PSYCHOMETRIKA—VOL. 64, NO. 3, 371–394
SEPTEMBER 1999

LOGIT MODELS AND LOGISTIC REGRESSIONS FOR SOCIAL NETWORKS: III. VALUED RELATIONS

GARRY ROBINS

DEAKIN UNIVERSITY

PHILIPPA PATTISON

UNIVERSITY OF MELBOURNE

STANLEY WASSERMAN

UNIVERSITY OF ILLINOIS

This paper generalizes the p^* model for dichotomous social network data (Wasserman & Pattison, 1996) to the polytomous case. The generalization is achieved by transforming valued social networks into three-way binary arrays. This data transformation requires a modification of the Hammersley-Clifford theorem that underpins the p^* class of models. We demonstrate that, provided that certain (non-observed) data patterns are excluded from consideration, a suitable version of the theorem can be developed. We also show that the approach amounts to a model for multiple logits derived from a pseudo-likelihood function. Estimation within this model is analogous to the separate fitting of multinomial baseline logits, except that the Hammersley-Clifford theorem requires the equating of certain parameters across logits. The paper describes how to convert a valued network into a data array suitable for fitting the model and provides some illustrative empirical examples.

Key words: social networks, p^* models, autologistic models, pseudo-likelihood estimation.

1. Introduction and Background

The goal of this paper is to extend the family of models termed p^*, presented in Frank and Strauss (1986), Strauss and Ikeda (1990) and Wasserman and Pattison (1996), from dichotomous to valued social relations. This paper is a companion paper to Wasserman and Pattison (1996), which presented p^* models for the case of a single dichotomous social network, and to Pattison and Wasserman (in press), which generalized p^* models to multiple binary networks.

The p^* family is a class of models for a single dichotomous social network relation, with parameters reflecting a wide variety of possible structural features. Yet measurements of social network relationships are often in valued form, being designed to reflect distinct levels of intensity of social ties. Here we present an extension of the p^* family to valued (polytomous) social network data, and so describe a general class of models that can be used to evaluate a wide range of hypotheses about structure in valued relations.

The statistical analysis of valued relations has a small literature. Wasserman and Iacobucci (1986) presented an extension of the dyad independence model p_1 to nondichotomous relations (see also chapter 15 of Wasserman & Faust, 1994); recently, their "valued p_1" models were generalized by Anderson and Wasserman (1995), and Wong and Wang (1995). Anderson and Wasserman added substantively interesting higher-order interactions to the models (multiplicative, rather than additive, effects) while Wong and Wang introduced a more elaborate parameter structure to a model in logit form. All these extensions, however, assume dyadic independence. Frank and Strauss (1986) briefly discussed Markov graph models for valued relations.

This research was supported by grants from the Australian Research Council, the National Science Foundation (#SBR96-30754), and the National Institute of Health (#PHS-1RO1-39829-01).

Requests for reprints should be sent to Garry Robins, Faculty of Health and Behavioural Sciences, School of Psychology, Deakin University, Geelong Victoria 3217, AUSTRALIA.

0033-3123/1999-3/1995-0433-A $00.75/0

Other approaches to statistical modeling of valued relations include correspondence analysis and RC(M) association and correlation models (see Agresti, 1990), which were developed for social networks in papers by Wasserman, Faust, and Galaskiewicz (1990), Wasserman and Faust (1989), and Faust and Wasserman (1993). These general purpose models can be useful but do not allow the parameterization of particular network features that may be of interest.

It would clearly be useful, then, to construct models for valued networks that possess a statistical basis and that allow for the parameterization of network features beyond the dyadic level. The purpose of this paper is to develop such models as an extension of the dichotomous case. The binary models first arose as models for lattice structures (Ising, 1925), and have found much use in spatial applications (Besag, 1972, 1974, 1977a; Strauss, 1992; Wasserman, 1987). Wasserman and Pattison (1996) elaborated Frank and Strauss's (1986) application of these models to social networks, and utilized the standard pseudo-likelihood estimation approach to fitting these models, first described by Besag (1975, 1977b) and applied to social networks by Strauss and Ikeda (1990).

After a presentation of notation in section 2, we briefly describe p^* models for dichotomous network data in section 3, and cover some important preliminaries including dependence graphs and the Hammersley-Clifford theorem. This theorem was first introduced in the context of spatial auto-logistic models by Besag (1974), was utilized by Frank and Strauss (1986) and Strauss and Ikeda (1990) in generalizing spatial statistical models to the case of social networks, and was also used by Wasserman and Pattison (1996) and Pattison and Wasserman (in press) to underpin the p^* class of models (see also Wasserman & Pattison, 1999). In section 4, we extend p^* models to polytomous data by converting valued data matrices to binary three-way arrays. We demonstrate that a simple transformation of valued data to a three-way binary array leads to a suitable binary version of the Hammersley-Clifford theorem, provided that certain (nonobserved) data patterns are excluded. This transformation amounts to a model for multiple logits derived from a pseudo-likelihood function. We show that estimation within this model is analogous to the separate fitting of multinomial baseline logits, except that the Hammersley-Clifford theorem requires the equating of certain parameters across logits. We explain how to convert a valued network into a data array suitable for fitting the model. In section 6, we present some empirical examples as illustrations.

2. Some Notation

2.1. Notation for Dichotomous and Polytomous Networks

We generalize the notation presented in Wasserman and Pattison (1996) and in Pattison and Wasserman (in press). A valued *social network* comprises a set $\mathcal{N} = \{1, 2, \ldots, g\}$ of g social *actors* and a valued *social relation*. If the relation is dichotomous, it can be regarded as comprising a set of ordered pairs recording pairs of actors for whom a relational tie is present. If the ordered pair (i, j) is in this set, then the first actor (i) has a relational tie to the second actor (j) in the pair. It will be convenient below to write ij as short-hand for the ordered pair (i, j).

Denote the set of all possible relational ties as $\mathcal{T} \subseteq \mathcal{N} \times \mathcal{N}$, where $\times$ indicates the Cartesian product of two sets. Normally, actors are not assumed to have relations with themselves, so $ii \notin \mathcal{T}$. Below, for $S \subseteq \mathcal{T}$, we shall use the short-hand $S - ij$ to indicate the set S excluding the ordered pair ij, that is $S \cap \overline{\{ij\}}$; and $S + ij$ to indicate $S \cup \{ij\}$.

A *valued social relation* arises when the ties are measured on an ordinal scale, taking integer values from the set $V = \{0, 1, \ldots, C - 1\}$; a dichotomous social relation may then be regarded as the special case when $C = 2$.

When a relational tie takes the integer value k, we say it has *strength* k. Any valued social relation can be represented by a $g \times g$ *matrix*, $\mathbf{X}$, where $X_{ij} = k$, if the tie ij has strength k. When $ij \notin \mathcal{T}$ the cell X_{ij} in the matrix can be coded to distinguish it from the strength values in V (for instance, X_{ij} could be set to -1). From the point of view of data analysis, such cells can be considered as missing values. (In our empirical example in Table 3, we have simply indicated such "missing" cells with a dash. In this example, the only "missing" cells relate to self-relations. In our second example, there are more "missing" cells because we exclude relations between members of separate groups.) Since the matrix **X** will be assumed to be a random quantity, we denote x_{ij} as a realization of X_{ij}, and the matrix **x** as a realization of **X**.

It is useful to define a relation derived by focusing only on ties at a particular strength. We define, for $k \neq 0$, the *k-strength relation*

$$\begin{aligned} Y_{(ij,k)} &= 1 \text{ if } X_{ij} = k \\ &= 0 \text{ if } X_{ij} \neq k, ij \in \mathcal{T} \end{aligned} \tag{1}$$

and where $Y_{(ij,k)}$ is regarded as missing if $ij \notin \mathcal{T}$. The point of adopting this additional notation is that it translates the valued social relation into a binary form. As will be seen below, our approach to dealing with valued relations is to work with this derived binary form in a similar manner to that discussed by Wasserman and Pattison (1996).

As before, we will denote a realization of $Y_{(ij,k)}$ as $y_{(ij,k)}$. We can represent the full set of $Y_{(ij,k)}$ as a three-way $g \times g \times (C-1)$ array, **Y**, and we will denote the three-way array **y** as a realization of **Y** such that the $y_{(ij,k)}$ constitute the cells of **y**. Alternatively, we can think of entries in **Y** as being indexed by elements in $\mathcal{T} \times V$ and below we shall use the notation (ij, v) to represent an element of this set, where $ij \in \mathcal{T}$ and $v \in V$.

2.2. *Notation for Conditional Logits*

As in Wasserman and Pattison (1996), we let ϖ represent conditional logits—log odds ratios, comparing the conditional probability of one outcome of a random variable to the conditional probability of another outcome, in a logarithm scale. In particular we let $\varpi_{ij,k}$ be a "baseline" conditional logit representing the log odds of X_{ij} taking the value k, compared to taking the value 0, conditional on all other cells in the matrix; that is

$$\varpi_{ij,k} = \log\left[\frac{P(X_{ij} = k \mid X_{st} = x_{st}, st \neq ij, st \in \mathcal{T})}{P(X_{ij} = 0 \mid X_{st} = x_{st}, st \neq ij, st \in \mathcal{T})}\right]. \tag{2}$$

We denote an analogous logit for the binary form of the k-strength relation (that is, expressed in terms of $Y_{(ij,k)}$) by $\varpi_{ij}^{[k]}$:

$$\varpi_{ij}^{[k]} = \log\left[\frac{P(Y_{(ij,k)} = 1 \mid Y_{(st,u)} = y_{(st,u)}, (st,u) \neq (ij,k), st \in \mathcal{T}, u \in V)}{P(Y_{(ij,k)} = 0 \mid Y_{(st,u)} = y_{(st,u)}, (st,u) \neq (ij,k), st \in \mathcal{T}, u \in V)}\right] \tag{3}$$

The second logit is not necessarily defined for all cells in the three-way array **Y**, since $P(Y_{(ij,k)} = 1 \mid Y_{(ij,u)} = 1, u \neq k) = 0$. This fact foreshadows that certain realizations of **Y** are impossible, as discussed below.

2.3. *Additional Notation*

In many of the derivations and algebraic expressions below, there is a need to set certain values in **X** and **Y** to zero, as well as a need at times to consider only a subset of non-missing cells in **X** and **Y**. For instance, in the conditional logit expression (2), the conditionality is specified by $X_{st} = x_{st}, st \neq ij, st \in \mathcal{T}$, that is, the conditionality is based on the observed values of non-missing cells in **X** excluding the cell ij. It is convenient to have a short-hand for these purposes.

First, we develop a simple means to express the setting to zero of a number of variables. Let **w** be an arbitrary vector $(w_1, w_2, \ldots, w_n)'$ of length n with J the index set $\{1, 2, \ldots, n\}$. Denote by $\mathbf{w}_A$, for $A \subseteq J$, the vector **w** with entries indexed by $J - A$ set to zero. For instance, $\mathbf{w}_{\{i\}} = (0, \ldots, 0, w_i, 0, \ldots, 0)'$ and $\mathbf{w}_{J-\{i\}} = (w_1, w_2, \ldots, w_{i-1}, 0, w_{i+1}, \ldots w_n)'$.

Secondly, we present a notation to represent a subset of variables. Let $|A| = m$. Denote by $\mathbf{w}^{\#}_{A}$ the ordered m-tuple derived from $\mathbf{w}$ by excluding the entries indexed by $J - A$, but retaining the natural ordering induced by J, so that $\mathbf{w}^{\#}_{A} = (w_{a_1}, w_{a_2}, \ldots w_{a_m})'$ where $\{a_1, a_2, \ldots a_m\} = A$ and $a_j < a_{j+1}$. For instance, $\mathbf{w}^{\#}_{\{i\}} = (w_i)$, and $\mathbf{w}^{\#}_{J-\{i\}} = (w_1, w_2, \ldots, w_{i-1}, w_{i+1}, \ldots w_n)'$.

