# Latent Structure in Multiplex Relations

Carter T. Butts*

3/30/2004

Social relations rarely occur alone, or in a vacuum: members of a social system relate to each other in various distinct ways, a fact whose importance has been recognized at least since Merton (1957). Conventionally, we speak of the structure formed by such a composite of overlapping social ties as being a *multiplex relation*, represented formally by a *multigraph*. Such multiplex relations are subject to – and can thereby reveal – a variety of structural constraints. In addition to the obvious definitional properties (certain relations being defined in terms of the negation or composition of other relations), physical constraints (such as the relationship between exposure and disease transmission ), cultural rules (e.g., courtship norms (King and Christensen, 1983)), and opportunity structures (e.g., informal ties and business deals (Lan, 2002)) can all impinge upon the structure of multiplex ties. Here, we demonstrate a likelihood-based modeling framework which both synthesizes previous approaches (e.g., Friedkin, 1990; Cross et al., 2001; Degenne and Lebeaux, 1996; Liebler and Sandefur, 2002) and facilitates the cross-application of techniques developed for other types of data (e.g., Wiley and Martin, 1999; Martin and Wiley, 2000) while using standard methods. Application of this framework is demonstrated on two "classic" data sets, and related results are discussed.

We begin by assuming a multigraph $\mathbb{G}$ formed from a set of $M$ distinct binary relations, $G_1, \ldots, G_M$, on a common vertex set $V$, such that $G_i = (E_i, V)$.[1] Typically, we will represent $\mathbb{G}$ via the $M \times N \times N$ adjacency array, $\mathbf{X}$. The relations of $\mathbb{G}$ may or may not be directed, and may or may not contain loops – to represent this, we define $\mathbb{D}$ as the set of all ordered pairs $(v_i, v_j) : v_i, v_j \in V$ such that $(v_i, v_j)$ is permissible; in the undirected case, we arbitrarily select one of the pairs $(v_i, v_j), (v_j, v_i)$ to be excluded from $\mathbb{D}$. Given this, it is clear that a given ordered pair in $\mathbb{D}$ must occupy one of $2^M$ distinct states. In particular, each such pair may be associated with a vector $\mathbf{z} \in \{0,1\}^M$ such that $\mathbf{z}_i = 1$ iff $(v_j, v_k) \in E_i$; we will here refer to this vector as the *profile* of $(v_j, v_k)$. Taking the set of all possible profiles, we then form the *profile matrix*, $\mathbf{Z} \in \{0,1\}^{2^M \times M}$. Every row of $\mathbf{Z}$ corresponds to a unique $\mathbf{z}$, and as such can be interpreted as a distinct (possible) dyad state. Clearly, then, $\mathbf{Z}$ is defined only up to a row permutation; for concreteness, we will assume that the rows of $\mathbf{Z}$ are sorted in ascending binary value, but this is not essential.

Given a collection of observations with the above form, it is now natural to inquire as to the frequency with which particular profiles are observed – given the tremendous number of profiles

*Department of Sociology and Institute for Mathematical Behavioral Sciences, University of California, Irvine; Irvine, CA 92697; buttsc@uci.edu

[1] Note that the assumption that $\mathbb{G}$ is formed from a composite of *distinct* relations implies that the multigraphs considered here are a subset of the set of all multigraphs (as considered in the graph theoretic literature).

which obtain for even a modest $M$, it seems quite unlikely to suspect that all possible profiles will be realized within the data (much less at equal rates!). Moreover, the various constraints mentioned above (definitional, physical, cultural, etc.) strongly suggest that patterns will be present in the profile distribution. To investigate this, we begin by defining the vector of *profile marginals*, $\mathbf{y}$, associated with the graph set as follows:

$$\mathbf{y}_i = \sum_{(j,k)\in\mathbb{D}} \prod_{l=1}^{M} \big( \mathbf{X}_{ljk}\mathbf{Z}_{il} + (1-\mathbf{X}_{ljk})(1-\mathbf{Z}_{il}) \big) \tag{1}$$

Clearly, $\sum_{i=1}^{2^M} \mathbf{y}_i = |\mathbb{D}|$. The profile marginals thus contain the frequency distribution for multiplex edge states – that is, they characterize the *types* of relationships which are present in the graph set, as opposed to the *configuration* of edges. The central focus of this paper is the analysis of these profile marginals, and hence of the distribution of relational types within $\mathbb{G}$. Although configural questions are also of interest in many settings, they have been treated in some detail elsewhere (Pattison and Wasserman, 1999), and will not be considered here.

