Journal of Mathematical Psychology **44**, 536–568 (2000)
doi:10.1006/jmps.1999.1261, available online at http://www.idealibrary.com on IDEAL®

Statistical Evaluation of Algebraic Constraints for Social Networks

Philippa Pattison

University of Melbourne

Stanley Wasserman

University of Illinois

Garry Robins

Deakin University

and

Alaina Michaelson Kanfer

University of Illinois

A multirelational social network on a set of individuals may be represented as a collection of binary relations. Compound relations constructed from this collection represent various labeled paths linking individuals in the network. Since many models of interest for social networks can be formulated in terms of orderings among these labeled paths, we consider the problem of evaluating an hypothesized set of orderings, termed *algebraic constraints*. Each constraint takes the form of an hypothesized inclusion relation for a pair of labeled paths. In this paper, we establish conditions under which sets of such constraints may be regarded as partial algebras. We describe the structure of constraint sets and show that each corresponds to a subset of consistent relation bundles between pairs of individuals. We thereby construct measures of fit for a given constraint set. Then, we show how, in combination with the assumption of various conditional uniform multigraph distributions, these measures lead to a flexible approach to the evaluation of fit of an hypothesized constraint set. Several applications are presented and some possible extensions of the approach are briefly discussed.

The support of the Australian Research Council and the National Science Foundation (Grants SBR93-10184 and SBR96-30754) is gratefully acknowledged, as is the programming assistance of Sarah Ardu. We thank Ece Batchelder, John Boyd, Ron Breiger, Ehtibar Dzhafarov, Laura Koehly, Felicia Trachtenberg, and Doug White for their comments on this research.

Address correspondence and reprint requests to Philippa Pattison, Department of Psychology, University of Melbourne, Parkville, Victoria S052, Australia. E-mail: pattison@psych.unimelb.edu.au, stanwass@uiuc.edu, glr@deakin.edu.au, alaina@ncsa.uiuc.edu.

0022-2496/00 $35.00

1. INTRODUCTION

A *social network* is defined as a set of g social *actors* and a collection of r social *relations* that specify how these actors are related to one another. Examples of social relations are "chooses as a friend," "gets advice from," and "has frequent contact with," recorded for each pair of actors in some set. The interdependence of these social relations is a main concern of researchers who study social networks (e.g., Katz & Powell, 1953; Hubert & Baker, 1978; White, 1984; Wasserman, 1987). Although there are many ways in which social relations can exhibit interdependence, equalities and inclusions can provide much insight into the structure of a social network (Boorman & White, 1976; Boyd, 1990; Pattison, 1993). Such forms of interdependence specify whether one social relation equals another or if actors linked by one social relation are always linked by a second. Are friendships observed at one point of time the same as friendships observed at a later time? Are friends of friends always friends? These equalities and inclusions can be regarded as a collection of *algebraic constraints* on relational ties. The presence of such constraints is of importance in the *algebraic analysis* of social networks.

Indeed, algebraic models based on equalities and inclusions among social relations have been argued to provide a useful, widely-applicable, and theoretically-based multirelational representation of structure in a social network. Recent monographs have reviewed such models in detail (Boyd, 1990; Pattison, 1993; see also Chapters 9–12 of Wasserman & Faust, 1994). It has been noted, however, that there are several practical problems with the direct application of algebraic procedures (Schwartz & Sprinzen, 1984; Pattison & Wasserman, 1995). In particular, algebraic models constructed from observed relational data are known to lack robustness and can change substantially if only a few relational ties change. In addition, algebraic models currently lack a statistical framework. Consequently, there is (at present) no general procedure for evaluating the plausibility of an algebraic model to an observed network; a framework for such evaluations is the primary goal of the paper.

After some necessary notation and definitions presented in Section 2, we describe some algebraic and statistical approaches to the study of interdependencies among social relations (Section 3). A framework for describing sets of equalities and inclusions (termed *constraint sets*) of possible interest is introduced in Section 4. Then, in Section 5, we define an index for quantifying the extent to which a particular constraint set is satisfied by a given set of observed social relations. In Section 6, a collection of probability distributions for relational structures is described, and, in Section 7, these probability distributions are used as the basis of a means of evaluating a constraint set of interest, Some illustrative applications are presented in Section 8.

2. SOME NOTATION AND DEFINITIONS

We let N denote a set of actors: $N = \{1, 2, ..., g\}$. A *dichotomous social relation*, $\mathbf{X}$, is a set of ordered pairs recording relational ties between pairs of actors. If the ordered pair (i, j) is in this set, then the first actor (i) in the pair has a relational *tie* to the second actor (j) in the pair; we write $(i, j) \in \mathbf{X}$. We assume throughout that relational ties are *directed* (i.e., a property of the *ordered* pair (i, j)), but note that it is straightforward to modify our account so as to deal with the case of *non directed* ties (i.e., ties that are a property of the *unordered* pair (i, j)). We also assume that any observed relation $\mathbf{X}$ has no *loops*, that is, ties of the form (i, i), $i \in N$.

Any directed social relation $\mathbf{X}$ can be represented by a (labeled) directed graph on g nodes in which there is a directed edge from i to j if and only if $(i, j) \in \mathbf{X}$ or by a $g \times g$ *adjacency matrix*, $\boldsymbol{X}$, where $X(i, j) = 1$ if $(i, j) \in \mathbf{X}$ and $X(i, j) = 0$, otherwise. In the general, multirelational case, relational ties are recorded for r relations with distinct labels. We assume that these relations, $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_r$, have associated matrices $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_r$ and we let $\mathfrak{R} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_r\}$ denote the entire collection of r relations. We also regard the collection $\mathfrak{R}$ as defining a (directed) *multigraph* $\mathbf{M}$ on N, represented by a 3-way matrix $\mathbf{M}$ in which $M(k, i, j) = X_{\mathbf{k}}(i, j)$. The multigraph comprises the set N of nodes and the set of labeled edges that link one node to another. (For each ordered pair, there are between 0 and r such labeled edges.) In matrix expressions, we use replacement of a subscript by the symbol "$+$" to denote summation over the subscript that the "$+$" replaces. Throughout the paper, we regard the matrices as random variables, with particular probability distributions (described later).

In some cases, it is useful to augment the collection $\mathfrak{R}$ (and hence the multigraph $\mathbf{M}$) with the converses and/or complements of relations in $\mathfrak{R}$. The *converse* of a binary relation $\mathbf{X}$, denoted by $\mathbf{X}'$, is the set of ordered pairs $\{(i, j): (j, i) \in \mathbf{X}\}$. The *complement* of $\mathbf{X}$, denoted by $\mathbf{X}^c$, is the set of ordered pairs $\{(i, j) : (i, j) \notin \mathbf{X}\}$.

Two binary relations $\mathbf{X}$ and $\mathbf{Z}$ are *equal* if they link exactly the same pairs of actors; that is, $\mathbf{X} = \mathbf{Z}$ if and only if $X(i, j) = Z(i, j)$ for all pairs (i, j). Social relations can also be partially ordered by an *inclusion* relation. One relation is *included* in another if all ordered pairs in the first are also included in the second (we also say that the second *entails* the first). Thus, we write $\mathbf{X} \subseteq \mathbf{Z}$ if and only if, whenever $(i, j) \in \mathbf{X}$, then also $(i, j) \in \mathbf{Z}$. In this case, $\mathbf{Z}$ is said to *include* $\mathbf{X}$. If two relations are included in each other then they are equal: that is, $\mathbf{X} \subseteq \mathbf{Z}$ and $\mathbf{Z} \subseteq \mathbf{X}$ if and only if $\mathbf{X} = \mathbf{Z}$. If $\mathbf{Z} \subseteq \mathbf{X}$, we also write $\mathbf{X} \supseteq \mathbf{Z}$; below, we use the two expressions interchangeably.

The *labeled* directed graphs of two relations on the same node set N are *isomorphic* if there is a directed edge from i to j in one graph if and only if there is a directed edge from i to j in the other (i.e., if the matrices of the relations are identical). The *unlabeled graph* of a relation $\mathbf{X}$ on N is defined as the labeled graph from which the node labels have been removed. We denote the unlabeled graph derived from the graph of the relation $\mathbf{X}$ as $A(\mathbf{X})$. The unlabeled graphs of two relations defined on g nodes are *isomorphic* if there are (possibly different) assignments of the labels 1, 2, ..., g to the nodes of each graph that make the resulting labeled graphs isomorphic.

We also define an *association table* $\boldsymbol{n}_{\mathbf{Y}_1\mathbf{Y}_2\cdots\mathbf{Y}_s}$ for a set $\{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_s\}$ of $g \times g$ discrete-valued matrices (Pattison & Wasserman, 1995). Let $\boldsymbol{Y}_k$ take values in the set

$\{0, 1, ..., c_k - 1\}$, and let $\mathbf{n}_{\mathbf{Y}_1\mathbf{Y}_2\cdots\mathbf{Y}_s}(m_1, m_2, ..., m_s)$ equal the number of pairs (i, j) such that

$$Y_1(i, j) = m_1,\ Y_2(i, j) = m_2, ..., Y_s(i, j) = m_s;$$
$$i, j \in N; \qquad m_k \in \{0, 1, ..., c_k - 1\}; \quad k \in \{1, 2, ..., s\}.$$

We note that the number of cells in the association table $\boldsymbol{n}_{\mathbf{Y}_1\mathbf{Y}_2\cdots\mathbf{Y}_s}$ is $c_1 c_2 \cdots c_s$. When the set comprises binary social relations on N, each of the 2^scells in the association table corresponds to a possible *relation bundle* linking one actor to another. A *relation bundle* (White & Reitz, 1983) is a vector of values for the relations in the set for the ties from i to j: a value of 0 in the vector means that the corresponding relational tie is absent, while a value of 1 indicates that the relational tie is present. Thus, each cell in $\boldsymbol{n}$ is a count of the number of pairs of nodes whose relational ties are described by a particular relation bundle (and the total of the counts in the association table is then $g(g-1)$).

Another case of interest (discussed below) arises when the set has the form $\{\boldsymbol{X}, \boldsymbol{K}\}$, where $\boldsymbol{X}$ is the matrix of a binary relation and $\boldsymbol{K}$ is a discrete-valued matrix with entries in the set $\{0, 1, ..., C-1\}$. Then $\boldsymbol{n}_{\mathbf{XK}}$ is a $2 \times C$ table whose entries count the number of ordered pairs of nodes with each possible profile of values from $\boldsymbol{X}$ and $\boldsymbol{K}$.

3. ASSESSING ASSOCIATION—EQUALITY AND INCLUSION

Both algebraic and statistical approaches to examining association among social relations have been developed. These developments have largely been motivated by the search for equalities and, to a lesser extent, by the attempt to identify inclusions.

Algebraic Approaches to Assessing Association

In social network analysis, an algebraic approach to describing the associations among social relations was pioneered by Lorrain and White (1971) and stems from the algebraic approach to modeling kinship relations (White, 1963; Boyd, 1969). This approach relies on a *semigroup* constructed from relations of equality or inclusion among labeled walks in the network. Hypothesized associations among these sets of labeled walks have been the basis of a number of characterizations of structure in networks (e.g., Boorman & White, 1976; Boyd, 1990; Cartwright & Harary, 1979; Granovetter, 1973; Pattison, 1993).

To be more specific, let $\mathbf{A}$ and $\mathbf{B}$ be any relations in a collection $\mathfrak{R}$ defined on the node set N. There is a *walk* with the *label* $\mathbf{AB}$ from actor i to actor j if there exists some actor m for whom $(i, m) \in \mathbf{A}$ and $(m, j) \in \mathbf{B}$; we write $(i, j) \in \mathbf{AB}$ and refer to the set of ordered pairs linked by a walk $\mathbf{AB}$ as the *composition* of the relation $\mathbf{A}$ with the relation $\mathbf{B}$ ($\mathbf{AB}$ is also termed a *compound* relation). Compositions can, in turn, be defined on compound relations and it can readily be shown that the composition operation is *associative*: that is, $(\mathbf{AB})\,\mathbf{C} = \mathbf{A}(\mathbf{BC})$ for any relations $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and we can write both products as $\mathbf{ABC}$ without ambiguity. In general, we let $\mathfrak{I}_p$ denote the set comprising all relations of the form $\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_h$ constructed as the composition of p or fewer relations ($\mathbf{A}_m \in \Re$, $m = 1, 2, ..., h$; $h \leqslant p$); the relation $\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_h$ corresponds to a labeled walk in the network in which actor i is linked by relation $\mathbf{A}_1$ to some actor m_1, actor m_1 is linked by relation $\mathbf{A}_2$ to some actor m_2, and so on, with the final step in the walk being a link of type $\mathbf{A}_h$ from some actor m_{h-1} to actor j. We also let $\Im$ denote the collection of all relations that are the composition of a finite number of relations from $\Re$. (Note that $\Im_1 = \Re$.)

As noted earlier, equalities and inclusions among members of $\Im$ have been used to investigate structure in social networks. The equality relation defined earlier, termed the *axiom of quality* by Boorman and White (1976), leads to the definition of a *partial semigroup* on $\Im_p$ and a *semigroup* on $\Im$. We refer to these algebraic structures as the partial semigroup or semigroup *generated by* the relation set $\Re$ (or the multigraph **M**). The inclusion relation gives rise to the definition of a partial partially ordered semigroup on $\Im_p$ and a partially ordered semigroup on $\Im$.[1] A *partially ordered semigroup* includes an *isotone* ordering relation $\supseteq$ as well as an associative binary operation. (The inclusion relation is *isotone* with respect to composition since $\mathbf{A} \supseteq \mathbf{B}$ implies $\mathbf{CA} \supseteq \mathbf{CB}$ and $\mathbf{AC} \supseteq \mathbf{BC}$ for any relation $\mathbf{C}$ in the relevant collection of relations.) Equalities and inclusions specify constraints among labeled walks in the network that often correspond (loosely) to natural-language claims about the possible structure of network relations (e.g., "the close friend of a close friend is unlikely to be a stranger") and they are expressed in terms of structures that are likely to be important to possible social processes occurring as a result of interactions among network members (i.e., in terms of network walks). As a result, it is not surprising that they have been used to formalize hypotheses about network structure.