More generally, the notation can be applied to higher-way arrays, whose entries can be regarded as forming a sequence. All that is required is some standard ordering on the set of nonmissing cells of the array. In our case, for instance, $\mathbf{X}_A$ for some $A \subseteq \mathcal{T}$ specifies the sequence of matrix cells X_{ij} such that $X_{ij} = x_{ij}$ for $ij \in A$ and $X_{ij} = 0$ for $ij \in \mathcal{T} - A$; while $\mathbf{X}^{\#}_{A}$ specifies the sequence of X_{ij} such that $ij \in A$. By these means, we express, respectively, that certain cells in $\mathbf{X}$ have been set to zero, or that we are considering only a subset of cells in $\mathbf{X}$.

3. The p^* Model

The original specification of the class of models p^* was for a single, dichotomous relation, as described by Wasserman and Pattison (1996) (see also Rennolls, 1995). Most of the early work (Frank & Strauss, 1986) focused on nondirected relations. Generalizations to valued relations were mentioned in passing (in concluding remarks) by Frank and Strauss (1986; sec. 6) and by Strauss and Ikeda (1990; sec. 5).

3.1. The Model for a Single Dichotomous Relation

Wasserman and Pattison (1996) introduced the p^* model in loglinear form:

$$P(\mathbf{X} = \mathbf{x}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{z}(\mathbf{x})\}}{\kappa(\boldsymbol{\theta})}, \tag{4}$$

where $\mathbf{X}$ represents the matrix of a dichotomous relation, $\boldsymbol{\theta}$ is a vector of model parameters, $\mathbf{z}(\mathbf{x})$ is a vector of network statistics and the function κ normalizes the expression for a given $\boldsymbol{\theta}$. This model allows the probability of the network $\mathbf{X}$ to be expressed in terms of various structural features that can be represented in terms of network statistics, such as reciprocity and transitivity. The difficulty with the model in this form, however, is that in any but the simplest cases calculation of κ is intractable. Following Strauss and Ikeda (1990), Wasserman and Pattison (1996) converted the model of (4) to a logit form:

$$\varpi_{ij} = \log\left\{\frac{P(X_{ij} = 1 \mid \mathbf{X}^{\#}_{\mathcal{T}-ij} = \mathbf{x}^{\#}_{\mathcal{T}-ij})}{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{\mathcal{T}-ij} = \mathbf{x}^{\#}_{\mathcal{T}-ij})}\right\} = \boldsymbol{\theta}'\left[\mathbf{z}(\mathbf{x}^{+}_{ij}) - \mathbf{z}(\mathbf{x}^{-}_{ij})\right], \tag{5}$$

where $\mathbf{x}^{+}_{ij}$ is the matrix $\mathbf{x}$ but with x_{ij} forced to take the value 1, and $\mathbf{x}^{-}_{ij}$ is the matrix $\mathbf{x}$ but with x_{ij} forced to take the value 0. This logit version of the model does not include κ. Maximum likelihood estimation of parameters is problematic but, as Strauss and Ikeda (1990) showed, approximate estimation under maximum pseudo-likelihood estimation is possible using standard logistic regression programs.

The accuracy of pseudo-likelihood estimation is difficult to assess. Strauss and Ikeda (1990) compared maximum likelihood and maximum pseudo-likelihood estimation in a simulation study and in a simple network case, and found that the two procedures performed similarly. Geyer and Thompson (1992), on the other hand, analyzed (nonnetwork) data for which pseudo-likelihood estimation performed poorly compared with Monte Carlo maximum likelihood techniques. One of the problems with pseudo-likelihood estimation is that standard errors of parameter estimates are uncertain, so the usual tests in logistic regression are at best approximate.

For the purposes of this paper, pseudo-likelihood estimation is used, bearing in mind that it is an approximate procedure that, at least in some cases, may be problematic. Our pseudo-likelihood function assumes that conditional logits are independent. In the longer run, pseudolikelihood procedures may be an interim phase. The development of Monte Carlo techniques, such as those suggested by Besag and Clifford (1989) and Geyer and Thompson (1992) may provide either an alternative method for the estimation of models discussed in this paper, or allow the conditions to be elaborated under which maximum pseudo-likelihood estimates can be used with confidence. (The Preisler, 1993, approach of bootstrapping standard errors also suggests another useful direction.) The development of Monte Carlo-based estimation specifically for p^* models is a matter for ongoing work (see Crouch & Wasserman, 1998).

3.2. Dependence Graphs

Frank and Strauss (1986) introduced the construct of a *dependence graph* for a social network as a means of representing the dependencies among possible network ties. The dependence graph indicates which relational ties (or subsets of relational ties) are assumed to be conditionally dependent. That is, the dependence graph has as nodes the set of all possible relational ties in $\mathcal{T}$; an edge between two nodes in the dependence graph signifies that the two corresponding ties are assumed to be dependent, conditional on all other ties in the network. Dependence graphs are analogous to the *independence graphs* used in the graphical modeling literature (see, for instance, Lauritzen, 1996).

Wasserman and Pattison (1996) largely followed the dependence structures discussed by Frank and Strauss (1986) (for instance, *Markov directed graphs,* which assume dependencies for any pair of relational ties that have an actor in common), although they briefly introduced some more complex dependence structures. More recently, Pattison and Wasserman (in press) provided a fuller presentation of dependence structures for multivariate social networks in terms of theoretical themes discussed in the literature on social structural models.

3.3. The Hammersley-Clifford Theorem

The link between dependence graphs and the models of (4) and (5) is the *Hammersley-Clifford theorem* (Besag, 1974). Besag, who was interested in stochastic models for spatial processes, used the theorem to formulate a conditional probability model for a finite system of spatially interacting random variables. Besag construed a lattice arrangement of *sites* to represent the spatial distribution associated with the random variables and introduced the concept of *neighboring sites* to specify dependencies among the variables. A *clique* was a set of sites which were all neighbors of each other. In these terms, the theorem can be stated as follows:

Hammersley-Clifford theorem. For a system of n interacting variables $\mathbf{W} = (W_i)$, where each variable can take only a finite number of values:

$$\log\left[\frac{P(W_i = w_i \mid \mathbf{W}^{\#}_{J-\{i\}} = \mathbf{w}^{\#}_{J-\{i\}})}{P(W_i = 0 \mid \mathbf{W}^{\#}_{J-\{i\}} = \mathbf{w}^{\#}_{J-\{i\}})}\right] = w_i \sum_{S \subseteq J-\{i\}} \Gamma_{S\cup\{i\}}(\mathbf{w}^{\#}_{S\cup\{i\}}) \prod_{s\in S} w_s \tag{6}$$

where $\Gamma_{S\cup\{i\}}$ is a function of the observed values of those variables indexed by the subset $S \cup \{i\}$ of $J = \{1, 2, \ldots n\}$, and where, if $S = \emptyset$, the term $\prod_{s\in S} w_s$ is taken to have the value 1. Moreover, $\Gamma_{S\cup\{i\}} = 0$ unless $S \cup \{i\}$ are all neighboring sites of each other (that is, a clique).

When the system is dichotomous, for each $S \subseteq J-\{i\}$ the expression, $\Gamma_{S\cup\{i\}}(\mathbf{w}^{\#}_{S\cup\{i\}}) \prod_{s\in S} w_s$ becomes zero unless $w_s = 1$ for all $s \in S$. In this instance, the term is nonzero only when the arguments of $\Gamma_{S\cup\{i\}}$ are all 1, in which case the Γ-functions can be taken as constants. In the dichotomous case, then, equation (6) becomes

$$\log\left[\frac{P(W_i = 1 \mid \mathbf{W}^{\#}_{J-\{i\}} = \mathbf{w}^{\#}_{J-\{i\}})}{P(W_i = 0 \mid \mathbf{W}^{\#}_{J-\{i\}} = \mathbf{w}^{\#}_{J-\{i\}})}\right] = \sum_{S\subseteq J-\{i\}} \gamma_{S\cup\{i\}} \prod_{s\in S} w_s, \tag{7}$$

where the γ-terms are constants. In other words, the theorem establishes that, in the dichotomous case, the γ-terms will be zero unless the variables indexed by the set $S \cup \{i\}$ form a clique of the dependence graph. (Apart from this constraint, Besag, 1974, noted that in the non-dichotomous case any set of Γ-functions gives rise to a valid probability distribution.)

In the case of social networks (i.e., where the binary vector $\mathbf{W}$ is replaced by a dichotomous matrix $\mathbf{X}$) the theorem establishes that a probability model for the collection of random variables $\{X_{ij}\}$ depends only on the *complete subgraphs*, or *cliques*, of the dependence graph. (A subgraph is *complete* if every pair of nodes in the subgraph is connected by an edge. A subgraph comprising a single node is also regarded as complete.) The model in (7) may be expressed in the form of (5) and hence also in the form of (4). Accordingly, the specification of dependencies through a dependence graph determines the nonzero terms in the Hammersley-Clifford expansion, which is in fact the logit p^* model (5) implied by that set of dependencies.

3.4. *Polytomous Models*

For polytomous data, (6) enables the specification of a p^* model. Unlike the dichotomous case, however, the terms $\Gamma_{S\cup\{i\}}(\mathbf{w}^{\#}_{S\cup\{i\}}) \prod_{s\in S} w_s$ are not necessarily zero for a particular value of w_s, and so instead of a set of constants as in (7), we retain a set of unspecified functions. As a result, (6) leads to versions of (5) and (4) but with the parameters $\boldsymbol{\theta}$ replaced by a set of functions.

For estimation purposes, however, there are simpler approaches than dealing with a set of functions. With polytomous data, there are only a finite number of arguments for any function $\Gamma_{S\cup\{i\}}$, so that $\Gamma_{S\cup\{i\}}$ will only take a finite range of values. Once we know these values, we have specified the RHS of all possible versions of equation (6), so that these values can be taken as the parameters of our conditional logit model.

More particularly, if the variables take values from the set $V = \{0, 1, \ldots, C-1\}$, the function $\Gamma_{S\cup\{i\}}$ will take the values $\Gamma_{S\cup\{i\}}(\mathbf{v})$ for all vectors $\mathbf{v}$ of length $|S| + 1$ with entries $v_t \in V$, that is, a total of $C^{|S|+1}$ possible values. Yet in (6) if any of the arguments of $\Gamma_{S\cup\{i\}}$ are zero, then the term $w_i \prod_{s\in S} w_s$ is zero. Accordingly, we only need consider nonzero values as arguments of $\Gamma_{S\cup\{i\}}$. There will be $(C-1)^{|S|+1}$ of these, each of which will be a parameter.

We can include all of these parameters in (6) by expanding the term $\Gamma_{S\cup\{i\}}(\mathbf{w}^{\#}_{S\cup\{i\}}) \prod_{s\in S} w_s$ into a summation across the possible values of $\Gamma_{S\cup\{i\}}$, provided that we incorporate indicator variables to indicate the actual observed values. For instance, suppose $S = \emptyset$ (so that $\prod_{s\in S} w_s = 1$, as noted above). Then $\Gamma_{\{i\}}$ can take C values, although, as noted above we can ignore the value 0. So the term $\Gamma_{\{i\}}(\mathbf{w}^{\#}_{\{i\}}) = \Gamma_{\{i\}}(w_i)$ can be expanded into the summation $\sum_{t=1}^{C-1} \Gamma_{\{i\}}(t) y_i^{[t]}$, where the indicator variable $y_i^{[t]} = 1$ if $w_i = t$ and $y_i^{[t]} = 0$ if $w_i \neq t$. In this way, we develop a model that includes all parameters.

