An Exponential Family of Probability Distributions for Directed Graphs

PAUL W. HOLLAND and SAMUEL LEINHARDT*

Directed graph (or digraph) data arise in many fields, especially in contemporary research on structures of social relationships. We describe an exponential family of distributions that can be used for analyzing such data. A substantive rationale for the general model is presented, and several special cases are discussed along with some possible substantive interpretations. A computational algorithm based on iterative scaling procedures for use in fitting data is described, as are the results of a pilot simulation study. An example using previously reported empirical data is worked out in detail. An extension to multiple relationship data is discussed briefly.

KEY WORDS: Random digraphs; Networks; Sociometry; Generalized iterative scaling.

1. INTRODUCTION

A directed graph or "digraph" (Harary, Norman, and Cartwright 1965) is specified by a (finite) set of points, or nodes, which we shall index by 1, 2, . . . , g, (g = total number of nodes) and a set of directed lines, or edges, that connect certain pairs of these nodes. We assume that there are no edges that connect a node to itself and that there is at most one edge connecting any two distinct nodes in a given direction. Figure 1 illustrates a digraph with five nodes ($g = 5$) and nine directed edges.

Directed graphs arise in many fields, but the applications that motivate our work are studies of social networks in anthropology, sociology, social psychology, and related disciplines (Leinhardt 1977; Holland and Leinhardt 1979b). In these applications, the nodes usually represent people, and the directed edges represent directed relationships that can obtain between these people. For example, some of the earliest quantitative research on social networks was done by Moreno (1934), who called his studies of the friendship patterns obtaining between group members "sociometric" studies. In this case there is a directed edge from node i to node j if individual i says that individual j is a friend. If we interpret Figure 1 as the digraph of friendship in a group of five people, then person 1 says that persons 2 and 5 are his or her friends, while person 2 says that person 3 is a friend, and so on.

The sociometric studies of Moreno have been generalized in a variety of ways; we use the term *sociometric* to refer to any study of the structure of social relationships, regardless whether the nodes represent people or other social actors such as corporations, government agencies, and other institutional entities. The many different kinds of scientific questions that are of interest in sociometric studies range from identifying patterns of regularities among the friendship choices in the original Moreno studies to relating communication patterns to the output of work groups. The key element of such studies is their focus on the *pattern of relationships* between the actors rather than on the *distribution of attributes* possessed by the actors. Sociometric studies have become quite common in the sociological, social psychological, anthropological, and educational literatures. Examples of the substantive concerns of recent sociometric studies include political, economic, and social elites (Moore 1979; Laumann and Pappi 1976; Alba and Moore 1978), scientific elites (Breiger 1976; Friedkin 1978; Burt 1980), interorganizational connections (Aldritch 1977; Galaskiewicz and Marsden 1978; Fennema and Schijf 1979), community structure (Freeman 1968; Fischer et al. 1977), ethnography (Wolfe 1978), acquaintance (de Sola Pool and Kochen 1978), job opportunities (Granovetter 1974; Boorman 1975), mental health (Burns 1974; Tolsdorf 1976), family organization (Bott 1971; Noble 1970), racial integration (Schofield and Sager 1977), political processes (Barnes 1969), diffusion of innovations (Rogers 1979), and mainstreaming or the integration of educable mentally retarded children in normal classrooms (Ballard et al. 1977). Leinhardt (1977) contains a selection of earlier studies, while Holland and Leinhardt (1979a) consists of reports of more recent research. The journal *Social Networks* regularly publishes research in this area.

* Paul W. Holland is Director, Program Statistics Research, Educational Testing Service, Princeton, NJ 08541, and Samuel Leinhardt is Professor of Sociology, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, PA 15213. This research is part of an ongoing project that is collaborative in every respect. Consequently, the authors' names appear in alphabetical order. Our research activities were supported in part by grants from the National Science Foundation (SOC 77–25821), the Foundation for Child Development, and the Department of Health, Education and Welfare (1 RO1 HD 12506–01) to Carnegie-Mellon University, and a National Science Foundation grant (SOC 77–25823) to Educational Testing Service. The opinions expressed do not necessarily reflect the positions or policies of the supporting agencies and organizations, and no official endorsement should be inferred. An earlier version of this report was presented at the Advanced Research Symposium on Stochastic Process Models of Social Structure, sponsored by the Mathematical Social Science Board at Carnegie-Mellon University, 1977. We are extremely grateful to Ariel Ish-Shalom and Kathy Blackmond for programming and general research assistance.

March 1981, Volume 76, Number 373
Invited Papers Section

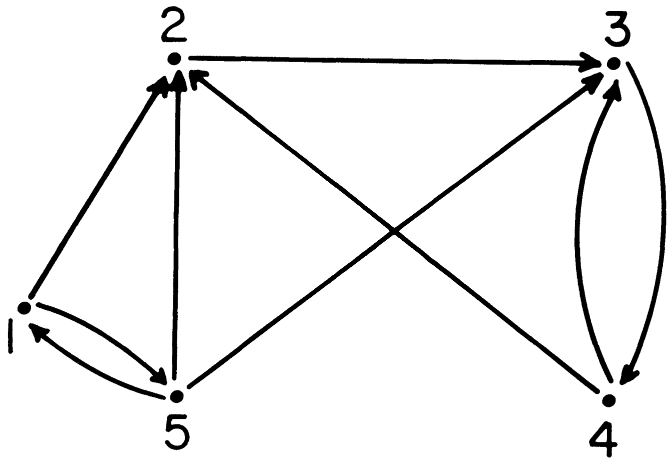


Figure 1. Digraph With Five Nodes and Nine Directed Edges

With all the substantive variety that is present in contemporary social network research, it is surprising to discover that there is a paucity of statistical tools available. The aim of this paper is to begin to fill this gap by providing a simple, yet flexible, family of probability distributions that can be used to analyze certain types of digraph data. In our opinion the most important aspect of the model we present is that it allows for the *simultaneous* estimation of parameters that measure both the amount of *reciprocation* of directed edges between nodes (i.e., our parameter ρ) and the amount of *differential attractiveness* exhibited by each node (i.e., our parameter β_j). Furthermore, these parameters are directly comparable across digraphs that differ in the number of nodes and directed edges they contain.