Algebraic models of social relations rely, however, on perfect associations and orderings among relations. For example. if two relations contain identical ordered pairs, except for one, they are not equal. Consequently, a single "error" in the network relations can obscure an association, result in the identification of an incorrect association, and possibly lead to the propagation of errors in many compound relations (Schwartz & Sprinzen, 1984). Yet, in many applications, it is unreasonable to assume error-free observations and so it is of interest to develop the means of investigating relational interdependencies in a stochastic framework.

One common solution to the lack of robustness of semigroup models has been to examine association among blockmodels (Breiger, Boorman, & Arabie, 1975; White, Boorman, & Breiger, 1976) of the relations, rather than among the relations themselves. A *blockmodel* for a collection of relations comprises an *a priori* partition on the node set N and an hypothesis about the presence of ties of each type

[1] A *semigroup* is a set S and a binary operation f: $S \times S \to S$ which is *associative* (that is, satisfies $f(f(s, t), u) = f(s, f(t, u))$). The element $f(s, t)$ is often termed the *product* of s and t. A *partial binary operation* is an operation that is defined for only a subset of $S \times S$. A *partial semigroup* is a set S and a partial binary operation f that satisfies $f(f(s, t), u) = f(s, f(t, u))$ whenever both sides are defined. A set S and a relation $\geqslant$ define a *partially ordered set* if the relation $\geqslant$ is reflexive (i.e., $s \geqslant s$, for all $s \in S$) and transitive (i.e., $s \geqslant t$ and $t \geqslant u$ implies $s \geqslant u$; $s, t, u \in S$). A *partially ordered semigroup* is a semigroup S with a partial order $\geqslant$ that is *isotone* with respect to the binary operation f, that is, $s \geqslant t$ implies $f(s, u) \geqslant f(t, u)$ and $f(u, s) \geqslant f(u, t)$. A *partial partially ordered semigroup* comprises a partial semigroup and a partial order that is isotone, whenever relevant products are defined.

between pairs of classes in the partition. (This form of blockmodel is often termed an *a priori* blockmodel and distinguished from an *a posteriori* blockmodel that is constructed from the network data; for example, see Wasserman & Anderson, 1987). In graph theoretical terms, an *a priori* blockmodel is simply a multigraph defined on the classes of the hypothesised partition of the node set (with loops permitted). We let φ represent the mapping from N onto the set $N_{\mathbf{B}}$ of hypothesized classes, and we represent an *a priori* blockmodel **B** by the three-way array $\boldsymbol{B}$.

Multigraph Homomorphisms and Blockmodel Algebras

One natural *a priori* blockmodel associated with any partition on the node set N is that in which a tie of some type is defined to link one class to another if there is a tie of that type from *any* node in the first class to *any* node in the second. In this case, the *a priori* blockmodel is completely determined by the node mapping φ and defines a multigraph *homomorphism* from **M** to **B**: $B(k, i, j) = 1$ if and only if $M(k, l, m) = 1$ for some $l, m \in N$ such that $\varphi(l) = i$ and $\varphi(m) = j$, and $B(k, i, j) = 0$, otherwise. A number of authors have examined the relationship between the algebra associated with a multigraph and the algebra generated by a homomorphism of its multigraph (e.g., Bonacich, 1983; Boyd, 1990; Kim & Roush, 1984; Lorrain & White, 1971; Pattison, 1982, 1993; White & Reitz, 1983). Although there is no necessary general relationship between the two algebras, Kim and Roush (1984) have established the most general conditions on a multigraph homomorphism under which all equations and inclusions that characterise the algebra of the multigraph also hold in the algebra associated with the multigraph homomorphism. Some special cases of these conditions that we invoke below include:

1. φ defines a *structural equivalence* (*SE*) on N iff $\varphi(i) = \varphi(j)$ implies $M(k, i, l) = M(k, j, l)$ and $M(k, l, i) = M(k, l, j)$, for all $l \in N$; $i, j \in N$;

2. φ defines an *automorphic equivalence* (*AE*) on N iff $\varphi(i) = \varphi(j)$ implies there is an automorphism α on the nodes of **M** such that $\alpha(i) = j$, and $M(k, i, l) = M(k, j, \alpha(l))$ and $M(k, l, i) = M(k, \alpha(l), j)$, for all $l \in N$; $i, j \in N$;

3. φ defines an *outdegree equivalence* (*OE*) on N iff $\varphi(i) = \varphi(j)$ implies $M(k, i, l) = 1$ iff $M(k, j, h) = 1$, for some h such that $\varphi(h) = \varphi(l)$; $i, j \in N$;

4. φ defines an *indegree equivalence* (*IE*) on N iff $\varphi(i) = \varphi(j)$ implies $M(k, l, i) = 1$ iff $M(k, h, j) = 1$, for some h such that $\varphi(h) = \varphi(l)$; $i, j \in N$; and

5. φ defines a *regular equivalence* (*RE*) on N iff it is both an outdegree equivalence and an indegree equivalence.

In the case of structural equivalence, the algebra of the blockmodel induced by the equivalence is necessarily identical to the algebra of the original multigraph but, in the other cases, the multigraph algebra may possess additional equations and inclusions. Pattison (1982, 1993) described a general procedure for decomposing a multigraph algebra into such simpler algebraic components, each of which could be associated with some multigraph homomorphism. We do not describe this decomposition in detail, but note that the analysis is relevant to the task of evaluating algebraic constraints in the following sense. Each of the algebraic components in the decomposition is described by equalities and inclusions which do not hold across the entire network; rather, these equalities and inclusions characterize relational ties within and between certain classes of nodes in the network (precisely, those classes defining the associated multigraph homomorphism). Thus, in developing a framework for the evaluation of a set of equalities and inclusions, we need to recognize that the equalities and inclusions may not hold universally, but may instead be descriptive of ties linking certain classes of nodes in the network. We return to this issue in Section 7.

Statistical Approaches to Examining Association

Despite our relatively detailed understanding of the relationship between the algebra of a multigraph and the algebra induced by certain multigraph homomorphisms, less is known about the algebra associated with a multigraph homomorphism, in general. Moreover, in replacing the multigraph by its homomorphic image, information about how well the image represents the original multigraph is lost. Schwartz and Sprinzen (1984) recognized the nonrobust nature of algebraic constructions and presented a series of indices designed to quantify the amount of equality and inclusion between two relations and to identify what they termed "structurally weak" ties. Their aim was to identify algebraic structures that were characteristic of many of the ties in the network. To do so, they introduced a statistical framework to the evaluation of equality and inclusion relations. Similarly, in developing *material entailment analysis*, White (1984, 1996; also White & McCann, 1988) sought to study inclusions (termed *entailments*) among a set of relational ties, stressing the need for a statistical approach to the complete system of associations.

Several other relevant statistical approaches have also been proposed. In an early contribution, Katz and Powell (1953) defined indices of conformity and concordance for two dichotomous relations from the 2×2 association table for the relations. Katz and Powell argued that, under the hypothesis of no association, the distribution of ties on one relation should be independent of the distribution on the other, so that association between the relations can be inferred if the hypothesis of independence in the 2×2 table is rejected (using either Fisher's exact test or a chi-squared approximation). Of course, this approach also entails the assumption that ties are independent of one another, which may be difficult to justify in many instances (for example, where an individual actor provides the information about all ties emanating from him-or herself).

Hubert and Baker's (1978) approach did not assume tie-independence. They constructed indices of conformity by regressing one relation on another and employed a permutation distribution to estimate the probability of the observed magnitude of an index, given an hypothesis of no association between the relations. They adopted the procedure developed by Mantel (1967) that builds a distribution by permuting the rows, and simultaneously the columns, of one of the matrices. The index of association is calculated for each of the new pairs of relations and, under the null hypothesis that the relations are independent of one another, all values of the index are equally likely. If the observed index is extreme in this distribution, then the hypothesis of no association is rejected.

Both of these approaches bear some similarity to the method proposed by Holland and Leinhardt (1975) for evaluating the tendency for certain triadic configurations (i.e., configurations involving three actors) to be observed in a social relation. Holland and Leinhardt examined the distribution of the frequency of these configurations in the set of all relations having the same *dyadic* structure as the observed relation. They argued that a particularly high (or low) frequency of a triadic configuration in the observed relation compared to other relations with the same dyadic structure provided evidence that the frequency of its occurrence could not be explained by the lower-order dyadic features of the relation.

Indeed, a number of authors have argued that evaluations such as these should take account of certain lower-order properties of the relations, such as differential relational density, dyadic structure, and individual differences in the reporting of certain relations (e.g., Schwartz & Sprinzen, 1984; Snijders, 1991; Snijders & Stokman, 1987; van de Bunt, van Duijn, & Snijders, 1995; Wasserman, 1987). Katz and Powell's (1953) method for assessing significance controls only for the densities of the relations. The permutation distribution used by Hubert and Baker controls for all properties of the relations that are independent of individual identities, including the relational densities, the degree to which ties are mutual or asymmetric in each relation, and the degree to which the relations exhibit transitivity. As Pattison and Wasserman (1995) observed, the rationale for controlling for various lower-order features may be either design-based or driven by the theoretical context and, as the above examples illustrate, these features may take a variety of forms. At this stage, we note that there exists no overarching strategy for assessing equality and inclusion among dichotomous relations, with the option of controlling for various lower order effects.

4. ALGEBRAIC CONSTRAINTS

Before outlining such a strategy, we first present a framework for the representation of a set of equalities and inclusions among a set of relations that we might be interested in evaluating.

Constraint Sets

A *constraint* is defined as an hypothesized inclusion relation between two relations on the set $\mathfrak{I}_p$ of compound relations of length no greater than p. A *constraint set* is a collection of such hypothesized inclusion relations and may be written in the form

$$\Gamma = \{(\mathbf{Y}_a, \mathbf{Z}_a) : \mathbf{Y}_a \supseteq \mathbf{Z}_a \text{ and } \mathbf{Y}_a, \mathbf{Z}_a \in \mathfrak{I}_p\}.$$

The subscript a simply indexes the pairs of relations that are hypothesized to be constrained in the set Γ. We regard Γ as a set of inclusion relations that are postulated to hold for members of a set $\mathfrak{R}$ of relations. Equality relations of the form $\mathbf{Y}_a = \mathbf{Z}_a$ are represented in Γ by the presence of the *pair* of ordered pairs $(\mathbf{Y}_a, \mathbf{Z}_a)$ and $(\mathbf{Z}_a, \mathbf{Y}_a)$. Some examples of constraint sets for some familiar network

TABLE 1

Constraints Sets for Some Networks Models

Model	$\mathfrak{R}$	k	Γ
Transitivity	**X**	2	(**X**, **XX**)
Strong-weak ties	**S**, **T**	2	(**S**, **SS**)(**SS**, **S**)(**T**, **ST**)(**T**, **TS**)(**ST**, **T**)(**TS**, **T**) (**TT**, **T**)(**T**, **S**)(**T**, **TT**)
Balance	**P**, **N**	2	(**P**, **PP**)(**PP**, **P**)(**P**, **NN**)(**NN**, **P**)(**N**, **PN**) (**PN**, **N**)(**N**, **NP**)(**NP**, **N**)($\mathbf{P}^c$, **N**)($\mathbf{N}^c$, **P**)
First Letter Law	**A**, **B**	2	(**A**, **AA**)(**A**, **AB**)(**B**, **BA**)(**B**, **BB**)(**AA**, **A**) (**AB**, **A**)(**BA**, **B**)(**BB**, **B**)
Last Letter Law	**A**, **B**	2	(**A**, **AA**)(**A**, **BA**)(**B**, **AB**)(**B**, **BB**)(**AA**, **A**) (**BA**, **A**)(**AB**, **B**)(**BB**, **B**)
Davis clustering model	**P**, **N**	2	(**P**, **PP**)(**PP**, **P**)(N, **PN**)(**PN**, **N**)(**N**, **NP**) (**NP**, **N**)(**NN**, **N**)($\mathbf{P}^c$, **N**)($\mathbf{N}^c$, **P**)

models (presented and defined in Pattison, 1993) are shown in Table 1. Two of the more familiar examples in the table are: (a) transitivity (which involves just a single relation, say **X**, and the relation compounded with itself, i.e., **XX**) and (b) constraints associated with Granovetter's (1973) *strength of weak ties* hypothesis (which involves the association of ties **S** hypothesized to be strong, such as close friendship, with ties **T**, such as acquaintanceship, hypothesized to be weak; see Breiger & Pattison, 1978). We illustrate some of the techniques described below on these two constraint sets.

As the examples in Table 1 indicate, constraint sets may overlap. Indeed, we can define a partial ordering on constraint sets using set-inclusion. Thus, if Γ_1 and Γ_2 are constraint sets for a set $\mathfrak{R}$ of relations, then we define

$$\Gamma_1 \subseteq \Gamma_2 \quad \text{iff} \quad (\mathbf{Y}, \mathbf{Z}) \in \Gamma_1 \quad \text{implies} \quad (\mathbf{Y}, \mathbf{Z}) \in \Gamma_2 \quad \text{for any} \quad (\mathbf{Y}, \mathbf{Z}) \in \Gamma_1.$$

We say that a model with the constraints of set Γ_2 is *nested* in a model with the constraints of set Γ_1. As we describe in the Appendix, constraint sets may be seen as generalizations of the algebras and partial algebras described by Boorman and White (1976), Boyd (1990), Pattison (1993), and Pattison and Wasserman (1995).