What this amounts to, of course, is a transformation to a higher-order binary array. In the network case, the corresponding transformation is achieved by using the k-strength relations in Definition (1).

In developing p^* models for the polytomous case, we choose to work with the three-way array $\mathbf{Y}$ rather than the original valued array $\mathbf{X}$. In part, this choice is made because it leads to a natural model parameterization, as the foregoing argument suggests. But as well, it permits us to determine precisely what data to include in the estimation of the $C-1$ conditional logits that specify the model (that is, one conditional logit for each possible nonzero value of X_{ij}). It also allows straightforward resolution of a problem that emerges in these auto-logistic models, that of the need to equate certain parameters across logits.

Models such as p^* that are derived from the expansion (6) also have an *autologistic* character, analogous to classical logistic models except that the explanatory variables are themselves observations on the process (Besag, 1974). In the logit for W_i, the i-th variable acts as a response, but in the logit for W_j it may be a predictor. There is an important implication of this property for equating parameters. In (7), there is one parameter for each clique of the dependence graph (i.e., one γ-term). So if a γ-term involves both the i-th and j-th variable, it will appear in both the logits for W_i and W_j. In the case of models for polytomous networks, we discuss below the need to equate parameters across different logits that estimate effects for different strengths of tie.

4. Models for the Three-Way Binary Array

Our approach is to construct a version of the Hammersley-Clifford equation (7) for the dichotomous three-way array $\mathbf{Y}$, constructed from the valued matrix $\mathbf{X}$. The purpose of this section is to show how the network version of the polytomous expansion for $\mathbf{X}$ in (6) can be converted to a dichotomous version for $\mathbf{Y}$. In making the transformation explicit, we establish consequences for the way in which the data should be structured for estimation. In particular, we show that certain patterns within the three-way array need to be excluded, as mentioned earlier. Estimation is addressed in section 5.

Before constructing an appropriate form of the Hammersley-Clifford equation for the three-way array $\mathbf{Y}$, it is important to note that Besag (1974) used a positivity condition in establishing the theorem. In Appendix A we show that Besag's condition can be replaced by a less stringent version that permits us to construct the analogue of (7) for the binary array $\mathbf{Y}$.

4.1. Transforming Polytomous to Binary Variables

We now formulate a version of the dichotomous Hammersley-Clifford expansion (7) for the three-way binary array $\mathbf{Y}$.

When in the case of a polytomous network, we want to compare the conditional log-odds of a tie taking the value k against there being no tie at all, the appropriate version of (6) is

$$\log\left[\frac{P(X_{ij}=k \mid \mathbf{X}^{\#}_{\mathcal{T}-ij}=\mathbf{x}^{\#}_{\mathcal{T}-ij})}{P(X_{ij}=0 \mid \mathbf{X}^{\#}_{\mathcal{T}-ij}=\mathbf{x}^{\#}_{\mathcal{T}-ij})}\right] = k \sum_{S\subseteq\mathcal{T}-ij} \Gamma_{S+ij}(\mathbf{x}^{\#}_{S+ij}) \prod_{st\in S} x_{st},$$

that is

$$\varpi_{ij,k} = k \sum_{S\subseteq\mathcal{T}-ij} \Gamma_{S+ij}(\mathbf{x}^{\#}_{S+ij}) \prod_{st\in S} x_{st}. \tag{8}$$

We now transfer to the three-way binary array $\mathbf{Y}$ by transforming $\mathbf{X}$ through (1). The basic result is given in the following theorem

Theorem 1. The expansion of equation (8) is equivalent to

$$\varpi_{ij}^{[k]} = \sum_{M\subseteq\mathcal{T}\times V-(ij,k)} \gamma_{M+(ij,k)} \prod_{m\in M} y_m, \tag{9}$$

where $m=(st,v)\in S\times V$ and where we force $y_m=0$ when $m=(ij,v)$ for $0<v\neq k$.

The proof of this theorem is given in Appendix B.

Equation (9) has exactly the same form as the Hammersley-Clifford equation (7).

The forced zeroes imply that certain patterns of data in **Y** are not permissible. Clearly, $Y_{(ij,v)} = 1$ and $Y_{(ij,k)} = 1$ are individually possible, but not jointly if $v \neq k$, as the network tie cannot have two values simultaneously. If $Y_{(ij,v)} = 1$, in estimating the logit $\varpi_{ij}^{[k]}$ for $v \neq k$, we require $Y_{(ij,v)} = 0$. To avoid this contradiction, we state that when $Y_{(ij,v)} = 1$ the logit $\varpi_{ij}^{[k]}$ is undefined and is to be excluded from the set of equations represented by (9). This is equivalent to stating that if $Y_{(ij,v)} = 1$, then $Y_{(ij,k)}$ is undefined for all $k \neq v$.

In so doing we have followed Besag's (1974) advice to restrict the sample space to those patterns of data in **Y** that are permissible. As discussed in Appendix A, our revised positivity condition allows the Hammersley-Clifford theorem to be proved in this case.

Note that the data patterns that need to be excluded imply that, when we estimate the polytomous logit for relations of strength k, observed relations with nonzero strengths other than k are to be excluded as "response" variables.

5. Estimation

We now show that estimation of the dichotomous model in (9) is analogous to the separate fitting of baseline logits for the polytomous relationship variable **X**. For the sake of a simple exposition we assume trichotomous valued data here but the argument can be generalized readily to relations with more than three values.

For trichotomous data, the transformation (1), together with the exclusion of impermissible data patterns described in the previous section, can be written as follows:

$$\begin{aligned}
&\text{if } X_{ij} = 0,\ Y_{(ij,1)} = Y_{(ij,2)} = 0;\\
&\text{if } X_{ij} = 1,\ Y_{(ij,1)} = 1 \text{ and } Y_{(ij,2)} \text{ is undefined;}\\
&\text{if } X_{ij} = 2,\ Y_{(ij,2)} = 1 \text{ and } Y_{(ij,1)} \text{ is undefined.}
\end{aligned} \tag{10}$$

Denote as P the subset of $\mathcal{T} \times V$ wherein $Y_{(ij,k)}$ is defined. Then for each $(ij, k) \in P$, we can write (9) as follows

$$\begin{aligned}
\varpi_{ij}^{[k]} &= \sum_{M \subseteq P-(ij,k)} \gamma_{M+(ij,k)} \prod_{m \in M} y_m \\
&= \sum_{M \subseteq P-(ij,1)} \gamma_{M+(ij,1)} z_{(ij,k)}^{M+(ij,1)} + \sum_{M \subseteq P-(ij,2)} \gamma_{M+(ij,2)}\, z_{(ij,k)}^{M+(ij,2)},
\end{aligned} \tag{11}$$

where we define

$$\begin{aligned}
z_{(ij,k)}^{Q} &= \prod_{q \in Q-(ij,k)} y_q \text{ when } (ij, k) \in Q \subseteq P \\
&= 0 \text{ when } (ij, k) \notin Q \text{ or when } Q \text{ is not a subset of } P.
\end{aligned}$$

Note that if both $(ij, 1), (ij, 2) \in Q \subseteq P$, then $Y_{(ij,1)} = Y_{(ij,2)} = 0$, so that $z_{(ij,k)}^{Q} = 0$. So $z_{(ij,k)}^{M+(ij,s)}$ can only be nonzero if $s = k$. In what follows, then, we need only consider those instances of $z_{(ij,k)}^{M+(ij,s)}$ where $s = k$ and where $(ij, v) \notin M$ for both $v = 1$ and $v = 2$. Furthermore, we know from the Hammersley-Clifford theorem that $\gamma_{M+(ij,k)} = 0$ unless the set $M + (ij, k)$ makes up a clique of the dependence graph for **Y**. This will place further constraints on the instances of the set M that we need to consider. (Homogeneity constraints, which we discuss below, introduce yet more constraints on what we need to consider in M.)

In practice, the statistics $z_{(ij,k)}^{M+(ij,s)}$ are readily computed. Consider, for instance, a clique in the dependence graph comprising a tie of strength 1 from i to j and a tie of strength 1 from j to i (that is, a clique corresponding to a strength 1 reciprocity effect). Then in the logits $w_{ij}^{[k]}$,

$M = \{(ji, 1)\}$, and the nonzero statistics are $z_{(ij,k)}^{M+(ij,s)} = z_{(ij,k)}^{\{(ji,1)(ij,1)\}} = y_{(ji,1)}$ as long as $k = 1$. Of course, $y_{(ji,1)}$ simply indicates whether there is a tie of strength 1 from j to i. So in the logit for relations at strength 1, there will be a reciprocity term $\gamma_{\{(ij,1),(ji,1)\}}y_{(ji,1)}$ where $\gamma_{\{(ij,1),(ji,1)\}}$ is the parameter for a strength 1 reciprocity effect. Given the way we have set up the z-statistics, there will be no such term in the logit for relations at strength 2, which is exactly what we require.

Equation (11) has the form

$$\varpi_{ij}^{[k]} = \boldsymbol{\gamma}'.\mathbf{z}_{(ij,k)} \tag{12}$$

where $\boldsymbol{\gamma}$ is a vector of parameters with γ_Q as vector cells for all $Q \subseteq P$, and $\mathbf{z}_{(ij,k)}$ is a vector of statistics with $z_{(ij,k)}^{Q}$ as vector cells. The (ij, k) suffix signifies that the statistics $\mathbf{z}_{(ij,k)}$ are calculated for the case (ij, k).

Equation (12), of course, remains an autologistic model, but for the purposes of pseudo-likelihood estimation, as Strauss and Ikeda (1990) and Wasserman and Pattison (1996) have shown, we can treat it as an ordinary logistic regression equation, considering $\mathbf{z}$ as a vector of explanatory variables and $\boldsymbol{\gamma}$ a vector of parameters.

5.1. *Baseline Logits*

It is a straightforward matter to show that (12) is equivalent to baseline logits for the original relational variables, $\mathbf{X}$. For all $(ij, 1) \in P$, we have

$$\begin{aligned} P(Y_{(ij,1)} = 1 \mid \mathbf{Y}^{\#}_{P-(ij,1)}) &= P(X_{ij} = 1 \mid X_{ij} \neq 2, \mathbf{X}^{\#}_{T-ij}) \\ &= \frac{P(X_{ij} = 1, X_{ij} \neq 2 \mid \mathbf{X}^{\#}_{T-ij})}{P(X_{ij} \neq 2 \mid \mathbf{X}^{\#}_{T-ij})} \\ &= \frac{P(X_{ij} = 1 \mid \mathbf{X}^{\#}_{T-ij})}{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{T-ij}) + P(X_{ij} = 1 \mid \mathbf{X}^{\#}_{T-ij})}. \end{aligned}$$

Similarly,

$$P(Y_{(ij,1)} = 0 \mid \mathbf{Y}^{\#}_{P-(ij,1)}) = \frac{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{T-ij})}{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{T-ij}) + P(X_{ij} = 1 \mid \mathbf{X}^{\#}_{T-ij})}$$

so that we have:

$$\varpi_{ij}^{[1]} = \log\left[\frac{P(X_{ij} = 1 \mid \mathbf{X}^{\#}_{T-ij})}{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{T-ij})}\right] = \boldsymbol{\gamma}_1'.\mathbf{z}_1 \tag{13}$$

where $\boldsymbol{\gamma}_1$ is a vector of parameters $\gamma_{M+(ij,1)}$ and $\mathbf{z}_1$ a vector of statistics $z_{(ij,1)}^{M+(ij,1)}$. These statistics were defined above in terms of the $\mathbf{Y}$ array, but of course can be equivalently expressed in terms of the $\mathbf{X}$ array. We have a similar logit for the second value

$$\varpi_{ij}^{[2]} = \log\left[\frac{P(X_{ij} = 2 \mid \mathbf{X}^{\#}_{T-ij})}{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{T-ij})}\right] = \boldsymbol{\gamma}_2'.\mathbf{z}_2 \tag{14}$$

So (12) is equivalent to two baseline logits in the trichotomous case. More generally, with polytomous values $\{0, 1, \ldots, C - 1\}$, there are $C - 1$ baseline logits.