Our analysis of the profile distribution begins by taking $\mathbf{y}$ as generated by a multinomial process, and then seeking parameterizations of that process which are substantively reasonable. We employ latent class analysis (Lazarsfeld and Henry, 1968) as a unifying framework for a range of latent structure models. In addition to the conventional alternatives of unconstrained LCA, independence, and Guttman scaling, we also demonstrate the applicability of a general latent lattice model originally developed by Martin and Wiley (2000) for the representation of belief systems. To extend this model to the profile distribution problem, we begin by postulating a set of $L$ binary *latent relations* on $V$, such that every ordered pair $(v_i, v_j) \in \mathbb{D}$ is associated with a unique latent profile $\mathbf{l} \in \{0,1\}^L$. Taking $\mathbf{r} \in \{0,1\}^M$ to be the corresponding profile of manifest relations, we then define a matrix $\mathbf{D} \in \{0,1\}^{L\times M}$ such that

$$\mathbf{r} = \left(\mathbf{l}^c \otimes \mathbf{D}\right)^c, \tag{2}$$

where $\otimes$ and $\cdot^c$ are the Boolean inner product and complement operators, respectively. $\mathbf{D}$ can be interpreted as expressing the *dependency* of the manifest edges (whose states are given by $\mathbf{r}$) on the latent edges (whose states are given by $\mathbf{l}$); in this, its role is roughly analogous to that of the loading matrix in a factor analytic context. For purposes of identifiability, we assume that the columns of $\mathbf{D}$ are 1-covered, and that no rows of $\mathbf{D}$ are identical – $\mathbf{D}$ is defined only up to a row permutation. Applying this model to the matrix $\mathbf{L}$ of all latent profiles then gives us the set of realizable manifest profiles:

$$\mathbf{R} = \left(\mathbf{L}^c \otimes \mathbf{D}\right)^c. \tag{3}$$

It can be shown from a theorem of Martin and Wiley (2000) that the rows of $\mathbf{R}$ are closed under Boolean intersection, and hence form a lattice. Alternately, given $\mathbf{R}$ we may obtain $\mathbf{D}$ using the Haertel-Wiley inversion (Haertel and Wiley, 1993),

$$\mathbf{D} = \mathbf{R}^{*c}, \tag{4}$$

where $\mathbf{R}^*$ is a matrix whose rows are the meet-irreducible elements of $\mathbf{R}$ (i.e., profiles which cannot be expressed as intersections of other distinct profiles).

Given the above latent structure model, we can infer the presence of latent structure in $\mathbf{y}$ using a confirmatory latent class approach directly analogous to that recommended by Martin and Wiley (2000) for belief data. Specifically, we posit the likelihood

$$p(\mathbf{y}|\theta, \mathbf{e}^+, \mathbf{e}^-, \mathbf{R}) = \prod_{i=1}^{2^M} \left( \sum_{r=1}^{|\mathbf{R}|} \theta_r \prod_{j=1}^{M} \left[ \mathbf{R}_{rj} B \left( \mathbf{Z}_{ij} \left| 1 - \mathbf{e}^-{}_j \right.\right) + (1 - \mathbf{R}_{rj}) B \left( \mathbf{Z}_{ij} \left| \mathbf{e}^+{}_j \right.\right) \right] \right)^{\mathbf{y}_i}, \quad (5)$$