5. ASSESSING CONSTRAINTS

In relational data that are subject to random variation, we do not expect constraint sets to be exactly satisfied. We therefore construct an index that quantifies adherence to the constraint sets. We can write the relations in $\mathfrak{I}_p$ as an ordered list $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_w\}$; the first r of these are the observed, primitive relations and clearly $w = r + r^2 + \cdots + r^p$. (The relations $\mathbf{X}_{r+1}$, $\mathbf{X}_{r+2}$, ..., $\mathbf{X}_w$ are all possible compound relations constructed as compositions of no more than p observed relations). The question we now consider is how to quantify the degree of conformity of relations in a given data set $\mathfrak{R}$ to the set of constraints Γ.

An Index for a Constraint Set

In order to assess the degree to which an observed set of relations $\mathfrak{R}$ satisfies the constraints in Γ, we use the 2^w association table $\mathbf{n}$ defined by the relations in $\mathfrak{I}_p$. (Note that we omit the subscript on $\mathbf{n}$ when it is clear which set of relations is used to construct the association table). Each cell in the array corresponds to a relation bundle and can be classified as being consistent with the constraints in Γ or not. In particular, we describe the cell $\mathbf{n}(m_1, m_2, ..., m_w)$ of the table as *consistent* with Γ if each constraint of the form $\mathbf{Y}_a \supseteq \mathbf{Z}_a$ in Γ holds for its relation bundle. For example, the constraint set for transitivity for a relation $\mathbf{S}$ is $\Gamma_T = \{(\mathbf{S}, \mathbf{SS})\}$, defined on $\mathfrak{I}_2 = \{\mathbf{S}, \mathbf{SS}\}$. The corresponding association table $\mathbf{n}_{\{\mathbf{S}, \mathbf{SS}\}}$ of counts cross-classifies $\mathbf{S}$ and $\mathbf{SS}$; it is a 2×2 table with cells $\mathbf{n}(0, 0)$, $\mathbf{n}(0, 1)$, $\mathbf{n}(1, 0)$, and $\mathbf{n}(1, 1)$. The cell $\mathbf{n}(0, 1)$ gives the number of relational ties for which an individual i chooses an individual j on $\mathbf{SS}$ but not on $\mathbf{S}$. The ties falling in this cell are not consistent with the constraint Γ_T, since the inclusion $\mathbf{S} \supseteq \mathbf{SS}$ fails to hold. The cells $\mathbf{n}(0, 0)$, $\mathbf{n}(1, 0)$, and $\mathbf{n}(1, 1)$, on the other hand, are consistent with the constraint $\mathbf{S} \supseteq \mathbf{SS}$.

We define an *indicator matrix* $\mathbf{I}$ to specify which cells of $\mathbf{n}$ are consistent with Γ and which are not. It is the same size as the association table and has cells:

$$\mathbf{I}(m_i, m_2, ..., m_w) = 1 \quad \text{if the cell } n(m_1, m_2, ..., m_w) \text{ is not consistent with } \Gamma$$
$$= 0 \quad \text{otherwise.}$$

We note that $\mathbf{I}$ may be regarded as the intersection of matrices $\mathbf{I}_a$, where $\mathbf{I}_a$ corresponds to the ath constraint $\mathbf{Y}_a \supseteq \mathbf{Z}_a$ in Γ:

$$\mathbf{I}_a(m_i, m_2, ..., m_w) = 1 \quad \text{if the cell } n(m_1, m_2, ..., m_w) \text{ is not consistent with } \mathbf{Y}_a \supseteq \mathbf{Z}_a \text{ and}$$
$$= 0 \quad \text{otherwise.}$$

Thus, if $\mathbf{Y}_a = \mathbf{X}_h$ and $\mathbf{Z}_a = \mathbf{X}_m$, then $\mathbf{I}_a(m_1, m_2, ..., m_w) = 0$ iff $m_h \geqslant m_m$; otherwise, $\mathbf{I}_a(m_1, m_2, ..., m_w) = 1$.

A simple index of the extent to which the constraints in Γ hold in a particular set $\mathfrak{R}$ of relations can be constructed from the indicator matrix $\mathbf{I}$ and the table $\mathbf{n}$ of counts. In particular,

$$v(\Gamma) = \Sigma \mathbf{I}(m_1, m_2, ..., m_w)\, n(m_1, m_2, ..., m_w)$$

is a count of the number of ordered pairs of actors whose relational ties are not consistent with the constraints in Γ.[2] Moreover, the constraints in Γ are exactly true if and only if $v(\Gamma) = 0$.

[2] The measure $v(\Gamma)$ is a simple count of the number of pairs of actors whose relation bundles are in violation of the constraint set Γ. More complex measures could also be constructed, reflecting the *degree* to which a relation bundle violates Γ (for example, distance to the closest nonviolating bundle). Ordered pairs of actors are chosen as the unit over which violations are summed, since constraints in Γ are assumed to hold for *every* possible tie. No single choice of a measure can be optimal for all situations.

The entire collection of cells of **n** can also be partitioned according to patterns of consistency with the individual constraints in Γ. That is, **I** can be expressed as the union of nonoverlapping subsets $\{T_j\}$ of cells of the association table **n**, each with exactly the same pattern of (in) consistencies with constraints in Γ. Cells of **n** that are in the set T_j are inconsistent with exactly the same subset of constraints of the form $(\mathbf{Y}_a, \mathbf{Z}_a)$ in Γ. Define for each such set T the statistic

$$v_T(\Gamma) = \Sigma_{\text{cells} \in T} \mathbf{I}(m_1, m_2, ..., m_w)\, n(m_1, m_2, ..., m_w)$$

which records the number of ordered pairs of actors whose relation bundles are in the set T and are in violation of Γ. If the statistic $v_T(\Gamma)$ is large, then the observed relations are inconsistent with the combination of constraints characterized by the cells in T. Clearly,

$$v(\Gamma) = \sum_{T \in \{T_j\}} v_T(\Gamma)$$

is a decomposition of the global statistic $v(\Gamma)$ and provides a detailed index of lack of fit of the constraint set Γ.

It is occasionally also useful to evaluate each constraint $(\mathbf{Y}_a, \mathbf{Z}_a)$ in Γ: we denote the corresponding statistic by $v_a(\Gamma)$. In general, the larger a component $v_T(\Gamma)$ of the vector $v(\Gamma)$, the more likely it is that the constraints that are inconsistent with the set T of cells do not hold for the observed network data. Similarly, a large value of $v_a(\Gamma)$ suggests little support for the ath constraint of Γ. Note, though, that the $v_a(\Gamma)$'s do not sum to $v(\Gamma)$, since the relation bundles that are inconsistent with the various constraints in Γ may overlap.

In the Appendix, we describe how the index $v(\Gamma)$ for a constraint set Γ defined on $\mathfrak{I}_p$ is related to the index for a partial algebra derived from Γ.

6. RANDOM GRAPHS AND MULTIGRAPHS: DISTRIBUTIONAL ASSUMPTIONS

To evaluate constraints in a situation where the relational data are assumed to be subject to random variation, we need to make some distributional assumptions. As Wasserman and Pattison (in press) have outlined, there are three major classes of multivariate random (directed) graph distributions in the literature. These are: (a) basic random graph distributions, including Bernoulli graphs and their generalizations (for example, Frank & Nowicki, 1993), and conditional uniform random graph distributions (Katz & Powell, 1957; Holland & Leinhardt, 1975; Wasserman, 1977; Snijders, 1991); (b) the dyad-independent p_1 model and its relatives (Anderson, Wasserman, & Faust, 1992; Fienberg & Wasserman, 1981; Fienberg, Meyer & Wasserman, 1985; Holland & Leinhardt, 1981; Holland, Laskey, & Leinhardt, 1983; Wang & Wong, 1987; Wasserman & Galaskiewicz, 1984); and (c) the p^* model, including Markov random graphs and their generalizations (Frank & Strauss, 1986; Pattison & Wasserman, 1999; Robins, Pattison, & Wasserman, 1999; Strauss, 1992; Strauss & Ikeda, 1990; Wasserman

& Pattison, 1996). We have chosen to work with the first class of distributions here and investigate the feasibility of the third class of models in future work.

We base our proposed procedures on conditional uniform random multigraph distributions. In order to describe these distributions, we assume that each multigraph **M** defined on N has a fixed set of r relation labels. The multigraph is regarded as a random variable, with realizations that are elements of some sample space of random (directed) multigraphs. We label this sample space Ω and we assume that every element of the space has the same probability of occurrence; that is, the distribution is *uniform.*

We allow the multigraphs in Ω to be restricted in some way; that is, they may be subject to some "marginal" constraints. For instance, they may all be assumed to have the same fixed densities, or the same outdegree vectors, or the same dyad censuses, and so on. The resulting distribution is uniform, conditional on a specific set of multigraph properties. We term such distributions *conditional uniform random* (*directed*) *multigraph distributions.* Univariate random graph distributions are described by Holland and Leinhardt (1975), Wasserman (1977); Wasserman and Pattison (in press) and Godehardt (1988) describe multigraph distributions. The purpose of working with conditional uniform random multigraph distributions is to be able to evaluate algebraic constraints in a way that is marginal to certain properties of a multigraph. We discuss some issues in the selection of useful properties in Section 7 after introducing the general evaluative framework.

We define $\boldsymbol{G}_d(g, r)$ to be the set of all multigraphs comprising r labeled directed graphs on the g nodes in N. The r relations are assumed to have the same set of relation labels for every multigraph. The set of (univariate) directed graphs on N corresponds to the set $\boldsymbol{G}_d(g, 1)$. We define a *property* $\boldsymbol{Q}$ of multivariate graphs to be any subset of $\boldsymbol{G}_d(g, r)$. For example, the property $\boldsymbol{Q}$ may correspond to the subset of multigraphs in $\boldsymbol{G}_d(g, r)$ that have a fixed total number of edges $L = M(+, +, +)$. In this case, we may describe $\boldsymbol{Q}$ as the property of having L edges (Bollobas, 1985; Wasserman & Pattison, in press). If we set the sample space Ω equal to $\boldsymbol{Q}$ and if we assume that every multigraph in Ω is equally likely, then we denote the resulting random multigraph distribution by $\mathbf{U} \mid \boldsymbol{Q}$: it is the uniform random multigraph distribution conditional on the property $\boldsymbol{Q}$.

The property $\boldsymbol{Q}$ may be simple or may take the form of the intersection of several more basic properties, $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$, ... In this latter case, $\boldsymbol{Q} = \boldsymbol{Q}_1 \cap \boldsymbol{Q}_1 \cap \cdots$ leads to a uniform distribution $\mathbf{U} \mid \boldsymbol{Q} = \mathbf{U} \mid \boldsymbol{Q}_1, \boldsymbol{Q}_2, ...$ that is conditional jointly on *each* of the listed properties. For instance, let $\boldsymbol{Q}_k$ be the subset of $\boldsymbol{G}_d(g, r)$ in which the number of ties in the kth relation $\mathbf{X}_k$ is constrained to be $M(k, +, +)$. Then the distribution $\mathbf{U} \mid \boldsymbol{Q}_1, \boldsymbol{Q}_2, ...,$ which may be written as $\mathbf{U} \mid \{M(1, +, +),\ M(2, +, +), ..., M(r, +, +)\} = \mathbf{U} \mid \{M(k, +, +)\}$, assigns equal probability to each multigraph having a fixed set of $M(k, +, +)$ edges in the kth relation.

In this last example, we can write the property $\boldsymbol{Q}$ as the intersection of properties, each of which depends only on *one* relation in the multigraph. In this special case, we may also write the sample space $\boldsymbol{Q}$ as the Cartesian product of sample spaces $Q^{[k]}$ defined for each relation: that is, $\boldsymbol{Q} = \boldsymbol{Q}^{[1]} \times \boldsymbol{Q}^{[2]} \times \cdots \times \boldsymbol{Q}^{[r]}$ (where $\boldsymbol{Q}^{[k]}$ is a subset of $\boldsymbol{G}_d(g, 1)$ referring to a graph with relation label k). We refer in this case to the properties $\boldsymbol{Q}^{[k]}$ as being *mutually independent*. More generally, a property

$\boldsymbol{Q}^{[k]}$ of the kth relation is *independent* in a property $\boldsymbol{Q}$ if $\boldsymbol{Q}$ can be written as a Cartesian product of the form $\boldsymbol{Q}^{[k]} \times \boldsymbol{Q}^{I \setminus \{k\}}$, where $\boldsymbol{Q}^{I \setminus \{k\}}$ is a subset of $\boldsymbol{G}_d(g, r-1)$ and refers to multigraphs with relation labels $\{1, 2, ..., k-1, k+1, ..., r\}$. A property $\boldsymbol{Q} = \boldsymbol{Q}^{[k_1, k_2, ..., k_r]}$ is *jointly dependent* on $\{k_1, k_2, ..., k_s\}$ if it cannot be written as a nontrivial product of independent properties.

Below we define properties of multigraphs on one or more relations that are useful for statistical evaluations of algebraic constraints.