5.2. *Pseudo-Likelihood Functions*

From (12), the pseudo-likelihood function, which assumes independence of baseline logits, is

$$l(\gamma) = \prod_{(rs,v)\in P} \left[\frac{\exp \gamma'.\mathbf{z}_{(rs,v)}}{1+\exp \gamma'.\mathbf{z}_{(rs,v)}}\right]^{Y_{(rs,v)}} \left[\frac{1}{1+\exp \gamma'.\mathbf{z}_{(rs,v)}}\right]^{1-Y_{(rs,v)}}$$

$$= \prod_{(rs,v)\in P} \frac{\left(\exp \gamma'.\mathbf{z}_{(rs,v)}\right)^{Y_{(rs,v)}}}{1+\exp \gamma'.\mathbf{z}_{(rs,v)}}.$$

Accordingly the log-pseudo-likelihood is

$$L(\gamma) = \sum_{(rs,v)\in P} \left[Y_{(rs,v)}\gamma'.\mathbf{z}_{(rs,v)} - \log\left(1+\exp \gamma'.\mathbf{z}_{(rs,v)}\right)\right]$$

so the pseudo-likelihood is maximized by solution of the equations, for all $Q \subseteq P$

$$\frac{\partial L(\gamma)}{\partial \gamma_Q} = \sum_{(rs,v)\in Q} \left[Y_{(rs,v)} z^Q_{(rs,v)} - \frac{z^Q_{(rs,v)} \exp \gamma'.\mathbf{z}_{(rs,v)}}{\left(1+\exp \gamma'.\mathbf{z}_{(rs,v)}\right)}\right] = 0. \tag{15}$$

If $(rs, 1) \in Q$, as noted above $z^Q_{(rs,v)} = 0$ if it is also the case that $(rs, 2) \in Q$. So we need only consider those cliques Q of the dependence graph which have as elements one but not both of $(rs, 1)$ and $(rs, 2)$ as elements. Moreover if $z^Q_{(rs,1)} \neq 0$ then $z^Q_{(rs,2)} = 0$, so that (15) decomposes into two sets of equations:

$$\sum_{(rs,1)\in Q} \left[Y_{(rs,1)} z^Q_{(rs,1)} - \frac{z^Q_{(rs,1)} \exp \gamma'.\mathbf{z}_{(rs,1)}}{\left(1+\exp \gamma'.\mathbf{z}_{(rs,1)}\right)}\right] = 0 \tag{16}$$

$$\sum_{(rs,2)\in Q} \left[Y_{(rs,2)} z^Q_{(rs,2)} - \frac{z^Q_{(rs,2)} \exp \gamma'.\mathbf{z}_{(rs,2)}}{\left(1+\exp \gamma'.\mathbf{z}_{(rs,2)}\right)}\right] = 0.$$

Note that the equations in (16) are not equivalent to joint maximum pseudo-likelihood estimation of (13) and (14). For joint estimation it is simple enough to show (Hosmer & Lemeshow, 1989) that the estimates are the solutions of the equations

$$\frac{\partial L(\gamma_1, \gamma_2)}{\partial \gamma_Q} = \sum_{(rs,v)\in Q} \left[Y_{(rs,v)} z^Q_{(rs,v)} - \frac{z^Q_{(rs,v)} \exp \gamma'_v.\mathbf{z}_{(rs,v)}}{\left(1+\exp \gamma'_1.\mathbf{z}_{(rs,1)} + \exp \gamma'_2.\mathbf{z}_{(rs,1)}\right)}\right] = 0,$$

which are not necessarily identical to (16).

The analogue of (16) in standard logistic regression is referred to as separate, rather than simultaneous, fitting of baseline logits. Begg and Gray (1984) proposed that separate fitting was a satisfactory alternative to simultaneous fitting. They commended the separate fitting procedure as enhancing analytic simplicity and flexibility, especially regarding variable selection, and noted that separate fitting of the logits overcame some practical difficulties that were, and still are, apparent in major computer statistical packages. Begg and Gray showed that the two models are parametrically equivalent and that the resulting estimates are consistent and efficient.

Hosmer and Lemeshow (1989) noted that in their experience coefficient estimates from separate fitting were close to those from polytomous fitting, although they urged caution in drawing definitive inferences from these estimates and their associated standard errors. (As we are working in a pseudo-likelihood framework, such caution has to be exercised in any case.) Begg and Gray (1984) noted that substantial losses of efficiency in separate fitting were generally restricted to configurations in which the probability of the baseline category was low. Agresti (1990) recommended that, unless there was a natural baseline category, the response category having the highest prevalence should be used as a baseline. This is not normally a problem in network data, as zeroes are usually a common, and often the most frequent, value in the network matrix.

5.3. Setting up Data and Equating Parameters

The previous section gives an indication of how to adapt a valued data matrix **X** to allow maximum pseudo-likelihood estimation (MPLE) of the model with standard statistical packages. The procedure is not dissimilar to the procedure for MPLE for multivariate network data (Pattison & Wasserman, in press), except that we have to allow for impermissible data patterns.

For trichotomous data, we establish two binary matrices in accordance with the transformation (1), one for the first logit pertaining to $X_{ij} = 1$, that is, (13), and the other for the second logit pertaining to $X_{ij} = 2$, that is, (14). Each cell in these matrices (excluding diagonals) becomes a separate case in the pseudo-logistic regression, with a "response" variable specified as the binary value of the cell except in those cases where it is deemed to be undefined (that is, missing) in accord with (10). For cell (i, j) in the k-th logit, as "explanatory variables" we calculate the statistics $z_{(ij,k)}^{M+(ij,k)}$ for each $M + (ij, k)$ hypothesized as a clique in the dependence graph.

However, the parameters for certain of these statistics have to be equated, because there is only one parameter $\gamma_{M+(ij,k)}$ for each clique $M + (ij, k)$ in the set of equations in (9). Suppose that the clique $T = M + (ij, k) = S + (st, v)$, where $st \neq ij$ and $v \neq k$, with $(st, v) \in M$ and $(ij, k) \in S$. Then in the logit for k-strength relations, the parameter $\gamma_{M+(ij,k)}$ appears, with statistic $z_{(ij,k)}^{M+(ij,k)}$, whereas in the logit for v-strength relations, the parameter $\gamma_{S+(st,v)}$ appears, with statistic $z_{(st,v)}^{S+(st,v)}$. Yet these parameters refer to the same clique T so they are in fact one and the same, the parameter γ_T. Accordingly, we have a single statistic, which in the data array we set to $z_{(ij,k)}^{M+(ij,k)}$ for the k-th logit, and to $z_{(st,v)}^{S+(st,v)}$ for the v-th logit.

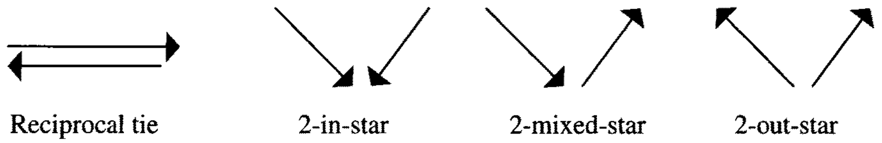
Before we show an example of how this works in practice, we introduce certain statistics that reflect important network configurations. To make the models identifiable, homogeneity constraints of some sort are required. Pattison and Wasserman (in press) introduced a general strategy of assuming that parameters corresponding to certain isomorphic configurations of array entries of **X** are equal.

Frank and Strauss (1986) used such an assumption for their univariate Markov directed graphs. In the first place, they assumed that possible ties were conditionally dependent if they had an actor in common; that is, ties ij and st are conditionally dependent if $\{i, j\} \cap \{s, t\} \neq \emptyset$. By imposing a homogeneity requirement that isomorphic directed graphs have the same probability, Frank and Strauss (1986) showed that the sufficient statistics for the model were counts of certain network configurations: *reciprocal ties*, *in-stars*, *out-stars*, *mixed-stars*, and certain triadic configurations. For our illustrative purposes, we also assume homogeneity and we restrict our attention to the following set of configurations illustrated in Figure 1 (we refer to the *order* of a configuration as the number of ties in it):

Configurations of order 2.

- *Reciprocal or exchange ties*: configurations in which ties ij and ji are both present. If one tie is present at strength $v \geq 1$ and the other at strength $w \geq 1$, the corresponding reciprocal or exchange parameter may be denoted by $\rho^{[v,w]}$, and we note that $\rho^{[v,w]} = \rho^{[w,v]}$.
- *2-in-stars*: configurations in which two separate ties are directed towards the same actor; for instance, ij and kj. The number of 2-in-stars for j is of course related to the number of choices of j made by other actors in the network (the *indegree* of j), and as such is related to the *popularity* of j. We shall use σ to denote star effects and σ_I to denote 2-in-stars in

Configurations of Order 2



Configurations of Order 3

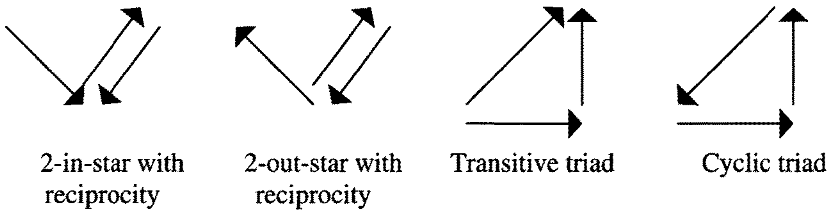


FIGURE 1.
Various network configurations.

particular. When one tie in a 2-in-star is at strength $v \geq 1$ and the other at strength $w \geq 1$, the corresponding parameter may be denoted by $\sigma_I^{[v,w]}$, and we note that $\sigma_I^{[v,w]} = \sigma_I^{[w,v]}$.

- *2-out-stars*: configurations in which two separate ties are directed away from the same actor; for instance, ij and ik. The number of 2-out-stars for i is of course related to the number of choices i makes of other actors in the network (the *outdegree* of i), and is related to the *expansiveness* of i. When one tie in a 2-out-star is at strength $v \geq 1$ and the other at strength $w \geq 1$, the corresponding parameter may be denoted by $\sigma_O^{[v,w]}$, and we note that $\sigma_O^{[v,w]} = \sigma_O^{[w,v]}$.
- *2-mixed-stars*: configurations in which a tie is directed away from an actor toward whom another tie is directed; for instance, ij and jk. When the first tie in a 2-mixed-star is at strength $v \geq 1$ and the other at strength $w \geq 1$, the corresponding parameter may be denoted by $\sigma_M^{[v,w]}$. For 2-mixed-stars $\sigma_M^{[v,w]} \neq \sigma_M^{[w,v]}$ unless $v = w$.

Configurations of order 3.