where $\theta \in \mathbb{R}^{|\mathbf{R}|}$, $\mathbf{e}^+, e^- \in \mathbb{R}^M$ are parameters, $|\mathbf{R}|$ is the number of rows in $\mathbf{R}$, and $B$ is the Bernoulli mass function. In this context, $\theta$ can be interpreted as the vector of base rates of occurrence for edges in the manifest relations, $\mathbf{e}^+$ can be interpreted as a vector of false positive rates, and $\mathbf{e}^-$ can be interpreted as a vector of false negative rates; for purposes of identifiability, various constraints are typically placed on $\mathbf{e}^+, \mathbf{e}^-$ (e.g., symmetry, homogeneity) in addition to the standard condition that $\mathbf{e}^+{}_i + \mathbf{e}^-{}_i < 1 \forall i \in 1, \ldots, M$. Maximum likelihood estimates and asymptotic standard errors for the above model can be obtained using standard methods (e.g., the EM algorithm), given $\mathbb{G}$ and a postulated lattice, $\mathbf{R}$. (Following Martin and Wiley (2000), we generally obtain various candidate lattices using exploratory methods, and select the lattice to be employed based on model fit.) Principled comparison of candidate models can then be performed using standard information indices (e.g., AIC, BIC, etc.), allowing evaluation of lattice models not only vis a vis each other, but also with respect to other profile models.

Application of the above modeling approach to two classic data sets (Roethlisberger and Dickson's Bank Wiring Room (Roethlisberger and Dickson, 1939) and Sampson's Monastery (Sampson, 1969)) demonstrates the potential of latent class analysis for uncovering latent structure in multiplex ties. Both data sets are shown to exhibit highly concentrated profile distributions, consistent with the hypothesis of constrained edge structure. In both cases, the observed profile distributions are consistent with lattice-like structure among manifest relations, with certain relations appearing as conjunctions of other relations in an "intersecting scale" pattern. Unconstrained latent class solutions ("profile clustering") and the independence model are shown to provide inferior fits to the observed data. Using the inferred latent lattice structure, approximate necessary and sufficient conditions are shown for the existence of particular relations given other relations in the multigraph; suggestions are made regarding the implications of these conditions for subsequent research.

# 1 References


Cross, R., Borgatti, S. P., and Parker, A. (2001). Beyond answers: Dimensions of the advice network. *Social Networks*, 23(3):215–235.

Degenne, A. and Lebeaux, M.-O. (1996). Boolean analysis of questionnaire data. *Social Networks*, 18(3):231–245.

Friedkin, N. (1990). A Guttman scale for the strength of an interpersonal tie. *Social Networks*, 12(3):239–252.

Haertel, E. H. and Wiley, D. E. (1993). Representations of ability structures: Implications for testing. In Frederiksen, N., Mislevy, R., and Bejar, I., editors, *Test Theory for a New Generation of Tests*, pages 359–384. Erlbaum, Hillsdale, NJ.

King, C. E. and Christensen, A. (1983). The relationship events scale: A Guttman scaling of progress in courtship. *Journal of Marriage and the Family*, 45(3):671–678.

Lan, P.-C. (2002). Networking capitalism: Network construction and control effects in direct selling sales and familiarizing sponsorship. *The Sociological Quarterly*, 43(2).

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. Houghton-Mifflin, New York.

Liebler, C. A. and Sandefur, G. D. (2002). Gender differences in the exchange of social support with friends, neighbors, and co-workers at midlife. *Social Science Research*, 31(3):364–391.

Martin, J. L. and Wiley, J. A. (2000). Algebraic representations of beliefs and attitudes II: Microbelief models for belief data. *Sociological Methodology*, 30:123–164.

Merton, R. K. (1957). *Social Theory and Social Structure*. Free Press, Gelncoe, IL.

Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52:169–193.

Roethlisberger, F. and Dickson, W. (1939). *Management and the Worker*. Cambridge University Press, Cambridge.

Sampson, S. (1969). *Crisis in a Cloister*. Cornell University, Doctoral Dissertation.

Wiley, J. A. and Martin, J. L. (1999). Algebraic representations of beliefs and attitudes: Partial order models for item responses. *Sociological Methodology*, 29:113–146.