Properties of a Single Relation

1. *An association table* $\mathbf{n}_{\mathbf{KX}}$. A quite general property of a relation with adjacency matrix $\boldsymbol{X}$ is based on the *association* between the relation and a fixed discrete-valued matrix $\boldsymbol{K}$. Some examples of possible matrices $\boldsymbol{K}$ are presented in Table 2 together with a hypothetical realization of the relation $\mathbf{X}$. The association tables $\mathbf{n}_{\mathbf{KX}}$ for the matrices in Table 2 are presented in Table 3.

The uniform distribution $\mathrm{U} \mid \mathbf{n}_{\mathbf{KX}}$ is conditional on the number of ties in the graph in positions corresponding to each possible value q of the matrix $\mathbf{K}$. This form of conditioning generalizes many cases of interest; for instance:

(a) If $K(i, j) = 1$ if $i \neq j$, and $K(i, j) = 0$ for $i = j$, there is just one possible value of q corresponding to possible ties of $\mathbf{X}$ (i.e., 1) so that $\mathbf{n}_{\mathbf{KX}}$ is a 1×2 table, with $n_{\mathbf{KX}}(1, 0)$ a count of the number of ties in $\mathbf{X}$ that are absent and $n_{\mathbf{KX}}(1, 1)$ the number that are present. We obtain the distribution $\mathrm{U} \mid X(+, +)$ (as for $\mathbf{K}_1$ in Tables 2 and 3). This random digraph distribution postulates that all graphs with a fixed number $X(+, +)$ of edges are equally likely to occur and was popularized by Erdős and Renyi (1960). There are $[g(g-1)]!/\{[g(g-1) - X(+, +)]!\ X(+, +)!\}$ elements in Ω.

(b) If $K(i, j) = i$ for all j, the entry $n_{\mathbf{KX}}(q, 1)$ of the association table $\mathbf{n}_{\mathbf{KX}}$ records the outdegree of node q. The distribution $\mathrm{U} \mid \{X(i, +)\}$ results (as for $\mathbf{K}_2$ in Tables 2 and 3). This distribution postulates that all directed graphs with the specified outdegrees $\{X(i, +)\}$ are equally likely.

(c) If φ is a mapping from N to a set $N_{\mathbf{B}}$ of classes, and if $K(i, j) = K(m, l)$ if and only if $\varphi(i) = \varphi(m)$ and $\varphi(j) = \varphi(l)$, then $\mathbf{K}$ is in *block form* and the distribution $\mathrm{U} \mid \mathbf{n}_{\mathbf{KX}}$ is conditional on the number of ties $X_{[p][q]}$ from each class p of $N_{\mathbf{B}}$ to each class q of $N_{\mathbf{B}}$ (see $\mathbf{K}_3$ in Tables 2 and 3). We refer to $\mathrm{U} \mid \mathbf{n}_{\mathbf{KX}}$ as the distribution conditional on the class-to-class tie frequency matrix $\mathbf{X}_{\varphi}$ (with entries $X_{[p][q]}$; see Fienberg & Wasserman, 1981).

(d) If $\mathbf{K}$ is a general binary relation, then the distribution $\mathrm{U} \mid \mathbf{n}_{\mathbf{KX}}$ is conditional on the number of $\mathbf{X}$ ties that match those in the relation $\mathbf{K}$ as well as the number of $\mathbf{X}$ ties that occur in the absence of $\mathbf{K}$ ties. The matrix of such a $\mathbf{K}$ is shown as $\mathbf{K}_4$ in Tables 2 and 3.

Other possible instances of $\mathbf{K}$ include matrices in block diagonal form that signify shared attributes (see also de Vries, 1993).

2. *The dyad census*. Denote by $M_{\mathbf{X}}$, $A_{\mathbf{X}}$, and $N_{\mathbf{X}}$, respectively, the number of mutual, asymmetric, and null dyads in the directed graph of a relation $\mathbf{X}$; the triple

TABLE 2

Valued Matrices K Corresponding to Some Conditional Uniform Random Graph Distributions Based on the Relation X

Label	Matrix **K**
X	0 1 1 0 0 0
	1 0 1 1 0 0
	1 1 0 1 0 0
	1 0 0 0 0 0
	0 1 0 0 0 1
	0 1 0 0 0 0
$\mathbf{K}_1$	0 1 1 1 1 1
	1 0 1 1 1 1
	1 1 0 1 1 1
	1 1 1 0 1 1
	1 1 1 1 0 1
	1 1 1 1 1 0
$\mathbf{K}_2$	0 1 1 1 1 1
	2 0 2 2 2 2
	3 3 0 3 3 3
	4 4 4 0 4 4
	5 5 5 5 0 5
	6 6 6 6 6 0
$\mathbf{K}_3$	0 1 1 2 2 2
	1 0 1 2 2 2
	1 1 0 2 2 2
	3 3 3 0 4 4
	3 3 3 4 0 4
	3 3 3 4 4 0
$\mathbf{K}_4$	0 2 2 2 2 2
	1 0 2 2 2 2
	1 1 0 2 2 2
	1 1 1 0 2 2
	1 1 1 2 0 2
	1 1 1 2 2 0

$(M_{\mathbf{X}}, A_{\mathbf{X}}, N_{\mathbf{X}})$ is termed the *dyad census* of the graph of **X** (for example, Holland & Leinhardt, 1975). If $\boldsymbol{Q}$ is the subset of directed graphs having the dyad census $(M_{\mathbf{X}}, A_{\mathbf{X}}, N_{\mathbf{X}})$ for fixed values $M_{\mathbf{X}}$, $A_{\mathbf{X}}$, and $N_{\mathbf{X}}$, then the distribution $\mathrm{U} \mid \boldsymbol{Q}$ assumes that all graphs with the specified dyad census are equally likely.

3. *Unlabeled directed graph structure*. Let **Q** comprise all relations whose unlabeled directed graphs are isomorphic to that of **X**, that is, to $A(\mathbf{X})$. We denote the uniform distribution that is conditional on this property **Q** by $\mathrm{U} \mid A(\mathbf{X})$. One additional useful restriction that may be placed on the graphs in the sample space is associated with a partition of the node set. Let $\varphi\colon N \rightarrow N_{\mathbf{B}}$ be a mapping from N

TABLE 3
Association Tables $\mathbf{n}_{\mathbf{KX}}$ for the Matrices K and X in Table 2

		$\mathbf{n}_{\mathrm{KX}}$	
K	q	$n_{\mathrm{KX}}(q, 0)$	$n_{\mathrm{KX}}(q, 1)$
K_1	1	18	12
K_2	1	3	2
	2	2	3
	3	2	3
	4	4	1
	5	3	2
	6	4	1
K_3	1	0	6
	2	7	2
	3	6	3
	4	5	1
K_4	1	6	6
	2	12	6

onto a set of classes $N_{\mathbf{B}}$ assumed, without loss of generality, to be labelled with the integers $\{0, 1, ..., C-1\}$. We can think of φ as a (*node-coloring*) function assigning one of C distinct values, or *colors*, to each node in N; we define $N_q = \{i \in N : \varphi(i) = q\}$. We denote by $A_\varphi(\mathbf{X})$ the *color-labeled* directed graph obtained from the relation $\mathbf{X}$, that is, the directed graph obtained when each node label i is replaced by the *color* $\varphi(i)$ of its class. Let $\mathbf{Q}$ comprise all relations whose color-labeled graphs are isomorphic. Then the distribution $\mathrm{U} \mid A_\varphi(\mathbf{X})$ is conditional on the color-dependent structural features of the relation $\mathbf{X}$ that are associated with some hypothesized partition φ.

Multigraph Properties

1. *Completely independent properties* $\mathbf{Q}^{[1]}, \mathbf{Q}^{[2]}, ..., \mathbf{Q}^{[r]}\}$. Suppose that $\mathbf{Q}^{[k]}$ is a property of the kth relation in a random multigraph and that the kth relation has a conditional uniform distribution $\mathrm{U} \mid \mathbf{Q}^{[k]}$. If we assume that the r relations in the multigraph are completely independent, then

$$\Pr(\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_r\} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_r\}) = \Pi_{k=1, 2, ..., r} \Pr(X_k = x_k),$$

where $\mathbf{x}_k$ is a realisation of the random relation $\mathbf{X}_k$. It follows that the multigraph distribution is uniform with sample space $\mathbf{Q}^{[1]} \times \mathbf{Q}^{[2]} \times \cdots \times \mathbf{Q}^{[r]}$. In this case each property $\mathbf{Q}^{[i]}$ is independent in $\mathbf{Q}$ and the multigraph distribution may be denoted by $\mathbf{U} \mid \mathbf{Q}^{[1]}, \mathbf{Q}^{[2]}, ..., \mathbf{Q}^{[r]}$. In a number of examples presented in Section 8 of the paper, completely independent properties are assumed, with the $\mathbf{Q}^{[k]}$ taking the form of one or more of the properties for a single relation listed above.

2. *Multivariate association properties.* More generally, we consider properties that are conditional on some form of joint association between relations. Three special cases of such distributions are:

(a) *The multigraph distribution* $\mathbf{U} \mid \mathbf{n}_{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s}$ *for an s-way association table.* One important case of a multivariate association distribution is based on the association table for a set of relations. Here the property $\mathbf{Q}^{[1, 2, ..., s]}$ specifies the set of multigraphs having the same table of cross-classified ties. We denote this distribution by $\mathbf{U} \mid \mathbf{n}_{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s}$ and refer to it as the distribution *conditional on the association table* for relations indexed by 1, 2, ..., *s*.

(b) *The multigraph distribution* $\mathbf{U} \mid A(\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\})$. A second type of joint property of several relations is the multivariate generalization of the property $A(\mathbf{X})$. As before, we let $A(\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\})$ denote the multigraph of the relations $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\}$ from which the node labels have been removed. We define $\mathbf{U} \mid A(\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\})$ to be the conditional uniform random multigraph distribution defined by the subset of multigraphs of $\mathbf{G}_d(g, r)$ whose unlabeled multigraphs are isomorphic to $A(\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\})$ (note that only the *node* labels are removed from the multigraph and not the relation labels). We refer to the distribution as being conditional on the *unlabeled structure* of the multigraph $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\}$. The members of this distribution have the same algebra generated by the multigraph $\{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_s\}$—this algebra is a subalgebra of that of the entire multigraph $\mathbf{M}$.

(c) *The distribution* $\mathbf{U} \mid \{M(+, i, j)\}$. Finally, we define a multigraph distribution that is conditional on one of the 2-way margins of the 3-way $\mathbf{M}$ array. The distribution is conditional on the total number of relational ties, across all *r* relations, from actor *i* to actor *j*.

Many of the statistical approaches to evaluating equality and inclusion constraints among relations reviewed earlier may be seen as relying on the assumption of some conditional uniform multigraph distribution $\mathbf{U} \mid \mathbf{Q}$ in which $\mathbf{Q} = \mathbf{Q}^{[1]} \cap \mathbf{Q}^{[2]} \cap \cdots \cap \mathbf{Q}^{[r]}$ comprises a set of independent properties of each relation. For example, Katz and Powell (1953) based their evaluation of the conformity of two relations on the distribution $\mathbf{U} \mid \{M(1, +, +), M(2, +, +)\}$, Hubert and Baker (1978) used the distribution $\mathbf{U} \mid A(\mathbf{X}_1), A(\mathbf{X}_2)$, and Holland and Leinhardt's (1975) approach to evaluating the transitivity constraint for a relation was to use the distribution $\mathrm{U} \mid (M_{\mathbf{X}}, A_{\mathbf{X}}, N_{\mathbf{X}})$ conditional on the dyad census. Below, we present a framework for evaluating constraints that includes each of these approaches as a special case, but in which distributions conditional on multivariate association properties are argued to lead to a valuable extension of earlier methods.

7. EVALUATING THE CONSTRAINTS

We now consider the general problem of how to evaluate the conformity between an observed multigraph and a constraint set. Our approach is to evaluate the constraint set with respect to one or more conditional uniform random multigraph distributions. That is, we evaluate a constraint set against the assumption that the multigraph under study is random and governed by a particular conditional uniform distribution. As we discuss below, there is rarely only a single distribution $\mathbf{U} \mid \mathbf{Q}$ that is appropriate to the evaluation and we instead conduct the evaluation using several such distributions.

For a particular $\mathbf{U} \mid \mathbf{Q}$, we consider the distribution of the index $v(\Gamma)$ that is generated by this distributional hypothesis. We compute a p-value for a particular constraint set Γ, assuming the distribution $\mathbf{U} \mid \mathbf{Q}$, as a tail area of the distribution of the statistic $v(\Gamma)$. We interpret an extreme (small or large) observed value of $v(\Gamma)$ in this distribution as evidence that the (high or low) degree of conformity of the observed relations to the constraint set Γ is very unlikely *if the underlying stochastic mechanism is* $\mathbf{U}/\mathbf{Q}$.

We stress that, in general, this procedure is not statistical inference in the usual sense.[3] In particular, we do not claim that an extreme, small value of $v(\Gamma)$ assuming some distribution $\mathbf{U} \mid \mathbf{Q}$ provides a simple endorsement of the constraint set Γ, since we can never be sure that the underlying stochastic mechanism is indeed $\mathbf{U} \mid \mathbf{Q}$. Rather, we view the computation of p-values signifying the extremity of $v(\Gamma)$ assuming various distributions $\mathbf{U} \mid \mathbf{Q}$ as "thought-experiments," providing useful evidence on which a deeper understanding of the structural features of observed relations might be based (as we attempt to illustrate in Section 8). In particular, we see this type of evidence as informing the development of more precise parametric models of multigraph structure (e.g., those in the family of models described by Pattison & Wasserman, 1999).