- *2-in-stars with reciprocity*: configurations which contain a 2-in-star, but with a tie reciprocating one of the ties of the 2-in-star; in other words, configurations with ties of the form ij, jk and kj. When the tie ij is at strength $s \geq 1$, the tie jk is at strength $t \geq 1$, and the tie kj is at strength $v \geq 1$, we denote the corresponding parameter as $\sigma_{IR}^{[s,t,v]}$.
- *2-out-stars with reciprocity*: configurations which contain a 2-out-star, but with a tie reciprocating one of the ties of the 2-out-star; in other words, configurations with ties of the form ij, ik and ki. When the tie ij is at strength $s \geq 1$, the tie ik is at strength $t \geq 1$, and the tie ki is at strength $v \geq 1$, we denote the corresponding parameter as $\sigma_{OR}^{[s,t,v]}$.
- *Transitive triads*: configurations in which three ties comprise a transitive triad of the form ik, kj and ij. When the configuration is such that the tie ik is at strength $s \geq 1$, the tie jk is

at strength $t \geq 1$, and the tie ij is at strength $v \geq 1$, the corresponding parameter may be denoted by $\tau^{[s,t,v]}$.

- *3-cycles* or *cyclic triads*: configurations in which three ties comprise an intransitive cycle; for instance, ij, jk and ki. When the configuration is such that the tie ij is at strength $s \geq 1$, the tie jk is at strength $t \geq 1$, and the tie ki is at strength $v \geq 1$, the corresponding parameter may be denoted $\zeta^{[s,t,v]}$. We note that $\zeta^{[s,t,v]} = \zeta^{[t,v,s]} = \zeta^{[v,s,t]}$.
- In addition to these various effects, there is also a general *choice* or *density* effect for each logit, reflecting the propensity of actors to make choices at each strength level. (This can be considered a configuration of order 1.) This term becomes the intercept term in the conditional logistic regression equation for each logit. When the conditional logit is at strength $v \geq 1$, the corresponding parameter may be denoted by $\theta^{[v]}$.

In Tables 1 and 2, we have presented parameters and statistics pertaining to the configurations specified above, of orders 1 and 2, and of orders 3, respectively. Below, we sometimes refer to configurations in terms of their associated parameters.

With a data set derived along these lines from the original valued matrix **X**, MPLEs for model parameters can be obtained by utilizing any standard logistic regression package. We have written a set of commands in the SPSS *MATRIX* language (Norusis, 1990) that will convert a trichotomously valued matrix into an appropriate data set for the SPSS logistic regression procedure. This command set is available from the authors on request.

6. Empirical Examples

6.1. Student Friendship Network

We illustrate these methods with two examples. The first uses a data set from Vickers (1981) and Vickers and Chan (1981), which was also studied by Wasserman and Pattison (1996) and by

TABLE 1.
Parameters and associated statistics for configurations of orders 1 and 2

Parameter	Statistic (strength-1 logit)	Statistic (strength-2 logit)
$\theta^{[1]}$	1	0
$\theta^{[2]}$	0	1
$\rho^{[1,1]}$	$y_{ji,1}$	0
$\rho^{[1,2]}$	$y_{ji,2}$	$y_{ji,1}$
$\rho^{[2,2]}$	0	$y_{ji,2}$
$\sigma_I^{[1,1]}$	$\sum y_{ki,1}$	0
$\sigma_I^{[1,2]}$	$\sum y_{ki,2}$	$\sum y_{ki,1}$
$\sigma_I^{[2,2]}$	0	$\sum y_{ki,2}$
$\sigma_O^{[1,1]}$	$\sum y_{ik,1}$	0
$\sigma_O^{[1,2]}$	$\sum y_{ik,2}$	$\sum y_{ik,1}$
$\sigma_O^{[2,2]}$	0	$\sum y_{ik,2}$
$\sigma_M^{[1,1]}$	$\sum(y_{ki,1} + y_{jk,1})$	0
$\sigma_M^{[1,2]}$	$\sum y_{jk,2}$	$\sum y_{ki,1}$
$\sigma_M^{[2,1]}$	$\sum y_{ki,2}$	$\sum y_{jk,1}$
$\sigma_M^{[2,2]}$	0	$\sum(y_{ki,2} + y_{jk,2})$

Note. (Summations are over all $k \neq i, j$.)

TABLE 2.
Parameters and associated statistics for configurations of order 3

Parameter	Statistic (strength-1 logit)	Statistic (strength-2 logit)
$\sigma_{IR}^{[1,1,1]}$	$\sum(y_{kj,1}y_{jk,1} + y_{kj,1}y_{ji,1} + y_{ki,1}y_{ji,1})$	0
$\sigma_{IR}^{[1,1,2]}$	$\sum(y_{kj,2}y_{jk,1} + y_{ki,1}y_{ji,2})$	$\sum y_{kj,1}y_{ji,1}$
$\sigma_{IR}^{[1,2,1]}$	$\sum(y_{kj,1}y_{jk,2} + y_{kj,1}y_{ji,2})$	$\sum y_{ki,1}y_{ji,1}$
$\sigma_{IR}^{[2,1,1]}$	$y_{ji,1}\sum(y_{kj,2} + y_{ki,2})$	$\sum y_{kj,1}y_{jk,1}$
$\sigma_{IR}^{[2,2,1]}$	$\sum y_{kj,2}y_{ji,2}$	$\sum(y_{kj,1}y_{jk,2} + y_{ki,2}y_{ji,1})$
$\sigma_{IR}^{[2,1,2]}$	$\sum y_{ki,2}y_{ji,2}$	$\sum(y_{kj,2}y_{jk,1} + y_{kj,2}y_{ji,1})$
$\sigma_{IR}^{[1,2,2]}$	$\sum y_{kj,2}y_{jk,2}$	$y_{ji,2}\sum(y_{kj,1} + y_{ki,1})$
$\sigma_{IR}^{[2,2,2]}$	0	$\sum(y_{kj,2}y_{jk,2} + y_{kj,2}y_{ji,2} + y_{ki,2}y_{ji,2})$
$\sigma_{OR}^{[1,1,1]}$	$\sum(y_{ki,1}y_{ik,1} + y_{ik,1}y_{ji,1} + y_{jk,1}y_{ji,1})$	0
$\sigma_{OR}^{[1,1,2]}$	$\sum(y_{ki,2}y_{ik,1} + y_{ik,1}y_{ji,2})$	$\sum y_{jk,1}y_{ji,1}$
$\sigma_{OR}^{[1,2,1]}$	$\sum(y_{ki,1}y_{ik,2} + y_{jk,1}y_{ji,2})$	$\sum y_{ik,1}y_{ji,1}$
$\sigma_{OR}^{[2,1,1]}$	$y_{ji,1}\sum(y_{ik,2} + y_{jk,2})$	$\sum y_{ki,1}y_{ik,1}$
$\sigma_{OR}^{[2,2,2]}$	0	$\sum(y_{ki,2}y_{ik,2} + y_{ik,2}y_{ji,2} + y_{jk,2}y_{ji,2})$
$\sigma_{OR}^{[2,2,1]}$	$\sum y_{jk,2}y_{ji,2}$	$\sum(y_{ki,1}y_{ik,2} + y_{ik,2}y_{ji,1})$
$\sigma_{OR}^{[2,1,2]}$	$\sum y_{ik,2}y_{ji,2}$	$\sum(y_{ki,2}y_{ik,1} + y_{jk,2}y_{ji,1})$
$\sigma_{OR}^{[1,2,2]}$	$\sum y_{ki,2}y_{ik,2}$	$y_{ji,2}\sum(y_{ik,1} + y_{jk,1})$
$\tau^{[1,1,1]}$	$\sum(y_{ki,1}y_{kj,1} + y_{ik,1}y_{jk,1} + y_{ik,1}y_{kj,1})$	0
$\tau^{[1,1,2]}$	$\sum(y_{ki,1}y_{kj,2} + y_{ik,2}y_{jk,1})$	$\sum y_{ik,1}y_{kj,1}$
$\tau^{[1,2,1]}$	$\sum(y_{ik,1}y_{jk,2} + y_{ik,1}y_{kj,2})$	$\sum y_{ki,1}y_{kj,1}$
$\tau^{[2,1,1]}$	$\sum(y_{ki,2}y_{kj,1} + y_{ik,2}y_{kj,1})$	$\sum y_{ik,1}y_{jk,1}$
$\tau^{[1,2,2]}$	$\sum y_{ik,2}y_{jk,2}$	$\sum(y_{ki,1}y_{kj,2} + y_{ik,1}y_{kj,2})$
$\tau^{[2,1,2]}$	$\sum y_{ki,2}y_{kj,2}$	$\sum(y_{ik,2}y_{jk,1} + y_{ik,2}y_{kj,1})$
$\tau^{[2,2,1]}$	$\sum y_{ik,2}y_{kj,2}$	$\sum(y_{ki,2}y_{kj,1} + y_{ik,1}y_{jk,2})$
$\tau^{[2,2,2]}$	0	$\sum(y_{ki,2}y_{kj,2} + y_{ik,2}y_{jk,2} + y_{ik,2}y_{kj,2})$
$\zeta^{[1,1,1]}$	$\sum y_{ki,1}y_{jk,1}$	0
$\zeta^{[1,1,2]}$	$\sum(y_{ki,2}y_{jk,1} + y_{ki,1}y_{jk,2})$	$\sum y_{ki,1}y_{jk,1}$
$\zeta^{[1,2,2]}$	$\sum y_{ki,2}y_{jk,2}$	$\sum(y_{ki,2}y_{jk,1} + y_{ki,1}y_{jk,2})$
$\zeta^{[2,2,2]}$	0	$\sum y_{ki,2}y_{jk,2}$

Note. (Summations are over all $k \neq i, j$.)

Pattison and Wasserman (in press). Network data were obtained from 29 students in year 7 in a school in Victoria, Australia. Students were asked to nominate their classmates on a number of relations including the following:

- Who do you get on with in the class?
- Who are your best friends in the class?

Since these relations are ordered, with the "best friend" relation contained within the "get on with" relation, it makes sense to regard them as a single, valued Friendship relation, with 3 values:

TABLE 3.
Valued matrix for Friendship relation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	–	0	0	0	0	1	0	2	0	0	2	2	0	1	0	1	1	0	1	0	1	1	0	0	0	1	1	0	0
2	1	–	1	1	2	2	1	2	0	2	2	0	0	1	2	2	0	0	1	1	2	0	0	0	0	0	0	0	0
3	0	0	–	2	0	0	2	2	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	2	–	1	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	2	0	0	–	2	0	2	0	2	2	0	0	1	2	2	0	0	2	1	2	1	1	0	0	0	1	2	1
6	2	2	2	2	2	–	1	2	0	2	2	2	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	0	1
7	0	0	2	0	0	0	–	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	2	2	2	2	2	2	2	–	0	2	2	2	2	2	2	2	2	1	1	2	2	2	1	2	1	1	2	2	2
9	2	2	1	1	0	0	1	2	–	0	2	1	0	1	0	1	1	1	1	0	0	0	1	1	1	2	0	1	1
10	0	2	0	0	2	0	0	0	1	–	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	2	2	2	2	2	2	2	2	0	2	–	2	2	2	1	2	1	1	2	2	2	2	1	1	1	1	1	1	1
12	2	0	0	0	1	1	0	2	0	0	2	–	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	–	2	1	1	1	2	1	2	2	2	2	1	2	1	1	0	1
14	0	1	0	0	1	0	0	0	0	0	1	0	0	–	1	1	1	0	2	2	2	2	2	1	0	2	2	0	0
15	0	0	0	0	1	0	0	0	0	0	0	0	0	2	–	2	0	0	2	2	2	2	2	2	0	1	1	1	1
16	1	1	0	0	1	0	0	0	0	1	1	1	0	2	2	–	1	0	1	2	2	2	2	1	0	1	0	0	0
17	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	–	0	1	0	0	0	0	2	0	1	0	1	2
18	0	0	0	0	0	0	0	1	1	0	0	0	2	0	0	1	0	–	0	0	0	0	0	0	2	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	1	–	2	2	2	2	2	0	1	1	0	0
20	0	1	0	0	1	0	0	0	0	0	1	0	1	2	2	2	0	0	2	–	2	2	2	1	1	1	2	0	0
21	1	1	0	0	1	0	0	0	0	0	1	0	0	2	0	2	0	0	2	2	–	2	2	0	0	0	0	0	0
22	0	1	0	0	1	0	0	0	0	0	1	0	0	2	2	1	0	0	2	2	2	–	2	0	0	2	2	0	1
23	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	0	0	2	2	2	2	–	0	1	1	1	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	1	2	1	2	1	1	1	2	–	1	2	2	2	2
25	0	0	0	0	0	0	0	1	1	0	1	0	2	0	0	0	0	2	0	0	0	0	0	0	–	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	2	0	2	1	0	0	–	2	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	1	1	1	1	0	0	0	2	–	1	1
28	0	1	0	1	1	0	0	1	2	0	0	1	2	0	2	0	1	0	0	1	1	0	2	2	0	1	1	–	2
29	0	1	0	1	0	0	0	1	1	0	0	0	2	0	0	0	1	0	0	0	0	0	2	2	0	1	2	2	–

Note. Derived from Vickers & Chan (1981). Numbers for rows and columns refer to the 29 actors.

