biometrika* **78**, 719–727.
Searle, S.R. (1982), *Matrix algebra useful for dtatistics*, John Wiley and Sons, New York.
Serfling, R.J. (1980), *Approximation theorems of mathematical statistics*. John Wiley and Sons, New York.
Snijders, T.A.B., and D. Kenny (1999), The Social Relations Model for family data: a multilevel approach, *Personal Relationships* **6**, 471–486.
Snijders, T.A.B. (2002), Markov Chain Monte Carlo estimation of exponential random graph models, *Journal of Social Structure* **3**(2). http://www.cmu.edu/-joss/
Snijders, T.A.B., and M.A.J. van Duijn (2002), Conditional Maximum Likelihood estimation under various specifications of exponential random graph models, in: J. Hagberg (ed.), *Contributions to social network analysis, information theory, and other topics in statistics*, University of Stockholm, Department of Statistics, Stockholm, 117–134.
Stiratelli, R., N. Laird, and J.H. Ware (1984), Random-effects models for serial observations with binary response, *Biometrics* **40**, 961–971.
Strauss, D., and M. Ikeda (1990), Pseudolikelihood estimation for social networks, *Journal of the American Statistical Association* **85**, 204–212.
Wang, Y.J., and G.Y. Wong (1987), Stochastic blockmodels for directed graphs, *Journal of the American Statistical Association* **82**, 8–19.
Warner, R.M., D.A. Kenny, and M. Stoto (1979), A new round robin analysis of variance for social interaction data, *Journal of Personality and Social Psychology* **37**, 1742–1757.
Wasserman, S. and K. Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.
Wasserman, S. and J. Galaskiewicz (1984), Some generalizations of $p_1$: external constraints, interactions and non-binary relations, *Social Networks* **6**, 177–192.
Wasserman, S. and D. Iacobucci (1986), Statistical analysis of discrete relational data, *British Journal of Mathematical and Statistical Psychology* **39**, 41–64.
Wasserman, S., and P. Pattison (1996), Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and $p^*$, *Psychometrika* **61**, 401–425.
Wasserman, S. and S.O. Weaver (1985), Statistical analysis of binary rela-

tional data: parameter estimation, *Journal of Mathematical Psychology* **29**, 406–427.
Wong, G.Y. (1987). Bayesian models for directed graphs. *Journal of the American Statistical Association*, **82**, 140–148.
Wright, D.B. (1997), Extra-binomial variation in multilevel logistic models with sparse structure, *British Journal of Mathematical and Statistical Psychology* **50**, 21–29.
Zeger, S.L., and M.R. Karim (1991), Generalized linear models with random effects: a Gibbs sampling approach, *Journal of the American Statistical Association* **86**, 79–86.
Zeggelink, E.P.H. (1993), *Strangers into Friends. The evolution of friendship networks using an individual oriented modeling approach*, Thesis Publishers, Amsterdam.
Zijlstra, B.J.H., and M.A.J. van Duijn (2003), *Manual $p_2$. Version 2.0.0.7*, iec ProGAMMA/University of Groningen, Groningen.

TABLE 1
Results of the $p_2$ analysis of Lazega's associates' friendship network

| | | Model 0 | Model 1 | | Model 2 | |
|---|---|---|---|---|---|---|
| | | Est. (S.E.) | Est. (S.E.) | $p$-value | Est. (S.E.) | $p$-value |
| Density | $\mu$ | -2.70 (0.23) | -0.64 (0.35) | | -1.62 (0.37) | |
| | Location[1] | | -2.33 (0.43) | <0.001 | -1.45 (0.31) | <0.001 |
| | Seniority[2] | | -0.58 (0.09) | <0.001 | -0.49 (0.09) | <0.001 |
| | Seniority[3] | | 0.18 (0.07) | 0.011 | | |
| | Gender[1] | | -0.55 (0.17) | 0.001 | -0.58 (0.18) | 0.002 |
| | Specialty[2] | | -0.51 (0.17) | 0.002 | | |
| | Advice | | | | 1.50 (0.21) | <0.001 |
| | Cowork | | | | 0.53 (0.24) | 0.027 |
| Reciprocity | $\rho$ | 3.29 (0.31) | 2.21 (0.36) | | 2.22 (0.35) | |
| | Location[1] | | 1.72 (0.94) | 0.068 | | |
| Sender Variance | $\sigma_A^2$ | 1.08 (0.25) | 1.19 (0.27) | | 1.36 (0.30) | |
| Receiver Variance | $\sigma_B^2$ | 0.75 (0.19) | 0.63 (0.17) | | 0.65 (0.18) | |
| Covariance | $\sigma_{AB}$ | -0.33 (0.16) | -0.01 (0.16) | | 0.16 (0.17) | |

[1] dichotomized difference of sending and receiving actor covariate values

[2] absolute difference of sending and receiving actor covariate values

[3] difference of sending and receiving actor covariate values