Of course, it may be possible to justify a more confirmatory approach in the case where an *a priori* choice of a specific $\mathbf{Q}$ may be made. Indeed, this was the approach used by Holland and Leinhardt (1975) and Hubert and Baker (1978). This general strategy has a long history in statistics, as well as in social network analysis (Wasserman & Faust, 1994, chapter 16). Hubert and Baker's technique has often been used in social network analysis (e.g., Hubert, 1987; Arabie & Hubert, 1992; Krackhardt, 1988).

Permutation Distributions

For each distribution $\mathbf{U} \mid \mathbf{Q}$, we assume that the observed multigraph is an element of the sample space $\mathbf{Q}$. Since the elements of the sample space are assumed to be equally likely, the distribution $\mathbf{U} \mid \mathbf{Q}$ can be simulated by picking elements from $\mathbf{Q}$ at random with equal probabilities. For many of the properties that we have described, the simulation can be achieved by randomly permuting some "constituents" of the multigraph (the constituents may be ties, dyads, nodes, or other higher-order structures; for further details, see Wasserman & Pattison, in press). For example, the distribution $\mathbf{U} \mid M(+, +, +)$ may be simulated by randomly permuting the entries $\{M(i, j, k)\}$ of the three-way array $\mathbf{M}$ and the distribution $\mathbf{U} \mid A(\mathbf{X}_1), A(\mathbf{X}_2), ..., A(\mathbf{X}_r)$ may be simulated by making independent random permutations of the nodes of each relation. This correspondence between the distribution of the statistic $v(\Gamma)$ and some permutation distribution invites inter-

[3] We concur with many of the criticisms that have been made of the interpretation of p-values, agreeing particularly with Good (1987) who suggests that a small value of p should be interpreted "not as a good reason for rejecting H_0 but as a reason for obtaining more evidence."

pretation of the proposed procedure as a *permutation test* (e.g., Good, 1994). The *exchangeability*[4] of observations on which the exactness and unbiasedness of permutation tests rely (Lehmann, 1986) is clearly satisfied if the assumption of $\mathbf{U} \mid \mathbf{Q}$ is justified.

If we are prepared to assume a particular $\mathbf{U} \mid \mathbf{Q}$, then we are likely to take the more confirmatory stance characterized above and the procedure can indeed be interpreted as a permutation test in the usual sense. If we are not prepared to commit to the assumption of a particular $\mathbf{U} \mid \mathbf{Q}$, though, we claim, as does Hubert (1987, p. iii), that our computations still have meaning *with respect to the relevant process of randomization* embodied in $\mathbf{U} \mid \mathbf{Q}$ and irrespective of how the data actually arose. Of course, randomization processes embodied in $\mathbf{Q}$ may take account of few (e.g., $\mathbf{U} \mid M(+, +, +)$) or many (e.g., $\mathbf{U} \mid A(\mathbf{X}_1), A(\mathbf{X}_2), ..., A(\mathbf{X}_r)$) hypotheses one might have of the manner in which the observations are actually generated. As we note below, the extent to which the most plausible hypotheses are captured in $\mathbf{U} \mid \mathbf{Q}$ affects the value of this referential meaning.

Choice of U/Q

It is clear, therefore, that an important aspect of the proposed framework for evaluating a constraint set Γ is the choice of the properties $\mathbf{Q}$ on which the evaluations are conditioned. A number of issues are relevant to the choice of properties.

First, as we have noted, many of the earlier approaches evaluated a particular constraint after controlling for some of the features of the observed relations (such as their density or aspects of their structure). Yet, the set of properties on which evaluations are conditioned varies markedly from one approach to another, and it is important to be clear about the rationale for the choice of relevant properties. We take the view here that an evaluation is generally conducted in the context of a broader argument about likely or possible structural issues. Thus, the set of relational features upon which a particular evaluation is conditioned is likely to depend on this broader context and, as we have observed, it is rarely possible, *a priori*, to specify such sets in advance. Of course, it is certainly important to accommodate any design constraints (such as limitations on the number of network ties emanating from a given node) or any well-supported hypothesis about the process that might underlie generation of the network data. At this stage, however, there are no strong claims for the latter (although we hope that they may emerge from the application of procedures such as we propose). For the present, the flexibility to evaluate constraints in a way that is marginal to any of a number of different relational features is a very useful aspect of the proposed framework.

Second, evaluation against a number of distributions may be desirable because it may not always be feasible to conduct evaluations that are marginal to a number of properties *simultaneously*. In order to condition simultaneously on a number of properties, the relevant subset of multigraphs is the intersection of the properties concerned and may be extremely small. For instance, the property that corresponds to a fixed unlabeled directed graph structure and a fixed set of outdegrees may

[4] Observations are *exchangeable* if the probability of any joint outcome is the same, regardless of the order in which the observations are considered (e.g., Lehmann, 1986).

define a subset of multigraphs of size 1. In such a case. simultaneous conditioning is clearly not possible. (Indeed, the size of the relevant subset can be regarded as an indicator of the feasibility of controlling simultaneously for a set of properties.)

Third, we note that comparisons among alternative candidate properties may be assisted by considering the natural partial order of properties induced by set-inclusion. For instance, the feature controlled in Katz and Powell's (1953) approach to evaluating conformity (namely, the number of ties in each relation) is a subset of those controlled in Hubert and Baker's (1978) approach (namely, the unlabeled directed graph structure of each relation). As a result, the latter provides a more compelling evaluation than the former over a wide range of situations (subject to consideration of the size of the sample space $\mathbf{Q}$).

Fourth, not all possible distributions are suitable for the evaluation of all constraint sets since there may, in special cases, be a relationship between a constraint set and the properties of some conditional uniform multigraph distribution. For instance, if Γ is the transitivity constraint set (see Table 1), then $v(\Gamma)$ is constant for all members of the distribution $\mathrm{U} \mid A(\mathbf{X})$. In other words, it is not useful to evaluate the transitivity constraint set using the distribution that conditions on all label-independent structural features of a graph.

Fifth, as we have noted, the use of multigraph properties exhibiting complete independence has been common for evaluating constraints involving two or more relations, and it is sufficiently flexible to control for a range of features of each relation. Nonetheless, the distribution associated with a completely independent set of properties generally serves the role of a null distribution for the evaluation of interdependencies among social relations—a distribution that is in many cases likely to be implausible. In the event that evidence renders such a null distribution unlikely, we may be interested in pursuing evaluations using more plausible random multigraph assumptions. We argue that there are at least two main types of random multigraph distributions that are likely to be of interest and that extend the value of the approach developed by Holland and Leinhardt (1975) and Hubert and Baker (1978) in a particularly useful way: distributions based on conditional independence and distributions based on uniform or partial ordering.

Conditional Independence Models

The first type of assumption specifies some form of conditional independence, such as the property that two relations **A** and **B** are independent, conditional on a third relation **C** (or set of relations). Two examples illustrate possible cases of interest. The first is illustrated by the assumption that two relations (e.g., help-seeking **A** and sharing leisure-time **B**) are conditionally independent, given a third (e.g., friendship **C**).[5] In this case, we might find it informative to evaluate the conformity of two relations with respect to a uniform random graph distribution that conditions on the association between each of the first two relations and the third (such as $\mathbf{U} \mid \mathbf{n}_{\mathbf{AC}}, \mathbf{n}_{\mathbf{BC}}$), that is, on the assumption that **A** and **B** are

[5] Such an assumption underlies the application of cultural consensus theory to networks (e.g., Romney & Weller, 1984; Romney, Weller & Batchelder, 1987) although, in this context, the third relation is assumed to be unobserved.

conditionally independent, given **C**. Our expectation would be that the statistic for the conformity constraint would not be extreme in this randomization distribution, although it would be extreme with respect to a distribution that does not condition on the association of each relation with **C**.[6]

The second example is that some algebraic constraint may be hypothesized to be associated with an *a priori* blockmodel, in the manner described in Section 3. That is, the constraints in Γ might be hypothesized to characterize constraints among relational ties *between* the classes of some hypothesized partition of the nodes. As a result, conformity to the constraints would be expected to be extreme with respect to a random multigraph distribution that is not marginal to aspects of the hypothesized partition, but would not be expected to be extreme with respect to a distribution for which association of the relational ties with the partition had been taken into account. To formalize these notions, we define several different stochastic mechanisms that are represented as properties of conditional uniform multigraph distributions.

Let the mapping φ: $N \to N_{\mathbf{B}}$ correspond to an hypothesized partition on the node set N, with $N_{\mathbf{B}}$ denoting the set of $g_{\mathbf{B}}$ classes of the partition. Define discrete-valued $g \times g$ matrices $\mathbf{K}$, $\mathbf{K}_o$, and $\mathbf{K}_i$, by

$$K(k, l) = (m-1)\, g_{\mathbf{B}} + h,$$

$$K_o(k, l) = (k-1)\, g_{\mathbf{B}} + h,$$

$$K_i(k, l) = (m-1)\, g + l,$$

where $\varphi(k) = m$ and $\varphi(l) = h$, and consider the properties $\mathbf{Q}_{\mathrm{SE}}$, $\mathbf{Q}_{\mathrm{AE}}$, $\mathbf{Q}_{\mathrm{OE}}$, $\mathbf{Q}_{\mathrm{IE}}$, and $\mathbf{Q}_{\mathrm{RE}}$ defined in Table 4. Each property is conditional on some feature associated with a proposed partition on N and can be regarded as defining a possible stochastic mechanism in association with the partition. For instance, $\mathbf{Q}_{\mathrm{SE}}$ specifies the collection of multigraphs having the same matrix of frequency of ties between and among classes of the partition (defined in Section 6) and so provides a plausible stochastic mechanism for an hypothesis of an *a priori* blockmodel, in which the probability of a tie between actors in two given classes depends only on the classes and not on the particular actors (see the definition of stochastic equivalence in Fienberg & Wasserman, 1981). Indeed, the implications of each property for tie probabilities are presented in Table 4 and indicate the way in which the properties may be seen as stochastic generalizations of the conditions governing the equivalence definitions given in Section 3.[7]

If Θ denotes one of the forms of equivalence (SE, AE, etc.) associated with a partition φ on the node set of a multigraph $\mathbf{M}$, and if $\mathbf{Q}_\Theta$ denotes the corresponding property defined in Table 4, then it is readily established that if φ induces a Θ-equivalence on $\mathbf{M}$, then φ induces a Θ-equivalence on $\mathbf{M}'$, for all $\mathbf{M}' \in \mathbf{U} \mid \mathbf{Q}_\Theta$. Further, in this case, the multigraph homomorphism induced by the partition φ is

[6] The simulation of the distribution is relatively straightforward; details are given by Wasserman and Pattison (in press).

[7] Note that the simulation of $\mathrm{U} \mid \mathbf{Q}_{\mathrm{RE}}$ is not straightforward, requiring the fixing of row and column sums within each block. The importance sampling approach described by Snijders (1991) may be used.

TABLE 4

Conditional Uniform Multigraph Distributions Associated with an Hypothesized Partition φ

Property label $\mathbf{Q}_{\Theta}$	Fixed marginal properties	Implications[a] (for all $l \in N$)
$\mathbf{Q}_{\mathrm{SE}}$	$\mathbf{n}_{\mathrm{KX}_1}, \mathbf{n}_{\mathrm{KX}_2}, ..., \mathbf{n}_{\mathrm{KX}_r}$	$\Pr(X_{\mathrm{k}}(i, l)=1)=\Pr(X_{\mathrm{k}}(j, l)=1)$ $\Pr(X_{\mathrm{k}}(l, i)=1)=\Pr(X_{\mathrm{k}}(l, j)=1)$
$\mathbf{Q}_{\mathrm{AE}}$	$A_{\varphi}(\mathbf{X}_1), A_{\varphi}(\mathbf{X}_2), ..., A_{\varphi}(\mathbf{X}_r)$	$\Pr(X_{\mathrm{k}}(i, l)=1)=\Pr(X_{\mathrm{k}}(j, \varphi(l))=1)$ $\Pr(X_{\mathrm{k}}(l, i)=1)=\Pr(X_{\mathrm{k}}(\varphi(l), j)=1)$
$\mathbf{Q}_{\mathrm{OE}}$	$\mathbf{n}_{\mathrm{K}_o\mathrm{X}_1, \mathrm{K}_o\mathrm{X}_2, ..., \mathrm{K}_o\mathrm{X}_r}$	$\Pr(X_{\mathrm{k}}(i, l)=1)=\Pr(X_{\mathrm{k}}(i, \varphi(l))=1)$
$\mathbf{Q}_{\mathrm{IE}}$	$\mathbf{n}_{\mathrm{K}_i\mathrm{X}_1, \mathrm{K}_i\mathrm{X}_2, ..., \mathrm{K}_i\mathrm{X}_r}$	$\Pr(X_{\mathrm{k}}(l, i)=1)=\Pr(X_{\mathrm{k}}(\varphi(l), i)=1)$
$\mathbf{Q}_{\mathrm{RE}}$	$\mathbf{n}_{\mathrm{K}_o\mathrm{X}_1, \mathrm{K}_o\mathrm{X}_2, ..., \mathrm{K}_o\mathrm{X}_r}$ $\mathbf{n}_{\mathrm{K}_i\mathrm{X}_1, \mathrm{K}_i\mathrm{X}_2, ..., \mathrm{K}_i\mathrm{X}_r}$	$\Pr(X_{\mathrm{k}}(i, l)=1)=\Pr(X_{\mathrm{k}}(i, \varphi(l))=1)$ $\Pr(X_{\mathrm{k}}(l, i)=1)=\Pr(X_{\mathrm{k}}(\varphi(l), i)=1)$

[a] It is assumed that $\varphi(j)=\varphi(i)$ under the hypothesized partition φ.

isomorphic for all $\mathbf{M}' \in \mathbf{U} \mid \mathbf{Q}_{\Theta}$. Thus, if some algebraic constraint Γ is exactly associated with an induced multigraph under an hypothesized partition φ, then the distribution of the statistic $v(\Gamma)$ will be degenerate. More generally, if the algebraic constraint Γ is associated with an hypothesized partition φ and if the stochastic mechanism is $\mathbf{U} \mid \mathbf{Q}_{\Theta}$, where $\mathbf{Q}_{\Theta}$ fixes some aspect of the multigraph induced by φ, then we would not expect the statistic $v(\Gamma)$ to be extreme in the randomization distribution associated with $\mathbf{U} \mid \mathbf{Q}_{\Theta}$.