- $X_{ij} = 0$ - student *i* does not report either relationship with student *j*;
- $X_{ij} = 1$ - student *i* "gets on with" student *j*, but does not report that student *j* is a best friend;
- $X_{ij} = 2$ - student *i* reports that student *j* is a best friend.

The valued matrix for this relation is given in Table 3.

This model was fitted by a constrained stepwise elimination approach (where lower order effects were retained in the presence of related higher order effects). Initially, we fitted a model that included all parameters relating to first and second order configurations, with third order parameters then forward selected (in our examples, forward selection from second order configurations and backward selection from third order configurations made no difference to the final result.) We then developed a final model by retaining first order parameters, the selected third order parameters together with their lower second order parameters, and with other second order parameters then forward selected.

For this example, the criteria for forward selection were simple: An individual parameter was not included in the model if it did not contribute at least 10 to the change in the pseudo-likelihood deviance and as long as the resulting change in the mean absolute residual was less than 0.005. These particular criteria were chosen as simple "rules of thumb" and could be replaced by more or less stringent alternatives. The difficulty, of course, is that the pseudo-likelihood deviance cannot be assumed to have an asymptotic χ^2 distribution, and so the usual inferential techniques of standard logistic regression are not available here.

Our rationale for choosing these particular criteria is straightforward. The improvement in pseudo-likelihood deviance produced by an initial fit of all 43 parameters was 1031.7, and we decided that variables should contribute at least 1% of this figure to be regarded as important contributors to the model. (With maximum likelihood estimation, this would amount to testing using a χ^2 distribution with one degree of freedom at the 0.001 level, not an unreasonable level given the large number of "cases"—1263). Absolute residuals were used by Wasserman and Pattison (1996) as a guide to model fit, and in our experience with fitting p^* models, a parameter that is not important to the model is likely to change the mean of absolute residuals only in the third decimal place.

Parameter estimates for the final model, with parameters excluded according to these criteria, are presented in Table 4. The model includes 18 parameters with an improvement in pseudo-likelihood deviance of 975.1 (compared to the improvement of 1031.7 for 43 parameters). The mean of absolute residuals for the final model was 0.193, compared with 0.176 for the model with all 43 parameters. Table 4 includes, as an indicative guide only, approximate standard errors for each parameter estimate as calculated by normal logistic regression procedures. These standard error estimates need to be taken as a guide only, for they are not reliable for MPLE (although see Preisler, 1993). There are five parameters retained in the model simply because they are lower order to important higher order effects, rather than because they themselves make a substantial contribution to the pseudo-likelihood deviance. These parameters are $\sigma_I^{[1,2]}$, $\sigma_M^{[2,1]}$, $\sigma_M^{[2,2]}$, $\sigma_O^{[1,1]}$ and $\sigma_O^{[2,2]}$. Our interpretation of the model concentrates on the other thirteen parameters that make a substantial contribution to the pseudo-likelihood deviance.

TABLE 4.
Parameter estimates for final friendship model

Parameters	Pseudo-likelihood estimate	Standard error (approx)
$\theta^{[1]}$	−1.53	0.44
$\theta^{[2]}$	−4.11	0.84
$\rho^{[1,1]}$	1.71	0.25
$\rho^{[1,2]}$	1.39	0.30
$\rho^{[2,2]}$	7.75	1.21
$\sigma_I^{[1,2]}$	0.06	0.04
$\sigma_I^{[2,2]}$	−0.65	0.10
$\sigma_O^{[1,1]}$	−0.10	0.05
$\sigma_O^{[1,2]}$	0.27	0.05
$\sigma_O^{[2,2]}$	0.07	0.05
$\sigma_M^{[1,2]}$	−0.23	0.03
$\sigma_M^{[2,1]}$	−0.06	0.05
$\sigma_M^{[2,2]}$	−0.01	0.05
$\sigma_{OR}^{[1,2,2]}$	−0.28	0.08
$\tau^{[2,2,2]}$	0.77	0.09
$\tau^{[1,2,1]}$	0.33	0.05
$\tau^{[1,2,2]}$	0.43	0.08
$\zeta^{[2,2,2]}$	−0.93	0.20

6.1.1. *Interpretation of the Model*

To simplify our account of the model, we refer to strength 1 ties as weak friendships and to strength 2 ties as strong friendships.

The negative estimates for the density parameters ($\hat{\theta}^{[1]}$ and $\hat{\theta}^{[2]}$) simply indicate that, if students i and j are not jointly involved in any other higher order configurations of ties (i.e., reciprocity, stars or triads) both strong and weak friendships are quite unlikely. This suggests that ties are uncommon in isolation and tend to be involved in higher order configurations.

The strongest effects in the model relate to reciprocity. The large $\hat{\rho}^{[2,2]}$ estimate indicates a big reciprocity effect for strong friendships. If there are no other higher order configurations involved with the tie ij, the model predicts that the logit $\varpi_{ij}^{[2]}$ (i.e., predicting the log odds of a strong friendship for ij, conditional on the tie not being a weak friendship) will have a value of 7.75-4.11=3.64, implying an estimated conditional probability of 0.97. On the other hand, the positive $\hat{\rho}^{[1,2]}$ exchange effect suggests a smaller but still substantial tendency for a strong friendship to be reciprocated by weak friendship. Conditional on the tie ij not being strong, the estimated probability of a weak friendship ij reciprocating a strong friendship ji in the absence of higher order configurations is 0.47.

The estimate for the reciprocity effect for weak friendships ($\hat{\rho}^{[1,1]}$) is also large. Assuming again that no higher order configurations are involved, and conditional on the tie ij not being strong, the estimated probability of a weak ij reciprocating a weak ji is 0.54. On the other hand, if the tie ij is not a weak friendship, the estimated probability of a weak friendship ji being reciprocated by a strong ij is only 0.06.

These instances of the reciprocity/exchange effects illustrate our approach to interpretation, based on assessment of estimated conditional probabilities. What we have done is to use the estimates of the model parameters to compute the probability that the tie ij is present at a particular strength, given that a certain configuration is in place and that the tie is not present at a different strength. In other words, we estimate the probability of ij *completing* a higher order configuration conditional on the appropriate lower order configuration being observed. For instance, the $\rho^{[1,1]}$ effect relates to ij completing the configuration of reciprocated tie of strength 1, given that the configuration (of order 1) of a ji tie at strength 1 is observed.

Of course, when we move to configurations of order 3, it is possible that ij may complete more than one configuration, of orders 3 and 2. In this case, the relevant effects have to be added appropriately. Accordingly, it is often helpful to interpret lower and higher order parameters concurrently.

Interpretation of the negative estimate for the 2-out-star with reciprocity parameter ($\hat{\sigma}_{OR}^{[1,2,2]} = -0.28$) illustrates our approach well. Suppose that a strong reciprocated friendship exists between i and k. Then a weak ij would complete a $\sigma_{OR}^{[1,2,2]}$ configuration, as well as a $\sigma_{O}^{[1,2]}$ configuration. As the estimates for these two parameters are approximately equal but of different signs, they cancel each other out, so the $\hat{\sigma}_{OR}^{[1,2,2]}$ effect has little implication for weak friendships. (A weak ij tie also completes a $\sigma_{M}^{[2,1]}$ configuration, but since the estimate $\hat{\sigma}_{M}^{[2,1]}$ is small, -0.06, there is little effect on the conditional probability of a weak ij tie.) The same situation occurs if i has a weak tie to k and j a strong tie to i, where we are interested in whether i reciprocates the strong tie from j. Again the $\hat{\sigma}_{OR}^{[1,2,2]}$ and $\hat{\sigma}_{O}^{[1,2]}$ effects cancel each other out.

Accordingly, the $\hat{\sigma}_{OR}^{[1,2,2]}$ effect in this model bears on the likelihood of a reciprocated strong friendship, given the presence of a 2-out-star involving both strong and weak ties, for then the $\hat{\sigma}_{O}^{[1,2]}$ effect does not come into play. In this case, an ij tie that completes a $\sigma_{OR}^{[1,2,2]}$ configuration also completes a $\sigma_{M}^{[2,1]}$ configuration - although we noted above that parameter made little contribution to the model and is comparatively small - as well as, of course, completing a reciprocated strong tie ($\hat{\rho}^{[2,2]}$). Because the $\hat{\sigma}_{OR}^{[1,2,2]}$ effect is negative, the implication is that if j has many weak friendships then the tendency for i to reciprocate j's strong friendship choice will be lessened. The implication is that the very large effect for strong tie reciprocation is attenuated if one of the partners has many weak friendships.

Having illustrated our general approach to interpretation, we provide some broad conclusions for the remaining parameters. The negative effect for in-stars involving strong friends ($\widehat{\sigma}_I^{[2,2]} = -0.65$) indicates that in the absence of triads, the more a student is popular as a strong friend, the less likely he or she is to be a strong friend for yet others. This implies that there is something of a ceiling on the popularity of students as strong friends in the absence of triadic effects. This effect, however, is reduced in the presence of triads of strong friends ($\widehat{\tau}^{[2,2,2]} = 0.77$), indicating that the most popular strong friends are likely to be the choice of those who are strong friends with each other. This suggests that there are clusters of strong friends. The presence of an estimated negative cyclic effect ($\widehat{\zeta}^{[2,2,2]}$) implies that there is a degree of hierarchy in these clusters. Suppose, for instance, s and t are mutually strong friends, and that v considers both s and t strong friends. Then, the $\widehat{\zeta}^{[2,2,2]}(-0.93)$ effect makes it relatively less probable that either s or t will reciprocate v's expression of strong friendship.

The negative effect for mixed-stars involving weak and strong friends ($\widehat{\sigma}_M^{[1,2]} = -0.23$) indicates that, in the absence of triads, paths of length 2 with weak friendship as the first, and strong friendship as the second, tie are relatively less likely, but there are triadic effects ($\widehat{\tau}^{[1,2,1]}$ and $\widehat{\tau}^{[1,2,2]}$) that can reduce the mixed star effect. For instance, if j considers k as a strong friend, and if i also considers k a friend (either strong or weak), i is relatively more likely to consider j a weak friend, even though the tie completes the relevant mixed-star. (The $\widehat{\sigma}_O^{[1,2]}$ parameter also comes into play in some of these triads.) In this case there seems to be a balance effect operating, whereby two individuals who agree on their strong friendship choices are somewhat more likely to be friends (in this case, weak friends).

In general, it is notable that most structural effects include strong friendships, with no star or triadic effects involving solely weak friendships. This is hardly surprising, but it is reassuring that our intuitions—that friendship structure is likely to comprise primarily the effects of strong, rather than weak, friendships—are borne out by the model.

6.2. *Interaction Patterns in a Bank Branch*

Our second example is taken from a study of interaction patterns in a number of bank branches (Robins, Pattison & Langan-Fox, 1995). Among a number of sociometric questions, workers in fifteen bank branches in Victoria, Australia, were asked to indicate other individuals in their branch whom they regarded as friends and with whom they had a satisfying relationship at work. Through correspondence analysis, Robins et al showed that the friendship and satisfying working relationship had a similar pattern of responses. These two sociometric questions were combined into a single valued matrix, interpreted as an "affect" network. The "affect" relation between a respondent and each other fellow branch member was coded "0" if the respondent did not include the branch member in responses to either of the sociometric questions, "1" if the respondent included the branch member in response to one of the questions, and "2" if the respondent included the branch member in responses to both questions. For the valued affect matrix, "2" was interpreted as representing ties of strong affect and "1" ties of weak affect. (The full data set can be obtained from the authors on request.)

A trichotomous model was fitted to all bank branches simultaneously, with the set of all possible relational ties defined to include ties only between distinct members of the same branch (see Anderson, Wasserman and Crouch, in press). The number of parameters was reduced in accordance with the method and criteria of the previous example. The final model is presented in Table 5. The model includes 16 parameters and has a change of 1018 in the pseudo-likelihood deviance, with a mean absolute residual of 0.233. This compares with a change in pseudo-likelihood deviance of 1061 and a mean absolute residual of 0.222 for the model including all 43 parameters.

TABLE 5.
Parameter estimates for final affect model

Parameters	Pseudo-likelihood estimate	Standard error (approx)
$\theta^{[1]}$	−1.88	0.24
$\theta^{[2]}$	−2.45	0.35
$\rho^{[1,1]}$	1.60	0.24
$\rho^{[1,2]}$	1.98	0.24
$\rho^{[2,2]}$	2.33	0.37
$\sigma_I^{[1,2]}$	0.29	0.06
$\sigma_I^{[2,2]}$	0.58	0.10
$\sigma_O^{[1,1]}$	0.74	0.05
$\sigma_O^{[1,2]}$	0.51	0.05
$\sigma_O^{[2,2]}$	0.91	0.07
$\sigma_M^{[1,1]}$	−0.19	0.03
$\sigma_M^{[1,2]}$	−0.43	0.05
$\sigma_M^{[2,1]}$	−0.45	0.05
$\sigma_M^{[2,2]}$	−0.99	0.12
$\zeta^{[1,2,2]}$	0.72	0.12
$\zeta^{[2,2,2]}$	1.69	0.29

6.2.1. Interpretation

Four features of these parameter estimates deserve particular comment. First, the substantial reciprocity parameter estimates $(\hat{\rho}^{[1,1]}, \hat{\rho}^{[1,2]}, \hat{\rho}^{[2,2]})$ suggest the presence of strong dyadic dependencies of a kind commonly found for affect ties: the probability of a tie from one branch member to another is enhanced by the existence of a tie from the second to the first, and these effects are more substantial for strong ties.

Second, the positive in- and out-star parameter estimates $(\hat{\sigma}_I^{[1,2]}, \hat{\sigma}_I^{[2,2]}, \hat{\sigma}_O^{[1,1]}, \hat{\sigma}_O^{[1,2]}, \hat{\sigma}_O^{[2,2]})$ suggest that branch members are differentiated both in their tendency to report ties and in their tendency to receive them, with these tendencies more marked for configurations involving stronger ties. Thus, some individual differences in expansiveness and popularity are evident among branch members. It might be useful to look for individual characteristics that could underlie these differences. Third, the negative estimates for all of the mixed-star parameters $(\hat{\sigma}_M^{[1,1]}, \hat{\sigma}_M^{[1,2]}, \hat{\sigma}_M^{[2,1]}, \hat{\sigma}_M^{[2,2]})$ suggest that simple two-path or "bridging" structures (in which a member k receives an affect tie from i, and expresses one towards j) are relatively uncommon. Rather, to the extent that such structures are found, they tend to occur in combination with other configurations, such as in-stars, out-stars and reciprocal ties. Taken together, these estimates suggest some subtle distinctions in the conditions under which affect is expressed and point to the possibility of some interesting "signalling" effects. In particular, the estimates suggest that ties are relatively unlikely to be expressed towards individuals who are expansive in their expression of affect, whereas those who are nominated as recipients of affect ties tend to be more discriminating in the expression of affect.

Fourth, although two-path structures are unlikely, their occurrence is relatively more likely when they are arranged in three-cycles, as the two positive cyclic parameter estimates $(\hat{\zeta}^{[1,2,2]}, \hat{\zeta}^{[2,2,2]})$ suggest. Thus, there is a tendency for some clustering of affect, particularly at the higher value, as has been identified in a number of studies of affect relations (for instance, see the summary in Johnsen, 1986).

It is worth noting that, in combination with the other effects, this clustering takes a somewhat unfamiliar form. In fact, it is particularly noteworthy that positive estimates were obtained for cyclic parameters at the same time as values for transitivity parameters that were close to zero (and therefore excluded in the final model). This pattern suggests that work-related ties have more of a generalized exchange character (Bearman, 1997; Lazega & Pattison, 1998) than is customarily observed in other groups (e.g. see Johnsen, 1986). Indeed, the occurrence of cycles of affect ties suggests that affect ties can be seen in part as a "resource," a valued commodity that is exchanged among a small number of individuals (and, here, exchange is evident among subgroups of two or three people). Bearman (1997) has observed that generalized exchange of resources among three or more individuals leads to greater structural stability than direct exchange of resources in pairs; for affect ties in these small work groups, both forms of exchange are observed.

It should be noted, though, that a model with positive cycle parameters and zero transitivity may still display transitivity effects. For instance, suppose that there are strong ties from i to j and from j to k. Assuming the absence of other effects, the logit of the estimated conditional probability for a strong tie from i to k (completing the transitive triad) is $-2.45 + 0.58 + 0.91 = -0.96$, whereas the same computation for the tie from k to i (completing the cyclic triad) yields $-2.45 - 0.99 - 0.99 + 1.69 = -2.74$. Thus, as the literature on affect ties would predict, the tie from i to k is relatively more likely in these circumstances than the tie from k to i, and this is predicted by the model despite the zero transitivity parameter and the positive cyclic parameter. This discussion highlights the importance of a joint interpretation of parameters in these models.

7. Conclusions

This paper, together with Wasserman and Pattison (1996) and Pattison and Wasserman (in press), have described the class of p^* models for social networks for the cases of single binary networks, multiple binary networks, and single polytomous networks. A generalization to multiple polytomous networks follows directly from these papers.

These papers primarily concern structural effects arising from relational data, although some attribute data have also been introduced through the blockmodeling approaches discussed in the earlier two papers. The generalization of blockmodeling for polytomous networks again follows directly from the earlier work.

To add further to the richness of these models, the next step is to introduce attribute variables more explicitly into the modeling process. We have commenced some early work in this regard in relation to social influence models (Robins, 1997, 1998), where similarity in actor attributes is predicted from network ties. This work not only generalizes further the p^* class of models, but utilizes in a slightly different context the polytomous logit framework introduced here.

The polytomous logit framework also presents an alternative approach to modeling multivariate relations, which can be represented as colored graphs. If, for instance, we have bivariate network data, there are four possible states that ij can take: there can be no tie between i and j, a single tie of either type, or two ties. In the approach of Pattison and Wasserman (in press) to bivariate relations, the modeling is essentially based on the conditional log-odds of a tie of one type being predicted from configurations of ties of each type. The approach adopted here provides an alternative and arguably useful parameterization of these models. In particular, it makes a direct distinction between effects involving so-called *multiplex* ties (that is, configurations in which more than one type of tie links i to j) and those involving just one type of tie.

The discussion of the positivity condition in Appendix A, and the associated need to ignore certain patterns in the three-way array, was directed in a technical way towards the needs of this paper. There are, however, a number of potential applications beyond this immediate framework. For instance, network data have frequently been collected in fixed choice designs by asking respondents to specify a certain number of friends: perhaps, by asking them to list their three or four closest friends. This places a constraint on the resulting network that could be incorporated into models for the data by recognizing that patterns of results with more than three or four responses are impermissible. It may also be possible to utilize this approach to the modeling of missing data, an issue that hitherto has been quite difficult in network analysis.

Appendix A: Positivity Conditions of the Hammersley-Clifford Theorem

In his proof of the Hammersley-Clifford theorem, Besag (1974) imposed a *positivity condition* (first formulated by Hammersley & Clifford, 1971) whereby, if $P(W_i = w_i) > 0$ for each $i \in J$, then $P(W_1 = w_1, W_2 = w_2, \ldots, W_n = w_n) > 0$; that is, if variables individually can take certain values, then they can also take those values jointly. (He also needed the unproblematic assumption, $P(\mathbf{W} = \mathbf{0}) > 0$.) The positivity condition not only ensured that the logarithmic terms in the theorem were properly defined but was used specifically in the proof. We note that the positivity condition is unnecessarily stringent for Besag's proof of the theorem. The expansion of (6) is in fact an application of the Inclusion-Exclusion principle and merely requires that the logarithmic terms be properly defined. In particular, $P(\mathbf{W} = \mathbf{w}) > 0$ is needed, and this is indeed implied by the positivity condition.

However, in proving that only variables forming cliques in the dependence graph result in nonzero terms in (6), Besag (1974) takes various subsets of variables and sets all others to zero. In other words, the probability system has to allow any particular combination of zeroes. The proof works because Besag first considers a system whereby all variables are zero, except for two, W_i and W_j, that are conditionally independent; because of the conditional independence, the LHS of (6) should not be expressible in a functional form that includes W_j, a state of affairs that can only occur if, on the RHS, $\Gamma_{\{i,j\}} = 0$. Similar arguments apply to any subset of variables, resulting in the conclusion that the Γ-terms will be zero unless the variables constitute a clique. Obviously this process assumes that any subset of variables can be simultaneously set to zero.

Besag (1974) assumes that the value 0 is available to any variable. In that case, his positivity condition clearly allows him to set any combination of variables to zero, so that the proof can proceed along the lines of the previous paragraph. Yet the proof is still valid if a less stringent version is applied:

$$P(W_i = k \mid \mathbf{W}^{\#}_{S-\{i\}} = \mathbf{w}^{\#}_{S-\{i\}}, \mathbf{W}^{\#}_{J-S} = 0) > 0 \text{ for all } S \subseteq J \text{ such that } i \in S. \qquad \text{(A1)}$$

In fact, this version—which we shall call the *revised positivity condition*—is specifically what is required for the proof. How does it differ from the Besag (1974) version? It allows that certain patterns of nonzero data may be impossible, yet any subset of variables can be set to zero.

Besag (1974) noted that when some patterns of data were impossible—that is, when $P(\mathbf{W} = \mathbf{w}) = 0$ for some $\mathbf{w}$—then the sample space could be restricted to those patterns that were possible, and the proof of the theorem could then proceed. Yet in this case, the positivity condition does not necessarily hold. If, on the other hand, we adopt the less restrictive condition of equation (A1), we can exclude patterns of data that have zero probability, as long as we can set one or more variables to zero while leaving the remainder unchanged, resulting in a data pattern that has a nonzero probability. It is the capacity to set variables to zero that allows the proof to proceed. This is what the condition in equation (A1) lets us do, but we need bear in mind in these circumstances that the $w^{\#}_{S-\{i\}}$ in that equation may not necessarily take all possible values, depending on those patterns that are permissible.