As a specific example, consider the balance hypothesis of Table 1. It is well known (e.g., Cartwright & Harary, 1979) that the node set of a balanced multigraph may be partitioned into two classes, within which relational ties are exclusively positive and between which relational ties are exclusively negative. If we were to hypothesize that balance holds for a specified partition of the node set, then we would expect the statistic assessing the balance constraints to be small and extreme when evaluated with respect to a property that is insensitive to the hypothesized partition, but to be small and less extreme when evaluated with respect to a property that does respect the hypothesized partition.

Uniform and Partial Uniform Ordering

A second general form of the multivariate association model that may be helpful in assessing a set of algebraic constraints specifies that, for any pair i and j of nodes, ties of any type are equally likely to occur. This *uniform ordering* assumption implies that

$$\Pr(X_1(i, j)=1)=\Pr(X_2(i, j)=1)= \cdots =\Pr(X_r(i, j)=1) \qquad \text{for each } i, j \in N$$

and may be represented by the uniform random multigraph distribution $\mathbf{U} \mid \{M(+, i, j)\}$ that is conditional on the values $\{M(+, i, j)\}$. If the observed value of $v(\Gamma)$ for some constraint set Γ is small and extreme relative to the randomization distribution of $v(\Gamma)$, then there is evidence in favor of the hypothesized ordering constraints relative to the assumption of uniform ordering. Generalization of the uniform ordering hypothesis to properties that fix the number of ties from i to j for particular *subsets* of relations may also be useful.

To summarize, we suggest that it is often useful to conduct a series of evaluations of a particular constraint set Γ with respect to a variety of conditional uniform random multigraph distributions. We illustrate this approach below, using data from an empirical network study. We note, though, that in those (currently rare) contexts in which assumptions about underlying stochastic mechanisms can be strongly defended, a single evaluation of an hypothesized constraint set with respect to the corresponding distribution may be preferable because it permits a more standard permutation-test interpretation of the computed p-values.

8. APPLICATIONS: ORGANIZATIONAL STRUCTURE IN A BANK BRANCH

Our application is to data from a study of structure in a number of branches of a large Australian bank (Robins, 1994; Robins, Pattison, & Langan-Fox, 1997). The relations presented at the top of Table 5 are from one branch in response to questions: (1) With whom might you check out a course of action if an issue arises in your work? (the *advice-seeking* relation); (2) With whom do you feel that your work interactions are particularly satisfying? (the *satisfying interaction* relation); (3) In whom do you feel you would be able to confide if a problem arose that you did not want everyone to know about? (the *confiding* relation); (4) Whom do you consider to be a particularly close friend? (the *close friend* relation). We note that the first listed respondent is the Branch manager, the second is the deputy manager, the third, fourth, and fifth respondents are service advisers (a middle ranking position within the branch), and the remaining respondents are tellers.

1. *The transitivity of* Advice-Seeking *ties.* We first illustrate our approach by investigating the proposal that *advice-seeking* ties are transitive. We argue that individuals high in the organizational hierarchy are unlikely to seek advice from those at lower positions and that, since the workgroup is small, an individual is likely to seek advice from any individual at a higher level. This hypothesized pattern of advice-seeking implies transitivity in *advice-seeking* ties. We therefore choose $\Gamma_1 = \{(\mathbf{A}, \mathbf{AA})\}$ and examine $v(\Gamma_1)$ in relation to four conditional uniform random graph distributions: $\mathrm{U} \mid A(+, +)$, $\mathrm{U} \mid \{A(i, +)\}$, $\mathrm{U} \mid \mathbf{n}_{\mathrm{AK}}$, and $\mathrm{U} \mid \mathbf{n}_{\mathrm{AK}_o}$. The first two distributions are conditional, respectively, on the total number of *advice-seeking* ties, and on the number of *advice-seeking* ties expressed by each branch member. (Examination of the *advice-seeking* and other relations in Table 5 suggests that there are substantial individual differences in the expression of various kinds of tie, so that it is useful to conduct evaluations that are marginal to these variations.) The third and fourth distributions are conditional on a partition of group members according to organizational position. (The matrix $\mathbf{K}$ is shown in Table 5; the matrix $\mathbf{K}_o$ leads to a distribution that is conditional on the number of *advice-seeking* ties from each individual to those holding *each type* of formal

TABLE 5

Networks Relations in a Bank Branch

Label	Relation	Label	Relation
Advice – seeking	0 1 1 1 0 0 0 0 0 0 0	*Close*	0 1 1 1 1 0 0 0 0 0 0
A	1 0 0 0 0 0 0 0 0 0 0	*friendship*	0 0 0 0 0 0 0 0 0 0 0
	0 1 0 0 0 0 0 0 0 0 0	**F**	0 0 0 1 0 1 0 0 0 0 0
	1 0 0 0 0 0 0 0 0 0 0		0 0 1 0 0 1 0 0 0 0 1
	0 0 0 0 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0 0
	0 0 1 1 1 0 0 0 0 0 0		0 0 1 1 0 0 0 0 0 0 1
	1 1 1 1 1 1 0 0 0 0 0		0 0 1 1 0 1 0 0 1 0 0
	1 1 1 1 0 1 1 0 1 0 0		0 0 0 1 0 0 0 0 1 0 0
	1 1 1 0 0 0 0 0 0 0 0		0 0 0 0 0 0 0 1 0 0 0
	1 0 0 0 0 0 0 0 0 0 0		0 0 0 1 0 0 0 0 0 0 0
	0 0 1 1 0 1 1 0 0 0 0		0 0 0 0 0 0 0 0 0 0 0
Satisfying	0 0 0 0 0 0 0 0 0 0 0	*Confiding*	0 1 1 1 0 0 0 0 0 0 0
interaction	1 0 1 1 1 1 1 1 1 1 1	**C**	1 0 0 0 0 0 0 0 0 0 0
S	1 1 0 0 0 0 0 0 0 0 0		0 0 0 0 0 0 0 0 0 0 0
	1 1 1 0 1 1 1 1 1 1 1		1 0 0 0 0 0 0 0 0 0 0
	1 1 1 1 0 1 1 1 1 1 1		1 1 1 1 0 1 1 0 0 0 0
	0 0 1 1 0 0 0 0 0 0 0		0 0 1 1 0 0 0 0 0 0 0
	1 0 1 1 1 1 0 1 0 0 0		0 0 1 1 1 0 0 0 0 0 0
	0 0 1 1 1 0 1 0 1 0 1		1 0 0 1 0 0 0 0 1 0 0
	0 0 0 0 0 0 0 0 0 0 0		1 0 0 0 0 0 0 1 0 0 0
	0 0 0 1 0 0 0 0 0 0 0		1 0 0 1 0 0 0 0 0 0 0
	0 0 0 1 0 1 1 0 0 1 0		0 0 0 1 0 1 1 0 0 1 0
The formal	1 2 3 3 3 4 4 4 4 4 4		
structure	5 6 7 7 7 8 8 8 8 8 8		
K	9 10 11 11 11 12 12 12 12 12 12		
	9 10 11 11 11 12 12 12 12 12 12		
	9 10 11 11 11 12 12 12 12 12 12		
	13 14 15 15 15 16 16 16 16 16 16		
	13 14 15 15 15 16 16 16 16 16 16		
	13 14 15 15 15 16 16 16 16 16 16		
	13 14 15 15 15 16 16 16 16 16 16		
	13 14 15 15 15 16 16 16 16 16 16		
	13 14 15 15 15 16 16 16 16 16 16		

position in the branch. Note also that the third and fourth distributions are $U \mid \mathbf{Q}_{SE}$ and $U \mid \mathbf{Q}_{OE}$, respectively, for the partition according to formal position.)

The value of $v(\Gamma_1)$ is 18 and small in the first two distributions ($p = 0.001$ and 0.009, respectively). We interpret these p-values as evidence that the *advice-seeking* relation exhibits a tendency towards transitivity that is unlikely if ties are completely random (as in $U \mid A(+,+)$) or if individuals merely have varying propensities for expressing advice-seeking ties but select targets at random (as in $U \mid \{A(i,+)\}$). When we examine $v(\Gamma_1)$ with respect to distributions that are conditional on aspects of the partition according to formal positions, we find that $v(\Gamma_1)$ is no longer extreme ($p = 0.328$ and 0.197 for the third and fourth distributions, respectively). Thus, there is little evidence in favor of transitivity of *advice-seeking* ties beyond that which is associated with the formal positional structure, in agreement with expectations.

2. *The strength of* close friendship *ties relative to* advice-seeking *ties*. The second constraint set that we evaluate is for the hypothesis that *close friendship* ties are strong ties (in Granovetter's, 1973, sense) in relation to *advice-seeking* ties. We use a version of the constraints argued by Breiger and Pattison to characterize Granovetter's (1973) claims about the associations among strong and weak ties (Breiger & Pattison, 1978; also Pattison, 1993). In particular, we choose

$$\Gamma_2 = \{(\mathbf{A}, \mathbf{F}), (\mathbf{A}, \mathbf{AF}), (\mathbf{A}, \mathbf{FA}), (\mathbf{AF}, \mathbf{A}), (\mathbf{FA}, \mathbf{A}), (\mathbf{F}, \mathbf{FF})(\mathbf{FF}, \mathbf{F})\}$$

in which $\mathbf{A} \supseteq \mathbf{F}$, $\mathbf{AF} = \mathbf{FA} = \mathbf{A}$, and $\mathbf{FF} = \mathbf{F}$. In Table 6, we report evaluations of the constraint set with respect to several distributions $\mathbf{U} \mid \mathbf{Q}$. The table indicates the estimated proportion of multigraphs in each $\mathbf{U} \mid \mathbf{Q}$ for which the value of $v(\Gamma_2)$ is as large or larger than the observed value; in addition, the statistics $v_i(\Gamma_2)$ associated with each of the constituent constraints in Γ_2 are also presented.

The first three distributions have the property of complete multivariate independence: they are the distributions $\mathbf{U} \mid \{M(k, +, +)\}$, $\mathbf{U} \mid A(\mathbf{A}), A(\mathbf{F})$, and $\mathbf{U} \mid \{M(k, i, +)\}$. These distributions are conditional, respectively, on the total number $M(k, +, +)$ of ties in each relation k, on the unlabeled structure of each relation (namely, $A(\mathbf{A})$ and $A(\mathbf{F})$), and on the set of outdegrees of each node for each relation (that is, $\{M(k, i, +)\}$). It can be seen that the observed value of the statistic $v(\Gamma_2)$ is reasonably small for each distribution ($p = 0.001$, 0.085, 0.001, respectively), suggesting greater conformity to the constraint set than expected given each of these possible stochastic mechanisms. A similar result is observed for several multivariate association models. First, the observed value of $v(\Gamma_2)$ is small for the distribution $\mathbf{U} \mid \mathbf{n_{AK}}, \mathbf{n_{FK}}$ that is conditional on the association between the formal organizational structure $\mathbf{K}$ and each of the *close friendship* ($\mathbf{F}$) and *advice-seeking* ($\mathbf{A}$) relations ($p = 0.001$). It is also small for the distribution $\mathrm{U} \mid \mathbf{n}_{\mathbf{AK}_o\mathbf{FK}_o}$ ($p = 0.016$). As outlined in Section 6, this distribution is conditional on the outdegree of each type of tie from each node to each block, where blocks correspond here to formal organizational positions. Second, the observed value of $v(\Gamma_2)$ is also small for the distribution $\mathbf{U} \mid \{M(+, i, j)\}$ that is conditional on the number of $\mathbf{A}$ and $\mathbf{F}$ ties linking each pair of branch members ($p = 0.034$). Thus, it may be argued that there is some overall supporting evidence that $\mathbf{A}$ and $\mathbf{F}$ are related in the manner described by the constraint set Γ_2 and that this relationship does not appear to be attributable to constraints associated with the formal organizational structure.