When might this less stringent condition be needed? In converting valued data to a three-way binary data structure, we clearly wish to avoid data patterns that imply the untransformed polytomous variable has more than one value at the same time. This can be done by excluding certain patterns of data from the binary structure. The proof of the theorem requires only that various combinations of variables can be set to zero given other variable values. The cases that we exclude, on the other hand, are those involving particular combinations of variables with simultaneous values of one, not zero. Whenever we have a permissible pattern of variable values, then one or more of those values can be set to zero, and the resulting pattern remains permissible.

Appendix B: Proof of Theorem 1

We begin with (6):

$$\varpi_{ij,k} = k \sum_{S \subseteq \mathcal{T}-ij} \Gamma_{S+ij}(\mathbf{x}^{\#}_{S+ij}) \prod_{st \in S} x_{st}.$$

We now transfer to the three-way binary array **Y** by transforming **X** through (1).

As X_{st} can take only one value, then either $Y_{(st,v)} = 0$ for all $v \in V$, in the case that $X_{st} = 0$, or $Y_{(st,v)} = 1$ for one and only one $v \in V$. As $y_{(st,v)}$ is nonzero for at most one value $v = x_{st}$, we can write $x_{st} = x_{st} y_{(st,x_{st})}$. We then have

$$\varpi_{ij,k} = k y_{(ij,k)} \sum_{S \subseteq \mathcal{T}-ij} \Gamma_{S+ij}(\mathbf{x}^{\#}_{S+ij}) \prod_{st \in S} x_{st} y_{(st,x_{st})}. \tag{B1}$$

We define a set of new functions indexed by subsets of $\mathcal{T} \times V$. In particular we define $\gamma_M(\mathbf{y}^{\#}_M) = \Gamma_S(\mathbf{x}^{\#}_S) \prod_{s \in S} x_s$ in the case where the set $M \subset \mathcal{T} \times V$ has the form $\{(st, x_{st}) | st \in S \subseteq \mathcal{T}\}$. If M does not take this particular form we define $\gamma_M(\mathbf{y}^{\#}_M)$ to be an arbitrary constant.

Then we have from (B1):

$$\begin{aligned}
\varpi_{ij,k} &= \sum_{S \subseteq \mathcal{T}-ij} \Gamma_{S+ij}(\mathbf{x}^{\#}_{S+ij}) \prod_{st \in S} k x_{st} \prod_{st \in S+ij} y_{(st,x_{st})} \\
&= \sum_{S \subseteq \mathcal{T}-ij} \Gamma_{S+ij}(\mathbf{x}^{\#}_{S+ij}) \prod_{st \in S+ij} x_{st} \prod_{st \in S+ij} y_{(st,x_{st})} \\
&= \sum_{\substack{M \subseteq \{(ef,x_{ef})\} \\ ef \in \mathcal{T}-ij}} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m \in M+(ij,k)} y_m \,.
\end{aligned}$$

When it is not the case that M is of the form $\{(st, x_{st})\}$, there must be an $l = (pq, v) \in M$ such that $v \neq x_{pq}$, and hence $y_{(pq,v)} = y_l = 0$, so that $\prod_{m \in M} y_m = 0$. This statement also applies when $l = (ij, v)$ where $v \neq k$. So we can break down $\mathcal{T} \times V$ as follows:

$$\begin{aligned}
\varpi_{ij,k} &= \sum_{\substack{M \subseteq \{(ef,x_{ef})\} \\ ef \in \mathcal{T}-ij}} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m \in M+(ij,k)} y_m \\
&\quad + \sum_{\substack{M \subseteq \{(ef,v): v \neq x_{ef}\} \\ ef \in \mathcal{T}-ij}} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m \in M+(ij,k)} y_m \\
&\quad + \sum_{M \subseteq \{(ij,v): v \neq k\}} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m \in M+(ij,k)} y_m \\
&= \sum_{M \subseteq \mathcal{T} \times V-(ij,k)} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m \in M+(ij,k)} y_m \\
&= y^{[k]}_{ij} \sum_{M \subseteq \mathcal{T} \times V-(ij,k)} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m \in M} y_m \,.
\end{aligned} \tag{B2}$$

We can reshape the LHS of equation (B2) as follows:

$$\varpi_{ij,k} = \log\left[\frac{P(X_{ij} = k \mid \mathbf{X}^{\#}_{\mathcal{T}-ij} = \mathbf{x}^{\#}_{\mathcal{T}-ij})}{P(X_{ij} = 0 \mid \mathbf{X}^{\#}_{\mathcal{T}-ij} = \mathbf{x}^{\#}_{\mathcal{T}-ij})}\right]$$

$$= \log\left[\frac{P(Y_{(ij,k)} = 1, \mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})} = 0 \mid \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}} = \mathbf{y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}})}{P(Y_{(ij,k)} = 0, \mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})} = 0 \mid \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}} = \mathbf{y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}})}\right].$$

As, from the definition of conditional probability,

$$P(Y_{(ij,k)}, \mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})} \mid \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}})$$
$$= P(Y_{(ij,k)} \mid \mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})}, \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}}) P(\mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})} \mid \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}})$$

we then have:

$$\varpi_{ij,k} = \log\left[\frac{P(Y_{(ij,k)} = 1 \mid \mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})} = 0, \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}} = \mathbf{y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}})}{P(Y_{(ij,k)} = 0 \mid \mathbf{Y}^{\#}_{ij\times(\mathcal{V}-\{k\})} = 0, \mathbf{Y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}} = \mathbf{y}^{\#}_{(\mathcal{T}-ij)\times\mathcal{V}})}\right].$$

This is just the conditional logit for the k strength relation variables $Y_{(ij,k)}$; that is, $\varpi_{ij}^{[k]}$ of (3) but with the value of $Y_{(ij,v)}$ required to be zero when $v \neq k$.

So, we can simply write

$$\varpi_{ij}^{[k]} = y_{(ij,k)} \sum_{M \subseteq \mathcal{T}\times\mathcal{V}-(ij,k)} \gamma_{M+(ij,k)}(\mathbf{y}^{\#}_{M+(ij,k)}) \prod_{m\in M} y_m, \qquad \text{(B3)}$$

where we force $y_m = 0$ when $m = (ij, v)$ for $0 < v \neq k$.

Given that equation (B3) is in binary form, the γ-terms become constants, rather than functions, and $y_{(ij,k)} = 1$, so we have

$$\varpi_{ij}^{[k]} = \sum_{M \subseteq \mathcal{T}\times\mathcal{V}-(ij,k)} \gamma_{M+(ij,k)} \prod_{m\in M} y_m.$$

References

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley and Sons.

Anderson, C.J., & Wasserman, S. (1995). Log multiplicative models for valued social relations. *Sociological Methods & Research, 24,* 96–127.

Anderson, C.J., Wasserman, S., & Crouch, B. (in press). A $p*$ primer: logit models for social networks. *Social Networks.*

Bearman, P. (1997). Generalized exchange. *American Journal of Sociology, 102,* 1383–1415.

Begg, C.B., & Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika, 71,* 11–18.

Besag, J.E. (1972). Nearest neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, Series B, 34,* 75–83.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B, 36,* 192–236.

Besag, J.E. (1975). Statistical analysis of nonlattice data. *The Statistician, 24,* 179–195.

Besag, J.E. (1977a). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Association, 47,* 77–92.

Besag, J.E. (1977b). Efficiency of pseudo-likelihood estimation for simple Gaussian random fields. *Biometrika, 64,* 616–618.

Besag, J.E., & Clifford, P. (1989, May). Generalized Monte Carlo significance tests. *Biometrika, 76,* 633–642.

Crouch, B., & Wasserman, S. (1998). Fitting p*: Monte Carlo maximum likelihood estimation. Paper presented at International Conference on Social Networks, Barcelona, Spain.

Faust, K., & Wasserman, S. (1993). Association and correlational models for studying measurements on ordinal relations. In P.V. Marsden (Ed.), *Sociological methodology 1993* (pp. 177–215). Cambridge, MA: Basil Blackwell.

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

Geyer, C. J., & Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B, 54*, 657–699.

Hammersley, J. M., & Clifford, P. (1971). *Markov fields on finite graphs and lattices.* Unpublished manuscript.

Hosmer, D.W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: John Wiley & Sons.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. [Contribution to the theory of ferromagnetism]. *Zeitschrift fur Physik, 31*, 253–258.

Johnsen, E.C. (1986). Structure and process: agreement models for friendship formation. *Social Networks, 8*, 257–306.

Lauritzen, S.L. (1996). *Graphical models.* Oxford: Clarendon Press.

Lazega, E., & Pattison, P. (1998). Social capital, multiplex generalized exchange and cooperation in organizations: A case study. Submitted to *Social Networks.*

Norusis, M.J. (1990). *SPSS advanced statistics user's guide.* Chicago: SPSS.

Pattison, P., & Wasserman, S. (in press). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology.*.

Preisler, H. (1993). Modeling spatial patterns of trees attacked by bark-beetles. *Applied Statistics, 42*, 501–514.

Rennolls, K. (1995). $p_{\frac{1}{2}}$. In M.G. Everett & K. Rennolls (Eds.), *Proceedings of the 1995 International Conference on Social Networks, Vol. 1.* (pp. 151–160). London: Greenwich University Press.

Robins, G.L. (1997, February). p^* models of social influence. Paper presented at the International Sunbelt Social Network Conference, San Diego, CA.

Robins, G.L. (1998). *Personal attributes in inter-personal contexts: Statistical models for individual characteristics and social relationships.* Unpublished doctoral dissertation, University of Melbourne, Australia.

Robins, G.L., Pattison, P., & Langan-Fox, J. (1995, July). Group effectiveness: A comparative analysis of interactional structure and group performance in organizational workgroups. Paper presented at International Social Networks Conference, London.

Strauss, D. (1992). The many faces of logistic regression. *The American Statistician, 46*, 321–327.

Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association, 85*, 204–212.

Vickers, M. (1981). *Relational analysis: An applied evaluation.* Unpublished Master of Science thesis, Department of Psychology, University of Melbourne.

Vickers, M., & Chan, S. (1981). *Representing classroom social structure.* Melbourne: Victoria Institute of Secondary Education.

Wasserman, S. (1987). Conformity of two sociometric relations. *Psychometrika, 52*, 3–18.

Wasserman, S., & Faust, K. (1989). Canonical analysis of composition and structure of social networks. In C.C. Clogg (Ed.), *Sociological methodology 1989* (pp. 1–42). Cambridge, MA: Basil Blackwell.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York: Cambridge University Press.

Wasserman, S., Faust, K., & Galaskiewicz, J. (1990). Correspondence and canonical analysis of relational data. *Journal of Mathematical Sociology, 15*, 11–62.

Wasserman, S., & Iacobucci, D. (1986). Statistical analysis of discrete relational data. *British Journal of Mathematical and Statistical Psychology, 39*, 41–64.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks. I: An introduction to Markov graphs and p*. *Psychometrika, 60*, 401–425.

Wasserman, S., & Pattison, P. (1999). *Multivariate random graph distributions.* (Springer Lecture Notes Series in Statistics). New York: Springer-Verlag.

Wong, G.Y., & Wang, Y.J. (1995). Exponential models for polytomous stochastic networks. Unpublished manuscript.

Manuscript received 9 NOV 1995
Final version received 14 DEC 1998