The statistics for the evaluation of the individual constituents of the constraint set Γ_2 provide evidence in support of only *some* of the individual constraints of the strong-weak tie model, however, and so suggest that there might be sharper characterizations of the interdependence of the relations $\mathbf{A}$ and $\mathbf{F}$ than that reflected in the constraint set Γ_2. Thus, although $v_i(\Gamma_2)$ is small for each constituent in the multivariate independence distribution that is conditional only on the density of $\mathbf{A}$

TABLE 6

Summary of Statistics for Evaluation of the Strong–Weak Tie Constraint Set (Γ_2) for *Advice-Seeking* (A) and *Close Friendship* (F) Relations (Γ_2 = {(FA, A), (A, FA), (AF, A), (A, AF), (FF, F), (F, FF), (A, F)})

Property Q	Statistic	Observed value	p	Statistic	Observed value	p
$\{M(k, +, +)\}$	$v(\Gamma_2)$	65	0.001	$v_1(\Gamma_2)$	15	0.087
$A(\mathbf{F}), A(\mathbf{A})$			0.085			0.442
$\{M(k, i, +)\}$			0.001			0.081
$\mathbf{n}_{\mathbf{FK}}, \mathbf{n}_{\mathbf{AK}}$			0.001			0.078
$\mathbf{n}_{\mathbf{FK}_o}, \mathbf{n}_{\mathbf{AK}_o}$			0.016			0.305
$\{M(+, i, j)\}$			0.034			—[a]
$\{M(k, +, +)\}$	$v_2(\Gamma_2)$	19	0.001	$v_3(\Gamma_2)$	11	0.003
$A(\mathbf{F}), A(\mathbf{A})$			0.191			0.016
$\{M(k, i, +)\}$			0.008			0.018
$\mathbf{n}_{\mathbf{FK}}, \mathbf{n}_{\mathbf{AK}}$			0.252			0.013
$\mathbf{n}_{\mathbf{FK}_o}, \mathbf{n}_{\mathbf{AK}_o}$			0.628			0.042
$\{M(+, i, j)\}$			—[a]			—[a]
$\{M(k, +, +)\}$	$v_4(\Gamma_2)$	25	0.026	$v_5(\Gamma_2)$	7	0.001
$A(\mathbf{F}), A(\mathbf{A})$			0.462			—[b]
$\{M(k, i, +)\}$			0.223			0.002
$\mathbf{n}_{\mathbf{FK}}, \mathbf{n}_{\mathbf{AK}}$			0.066			0.001
$\mathbf{n}_{\mathbf{FK}_o}, \mathbf{n}_{\mathbf{AK}_o}$			0.075			0.005
$\{M(+, i, j)\}$			—[a]			—[a]
$\{M(k, +, +)\}$	$v_6(\Gamma_2)$	17	0.031	$v_7(\Gamma_2)$	10	0.018
$A(\mathbf{F}), A(\mathbf{A})$			—[b]			0.032
$\{M(k, i, +)\}$			0.094			0.018
$\mathbf{n}_{\mathbf{FK}}, \mathbf{n}_{\mathbf{AK}}$			0.118			0.069
$\mathbf{n}_{\mathbf{FK}_o}, \mathbf{n}_{\mathbf{AK}_o}$			0.260			0.171
$\{M(+, i, j)\}$			—[a]			0.050

[a] Only component constraints that refer directly to observed relations are evaluated using the uniform ordering distribution.

[b] The distribution that is conditional on unlabeled directed graph structure cannot be used for constraints involving a singe relation (see Section 7)

and **F** ties, the same pattern is not observed in a number of other distributions. In particular, the statistic $v_4(\Gamma_2)$ for the constraint $\mathbf{A} \supseteq \mathbf{AF}$ is not extreme in the distribution that is conditional on unlabeled graph structure nor in that conditional on the outdegrees of each member for each type of tie; similarly for the statistic $v_1(\Gamma_2)$ pertaining to the constraint $\mathbf{FA} \supseteq \mathbf{A}$. The constraints with small and extreme statistics in most distributions include $\mathbf{AF} \supseteq \mathbf{A}$ and $\mathbf{FF} \supseteq \mathbf{F}$; the findings suggest that a sharper characterization of the interdependence of **A** and **F** ties might be developed as follows.

Although *close friendship* ties do not exhibit transitivity as the "strong tie" hypothesis would suggest, they do display a redundant embedding in the network, in the sense that any *close friendship* tie linking a pair of individuals is likely to be accompanied by a walk comprising two *close friendship* ties between the pair (in agreement with $\mathbf{FF} \supseteq \mathbf{F}$). Similarly, an *advice-seeking* tie from one individual to another tends to be accompanied by at least one *advice-seeking* tie from the first individual to someone having a *close friendship* tie to the second (as suggested by $\mathbf{AF} \supseteq \mathbf{A}$). In other words, both types of tie tend not to occur in isolation, with *close friendship* ties providing the "glue" that tends to link at least two of the recipients of an individual's ties to one another. We note that this is a weaker characterization of structural interdependence than that embodied in the strong–weak tie hypothesis (wherein *close friendship* ties tend to be involved in transitive triads, that is, triads in which a tie from i to m and a tie from m to j are inevitably accompanied by a tie from i to j). Here, the claim is simply that these kinds of transitive triads (involving *some* actor m) exist for any pair of actors i and j who are linked by a close friendship tie, rather than that they are a *necessary* triadic form.

3. *The possible bases of* satisfying interactions. Under what conditions will a member of the organization describe work interactions with other members as satisfying? One likely condition is when there is a strong affective basis for the relationship, sufficiently strong for the respondent to use the label *close friend*; another is when there is enough trust on the part of the respondent to confide in the other member about work-related problems. Thus, a plausible set of constraints for the associations among relations of *close friendship*, *confiding*, and *satisfying interactions* is

$$\Gamma_3 = \{(\mathbf{S}, \mathbf{C})(\mathbf{S}, \mathbf{F})\}.$$

Statistics relevant to the evaluation of this constraint set are reported in Table 7. The observed value of the statistic $v(\Gamma_3)$ is quite small (12) and is extreme for each of the assumed conditional uniform random multigraph distributions. Thus, no matter whether we examined the distribution of the statistic assuming complete multivariate independence (with fixed density of each type of tie, fixed outdegree of each type of tie for each node, or fixed unlabeled structure of each type of tie) or some form of multivariate association (fixed relationship of each type of tie with the formal organizational structure, a fixed distribution of each type of tie from each node to each formal position, or a fixed number of ties between every pair of organization members), the p-value associated with the observed statistic is small ($p = 0.001$).

One might conclude, therefore, that evidence in favor of the constraints expressed in the set Γ_3 is strong. This conclusion is modified when one examines the statistics for the constituent constraints, $(\mathbf{S}, \mathbf{C})$ and $(\mathbf{S}, \mathbf{F})$: although support in favor of $\mathbf{S} \supseteq \mathbf{C}$ is strong, that for $\mathbf{S} \supseteq \mathbf{F}$ is weaker (indeed, it is strong only in relation to the multivariate association distributions). Thus, whereas there is a strong tendency for members of the branch to describe as satisfying the relational ties with those in whom they confide, a tendency for relational ties to close friends to be described in this way is discernible only when the evaluation is marginal to the formal organizational structure. In fact, a possible modification to the constraint is suggested by the observation that two-thirds of the observed violations to $\mathbf{S} \supseteq \mathbf{F}$ are associated with

TABLE 7

Summary of Statistics for Evaluation of the Constraint Set (Γ_2) for *Satisfying Working Relations* (S), *Confiding* (C), and *Close Friendship* (F) ($\Gamma_3 = \{(S, C), (S, f)\}$)

Property Q	Statistic	Observed value	p
$\{M(k, +, +)\}$	$v(\Gamma_3)$	12	0.001
$A(\mathbf{F}), A(\mathbf{S}), A(\mathbf{C})$			0.001
$\{M(k, i, +)\}$			0.001
$\mathbf{n_{FK}}, \mathbf{n_{SK}}, \mathbf{n_{CK}}$			0.001
$\mathbf{n}_{\mathrm{FK}_o}, \mathbf{n}_{\mathrm{SK}_o}\, \mathbf{n}_{\mathrm{CK}_o}$			0.001
$\{M(+, i, j)\}$			0.001
$\{M(k, +, +)\}$	$v_1(\Gamma_3)$	7	0.001
$A(\mathbf{F}), A(\mathbf{S}), A(\mathbf{C})$			0.003
$\{M(k, i, +)\}$			0.001
$\mathbf{n_{FK}}, \mathbf{n_{SK}}, \mathbf{n_{CK}}$			0.001
$\mathbf{n}_{\mathrm{FK}_o}, \mathbf{n}_{\mathrm{SK}_o}\, \mathbf{n}_{\mathrm{CK}_o}$			0.001
$\{M(+, i, j)\}$			0.001
$\{M(k, +, +)\}$	$v_2(\Gamma_3)$	9	0.162
$A(\mathbf{F}), A(\mathbf{S}), A(\mathbf{C})$			0.305
$\{M(k, i, +)\}$			0.187
$\mathbf{n_{FK}}, \mathbf{n_{SK}}, \mathbf{n_{CK}}$			0.028
$\mathbf{n}_{\mathrm{FK}_o}, \mathbf{n}_{\mathrm{SK}_o}\, \mathbf{n}_{\mathrm{CK}_o}$			0.037
$\{M(+, i, j)\}$			0.001

ties from branch members 1 and 3, neither of whom regards ties to those at lower levels in the branch (whether close friends or not) as satisfying.

4. Confiding *as local* advice-seeking. The final constraint set that we evaluate for the bank branch is one that pertains to a possible association between *advice-seeking* and *confiding* relations. We might expect that those individuals in whom one confides are a subset of those from whom one seeks advice; hence $\mathbf{A} \supseteq \mathbf{C}$ is a likely constraint. Further, since advice-seeking is likely to cross several hierarchical levels of the formal structure in a group as small as the branch under study, one might argue that *confiding* relations are likely to be more local or restricted in their span, linking individuals from one level in the organization to those in the next. Indeed, it is likely that one seeks advice from the confidants of one's confidants (possibly with the encouragement of the latter) and that such links may span more than two adjacent levels of the organization. Consequently, a second likely constraint is $\mathbf{A} \supseteq \mathbf{CC}$. We therefore evaluate the constraint set

$$\Gamma_4 = \{(\mathbf{A}, \mathbf{C}), (\mathbf{A}, \mathbf{CC})\}.$$

The outcome of these evaluations is summarized in Table 8. The observed value of $v(\Gamma_4)$ (and of both $v_1(\Gamma_4)$ and $v_2(\Gamma_4)$) is small and extreme in the conditional

TABLE 8

Summary of Statistics for Evaluation of the Constraints Set (Γ_3) for *Confiding* (C) and *Advice-Seeking* (A) Relations ($\Gamma_4 = \{(A, CC), (A, C)\}$)

Property **Q**	Statistic	Observed value	p
$\{M(k, +, +)\}$	$v(\Gamma_4)$	26	0.001
$A(\mathbf{C}), A(\mathbf{A})$			0.001
$\{M(k, i, +)\}$			0.001
$\mathbf{n}_{\mathbf{CK}}, \mathbf{n}_{\mathbf{AK}}$			0.001
$\mathbf{n}_{\mathrm{AK}_o}\mathbf{n}_{\mathrm{CK}_o}$			0.003
$\{M(+, i, j)\}$			0.285
$\{M(k, +, +)\}$	$v_1(\Gamma_4)$	22	0.002
$A(\mathbf{C}), A(\mathbf{A})$			0.029
$\{M(k, i, +)\}$			0.001
$\mathbf{n}_{\mathbf{CK}}, \mathbf{n}_{\mathbf{AK}}$			0.002
$\mathbf{n}_{\mathrm{AK}_o}\mathbf{n}_{\mathrm{CK}_o}$			0.008
$\{M(+, i, j)\}$			—[a]
$\{M(k, +, +)\}$	$v_2(\Gamma_4)$	9	0.001
$A(\mathbf{C}), A(\mathbf{A})$			0.001
$\{M(k, i, +)\}$			0.001
$\mathbf{n}_{\mathbf{CK}}, \mathbf{n}_{\mathbf{AK}}$			0.001
$\mathbf{n}_{\mathrm{AK}_o}\mathbf{n}_{\mathrm{CK}_o}$			0.031
$\{M(+, i, j)\}$			0.344

[a] Only component constraints that referred directly to observed relations were evaluated using the uniform ordering distribution.

uniform multigraph distributions $U \mid \{M(k, +, +)\}$, $U \mid \{M(k, i, +)\}$, and $U \mid A(\mathbf{A})$, $A(\mathbf{C})$ and in the multivariate association models $U \mid \mathbf{n}_{\mathbf{AK}}, \mathbf{n}_{\mathbf{CK}}$ and $U \mid \mathbf{n}_{\mathrm{AK}_o\mathrm{CK}_o}$. Thus, with respect to these distributions, evidence in favor of the constraint set appears strong. The only distribution for which $v(\Gamma_4)$ is not extreme is $U \mid \{M(+, i, j)\}$. As the relations in Table 5 show, two thirds of the violations to the constraint $\mathbf{A} \supseteq \mathbf{C}$ are associated with the idiosyncratic confiding pattern of a single individual (branch member 5) but suggest, nonetheless, the need for some qualification to the general picture of support for Γ_4.

9. PROSPECTS

In conclusion, we argue that the framework that we have described is useful for evaluating quite complex hypotheses about forms of relational association. In this final section of the paper, we outline some developments of the approach that we believe will be worthwhile.

Exploratory Methods

In the event that an hypothesized set of constraints appears not to be supported by the evaluations, how can we generate alternative possibilities? Some potentially useful exploratory procedures for generating plausible candidate constraint sets include the following. First, separate evaluations of constituent constraints of the kind that we have conducted are clearly helpful. These evaluations allow us to identify whether there is evidence for each individual component of the hypothesized constraint set. Second, we can examine the distribution of components of the vector statistic $v(\Gamma)$ in the assumed multigraph distribution. Recall that the components of $v(\Gamma)$ are the sums of relation bundles that adhere to the constraints specified by Γ and that the indicator matrix $\boldsymbol{I}$ indicates which relation bundles are not consistent with Γ. As a result, a distribution of violations may be constructed, and those cells of the association table in which violations in the observed relations are, and are not, extreme in the multigraph distribution may be determined. This information may be used to compile more likely sets of constraints for the observed relational data, although it is not necessarily straightforward to generate a best candidate from this information. Indeed, this approach could conceivably be used to generate likely constraints without making any initial assumptions about their likely form.

Finally, another general exploratory technique that may prove useful is to evaluate all possible constraints comprising a single ordered pair of relations, where the pairs are constructed systematically from a list of relations comprising not only the observed relations, but also the converses and complements of the observed relations and compound relations constructed from all of these: this strategy is similar to that outlined in Pattison and Wasserman (1995).

Further Developments

There are three further developments of our approach that we believe may be useful from both theoretical and practical points of view. The first is a means of exploring variations in relational constraints in different regions of a multigraph. Although we have described the means for evaluating constraints given some hypothesis about the association of a set of constraints with a partition of the node set, it would be useful to develop the means for generating such hypotheses directly. Second, in some network studies, valued relational data are gathered (although see Haslam, 1994) and it would be valuable to develop analogues of the methods described here for the valued case. Pattison (1993) outlines possible procedures for generalizing algebraic constraints to the valued case using a form of composition for valued data (e.g., the so-called *max-min* rule); generalizations of the statistic $v(\Gamma)$ may also be readily generated. Several means of simulating valued multigraph distributions that are analogous to the ones we have described may also be outlined, although there are several practical limitations to such methods that require further exploration. Third, the exploration of alternative distributional assumptions for networks seems worthwhile, particularly models within the p^* class (e.g., Frank & Strauss, 1986; Pattison & Wasserman, 1999; Strauss & Ikeda, 1990).

APPENDIX: ALGEBRAS AND CONSTRAINT SETS

In order to describe the relationship between constraint sets and algebraic models, we define the closure of a constraint set. A constraint set Γ is *p-closed* if and only if, for all $\mathbf{Y}$, $\mathbf{Z} \in \Im_p$,

(a) $(\mathbf{Y}, \mathbf{Z}) \in \Gamma$ implies $(\mathbf{YU}, \mathbf{ZU}) \in \Gamma$, for any relation $\mathbf{U}$ such that $\mathbf{YU}$, $\mathbf{ZU} \in \Im_p$;

(b) $(\mathbf{Y}, \mathbf{Z}) \in \Gamma$ implies $(\mathbf{UY}, \mathbf{UZ}) \in \Gamma$, for any relation $\mathbf{U}$ such that $\mathbf{UY}$, $\mathbf{UZ} \in \Im_p$; and

(c) $(\mathbf{Y}, \mathbf{Z}) \in \Gamma$ and $(\mathbf{Z}, \mathbf{W}) \in \Gamma$ implies $(\mathbf{Y}, \mathbf{W}) \in \Gamma$, for any $\mathbf{Z} \in \Im_p$.

The *p-closure* of a constraint set Γ is the least constraint set Γ' such that Γ' is p-closed and $\Gamma \subseteq \Gamma'$. If a constraint set Γ is p-closed for all finite values of p then it is said simply to be *closed*; further, the *closure* of Γ is the least constraint set containing Γ that is closed. The *transitive p-closure* of Γ is the least constraint set $\Gamma' \supseteq \Gamma$ that satisfies property (c), while the *isotonic p-closure* is the least constraint set $\Gamma' \supseteq \Gamma$ satisfying both (a) and (b).

It follows directly from the definition that any semigroup or partially ordered semigroup constructed from a network defines a closed constraint set. Further, any partial partially ordered semigroup on $\Im_p$ gives rise to a p-closed constraint set (with $(\mathbf{Y}, \mathbf{Z})$ in the constraint set iff $\mathbf{Y} \geqslant \mathbf{Z}$ in the partial algebra); in addition, the p-closure of any given constraint set corresponds to a partial partially ordered semigroup on $\Im_p$. Pattison and Wasserman (1995) used similar closure constructions to generate algebraic models from exploratory statistical investigations of a set of relations.

It also follows from the definition above that if Γ_1 and Γ_2 are p-closed (closed) constraint sets, then $\Gamma' = \Gamma_1 \cap \Gamma_2$ is p-closed (closed). In addition, the constraint set $\mathbf{U}(\mathbf{U}_p)$ comprising all ordered pairs on $\Im(\Im_p)$ is closed (p-closed). As a result, p-closure (closure) is a *closure property* in the sense of Birkhoff (1967); that is, it is a property of the subsets of $\mathbf{U}(\mathbf{U}_p)$ such that (i) $\mathbf{U}(\mathbf{U}_p)$ has the property, and (ii) any intersection of subsets has the property. Since the subsets of any set which have a given closure property form a lattice (Birkhoff, 1967), it follows that the collection of constraints that are p-closed (closed) forms a lattice. The greatest lower bound of two p-closed (closed) constraint sets Γ_1 and Γ_2 is their intersection $\Gamma_1 \cap \Gamma_2$; the least upper bound of Γ_1 and Γ_2 is the least p-closed (closed) constraint set that contains both Γ_1 and Γ_2.

In some cases, there is a relationship between the index $v(\Gamma)$ for a constraint set Γ defined on $\Im_p$ and the index $v(\Gamma')$ of the p-closure Γ' of Γ. It is straightforward to establish that if Γ' is the transitive closure of Γ (i.e., $\Gamma' \supseteq \Gamma$ and Γ' differs from Γ only in having constraints of the form $(\mathbf{Y}, \mathbf{W})$, where $(\mathbf{Y}, \mathbf{Z}) \in \Gamma$ and $(\mathbf{Z}, \mathbf{W}) \in \Gamma$), then $v(\Gamma') = v(\Gamma)$. The same is not true in general, however, for constraints in the isotonic p-closure of Γ. Thus, even though $v(\Gamma) = 0$ implies that $v(\Gamma') = 0$, where Γ' is the p-closure of Γ, a large value of $v(\Gamma')$ can be associated with a small but nonzero value of $v(\Gamma)$. Indeed, adding (or deleting) a single network link can lead to a substantial change in the value of the statistic and some applications may require the construction of more robust measures (for example, see Schwartz & Sprinzen, 1984, for a strategy for constructing such measures). Our proposed framework for evaluation can be adapted readily to any proposed measure.

REFERENCES

Anderson, C., Wasserman, S., & Faust, K. (1992). Building stochastic blockmodels. *Social Networks*, **14**, 137–161.

Arabie, P., & Hubert, L. J. (1992). Combinatorial data analysis. *Annual Review of Psychology*, **43**, 169–203.

Birkhoff, G. (1967). *Lattice theory* (3rd ed.). Providence, RI: American Mathematical Society.

Bollobas, B. (1985). *Random graphs*. London: Academic Press.

Bonacich, P. (1983). Representations for homomorphisms. *Social Networks*, **5**, 173–192.

Boorman, S., & White, H. C. (1976). Social structures from multiple networks: II. Role structures. *American Journal of Sociology*, **81**, 1384–1446.

Boyd, J. P. (1969). The algebra of group kinship. *Journal of Mathematical Psychology*, **6**, 139–167.

Boyd, J. P. (1990). *Social semigroups: A unified theory of scaling and blockmodelling as applied to social networks*. Fairfax, VA: George Mason Univ. Press.

Breiger, R. L., Boorman, S., & Arabie, P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparisons with multi-dimensional scaling. *Journal of Mathematical Psychology*, **12**, 328–382.

Breiger, R. L., & Pattison, P. E. (1978). The joint role structure in two communities' elites. *Sociological Methods and Research*, **7**, 213–226.

van de Bunt, G., van Duijn, M., & Snijders, T. (1995). Friendship networks and rational choice. In M. G. Everett & K. Rennolls (Eds.), *Proceedings of the International Conference on Social Networks, Volume I, 6th–10th July, 1995*. London, UK: Univ. Greenwich Press.

Cartwright, D., & Harary, F. (1979). Balance and clusterability: An overview. In P. W. Holland, & S. Leinhardt (Eds.), *Perspectives on social network research* (pp. 25–50). New York: Academic Press.

Erdős, P., & Renyi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 17–61.

Fienberg, S. E., Meyer, M. M., & Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, **80**, 51–67.

Fienberg, S., & Wasserman, S. (1981). Categorical data analysis of single sociometric relations. In S. Leinhardt (Ed.), *Sociological methodology 1981* (pp. 156–192). San Francisco: Jossey-Bass.

Frank, O., & Nowicki, K. (1993). Exploratory statistical analysis of networks. In J. Gimbel, J. W. Kennedy & L. V. Quintas (Eds.), *Quo vadis graph theory? A source book for challenges and directions*. Amsterdam: North-Holland. Also *Annals of Discrete Mathematics*, **55**, 349–366.

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

Godehardt, E. (1988). *Graphs as structural models: The application of graphs and multigraphs in cluster analysis*. Braunschweig/Wiesbaden: Vieweg.

Good, I. J. (1987). Comment [on J. O. Berger & T. Selke (1987). Testing a point null hypothesis: the irreconcilability of *p* values and evidence.]. *Journal of the American Statistical Association*, **82**, 125–128.

Good, P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer-Verlag.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, **78**, 1360–1380.

Haslam, N. (1994). Categories of social relationship. *Cognition*, **53**, 59–90.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: some first steps. *Social Networks*, **5**, 109–137.

Holland, P. W., & Leinhardt, S. (1975). The statistical analysis of local structure in social networks. In D. R. Heise (Ed.), *Sociological methodology*, 1976, pp. 1–45. San Francisco: Jossey-Bass.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, **76**, 33–50.

Hubert, L. J. (1987). *Assignment methods in combinatorial data analysis*. New York: Dekker.

Hubert, L. J., & Baker, F. B. (1978). Evaluating the conformity of sociometric measurements. *Psychometrika*, **43**, 31–41.

Katz, L., & Powell, J. H. (1953). A proposed index of conformity of one sociometric measurement to another. *Psychometrika*, **18**, 249–256.

Katz, L., & Powell, J. H. (1957). Probability distributions of random variables associated with a structure of the sample space of sociometric investigations. *Annals of Mathematical Statistics*, **28**, 442–448.

Kim, K. H., & Roush, F. W. (1984). Group relationships and homomorphisms of Boolean matrix semigroups. *Journal of Mathematical Psychology*, **28**, 448–452.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analyses of dyadic data. *Social Networks*, **10**, 359–381.

Lehmann, E. (1986). *Testing statistical hypotheses*. New York: Wiley.

Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, **1**, 49–80.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Pattison, P. E. (1982). The analysis of semigroups of multirelational systems. *Journal of Mathematical Psychology*, **25**, 87–118.

Pattison, P. E. (1993). *Algebraic models for social networks*. New York: Cambridge Univ. Press.

Pattison, P., & Wasserman, S. (1995). Constructing algebraic models for local social networks using statistical methods. *Journal of Mathematical Psychology*, **39**, 57–72.

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

Robins, G. (1994). *Group effectiveness: A network analysis of group structure and organizational performance*. Unpublished B. A. Hons Thesis, Department of Psychology, University of Melbourne.

Robins, G., Pattison, P. E., & Langan-Fox, J. (1997). Group effectiveness: A comparative network analysis of interactional structure and group performance in organizational workgroups. Unpublished manuscript. Department of Psychology, University of Melbourne.

Robins, G., Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks, III. Valued relations. *Psychometrika*, **64**, 371–394.

Romney, A. K., & Weller, S. C. (1984). Predicting informant accuracy from patterns of recall among individuals. *Social Networks*, **6**, 59–77.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, **88**, 313–338.

Schwartz, J. E., & Sprinzen, M. (1984). Structures of connectivity. *Social Networks*, **6**, 103–140.

Snijders, T. A. B. (1991). Enumeration and simulation methods for 0-1 matrices with given marginals. *Psychometrika*, **56**, 397–417.

Snijders, T. A. B., & Stokman, F. N. (1987). Extensions of triad counts to networks with different subsets of points and testing underlying random graph distributions. *Social Networks*, **9**, 249–275.

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.

de Vries, H. (1993). The rowwise correlation between two proximity matrices and the partial rowwise correlation. *Psychometrika*, **58**, 53–69.

Wang, Y. Y., & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, **82**, 8–19.

Wasserman, S. (1977). Random directed graph distributions and the triad census in social networks. *Journal of Mathematical Sociology*, **5**, 61–86.

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

Wasserman, S., & Anderson, C. (1987). Stochastic *a posteriori* blockmodels: construction and assessment. *Social Networks*, **9**, 1–36.

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

Wasserman, S., & Galaskiewicz, J. (1984). Some generalizations of p_1: External constraints, interactions, and non-binary relations. *Social Networks*, **6**, 177–192.

Wasserman, S., & Pattison, P. E. in press. *Multivariate random graph distributions*. Lecture Note Series in Statistics. Berlin/New York: Springer-Verlag.

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

White, D. (1984). Material entailment analysis. *Practicing Anthropologist*, **6**, 10–11.

White, D. (1996). Statistical entailments and the Galois lattice. *Social Networks*, **18**, 20–215.

White, D., & McCann, H. G. (1988). Cites and fights: Material entailment analysis of the eighteenth-century chemical revolution. In B. Wellman & S. Berkowitz (Eds.), *Social structures*: *A network approach* (pp. 380–400). New York: Cambridge Univ. Press.

White, D., & Reitz, K. (1983). Graph and semigroup homomorphisms on networks of relations. *Social Networks*, **5**, 193–234.

White, H. C. (1963). *An anatomy of kinship*: *Mathematical models for structures of cumulated roles*. Englewood Cliffs, NJ: Prentice–Hall.

White, H. C., Boorman, S., & Breiger, R. L. (1976). Social structure from multiple networks, I. Blockmodels of roles and positions. *American Journal of Sociology*, **81**, 730–780.

Received August 17, 1995; revised November 17, 1997