In the discussion that follows we use the phrase "i relates to j" as shorthand for the more ponderous "i stands in the given relationship to j"; similarly, we use the terms *relation* and *relationship* interchangeably. While diagrams like Figure 1 are sometimes useful, for most analytic purposes the adjacency matrix of the digraph is more convenient. This is the g-by-g matrix of indicator variables, X, defined by

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ relates to } j, \\ 0 & \text{otherwise.} \end{cases}$$

We will always set $X_{ii} = 0$ by convention. The adjacency matrix for Figure 1 is given in Figure 2.

	1	2	3	4	5	X_{i+}
1	0	1	0	0	1	2
2	0	0	1	0	0	1
3	0	0	0	1	0	1
4	0	1	1	0	0	2
5	1	1	1	0	0	3
X_{+j}	1	3	3	1	1	$9 = X_{++}$

Figure 2. The Adjacency Matrix for Figure 1

In their applied substantive contexts social networks are complex social phenomena that exist over time and encompass actors who may be free to enter and leave the network at will and who do not necessarily share identical attributes. Social network research has exhibited the usual historical trend of developing procedures for analyzing more and more realistic situations. The currently available methods are typically based on deterministic models, and stochastic models are only just becoming available for many important applications. To locate our contribution in the context of social network analysis, we propose the following classification of digraph data. The scheme reflects, in varying degrees, the types of complications that obtain in real social network data.

a. *Single relationship data.* A single relationship observed on a set of nodes at a single point in time (e.g., "friendship" in a given school classroom observed during one week in November).
b. *Time series data.* There may be more than one point in time at which the relation on the set of nodes is observed (e.g., friendship in a school classroom observed on the first Monday of September, October, November, and December).
c. *Covariates.* There may be information about nodal attributes in addition to the relationship information (e.g., in a classroom study we may also observe each student's sex, race, etc., as well as his or her friendships).
d. *Valued relationship.* Some types of relationships exist in varying degrees or strengths rather than in an all-or-none fashion (e.g., children may be asked to rate the intensity of their friendship with each child in a classroom).
e. *Multiple relationships.* There may be more than one type of relationship studied on the same set of nodes (e.g., the relationship of friendship and the relationship of team membership[1]).

There are other complications that can arise, but the preceding list illustrates those that are important to the study of social networks (Davis and Leinhardt 1972). In other applied contexts, such as physics (Kinderman and Snell 1980) or sample surveys (Frank 1978), other complications may prove to be of greater relevance. Since the present work derives its motivation from social network research, we concern ourselves with developing an approach that facilitates the incorporation of the complications characterized by cases (b) through (e). We focus here on a fundamental framework, a family of parametric probability models that are appropriate for case (a), single relationship data. In Section 5 we briefly consider an extension to the multivariate case of several adjacency matrices defined on the same set of nodes, case (e).

There is a small amount of recent work that is related

[1] We will include symmetric relations as a special case of directed relations. In this case $X_{ij} = X_{ji}$ for all i, j.

to the approach we take here. For example, in Holland and Leinhardt (1977a,b) we proposed a class of stochastic process models that could serve as the basis for extending the work reported here to case (b), time series data. Wasserman (1977, 1979, 1980) and Galaskiewicz and Wasserman (1979) developed that work further. Sørensen and Hallinan (1976) and Hallinan (1978) reported related methodological and empirical research. Recently, Fienberg and Wasserman (1979, 1981) also considered probability models for social network data, using ideas that are closely related to ours. They studied cases (c) and (e), covariates and multiple relationships. Case (d), *valued relations*, was recently studied using analysis of variance models by Warner, Kenny, and Stoto (1979) and Kenny and Nasby (1980).

Although the literature contains little on the problem of fitting and estimating parametric probability models for digraph data, there is an extensive body of work on the analysis of social network data. This literature falls roughly into three types—tests of randomness, pattern detection, and measures of structure.

Examples of tests of randomness include the work of Katz (1951) on the distribution of the number of isolates in a random digraph, White's (1977) work on the random distribution of zero blocks in an adjacency matrix, and Holland and Leinhardt's (1978) work on the distribution of triads in a random graph. Examples of pattern detection methods include the many clique-finding algorithms (Nosanchuk 1963; Alba 1973; Roistacher 1974), blockmodeling procedures (White, Boorman, and Breiger 1976; and Boorman and White 1976; Arabie, Boorman, and Levitt 1978; Light and Mullins 1979) and spatial representations of digraphs (Levine 1972). Measures of structure are exemplified by structural measures of balance (Harary, Norman, and Cartwright 1965), connectivity (Luce 1950; Barnes 1966; Doreian 1974) and centrality (Moxley and Moxley 1974; Freeman 1977, 1979). Reviews of all three of these topics can be found in Burt (1980) and Leik and Meeker (1975).

2. THE p_1 DISTRIBUTION

We base our model on two empirical observations that have been made repeatedly in studies of social networks—from friendship among individuals to interlocks among the directors of corporations. To state these two observations precisely, we need to develop more notation. Let M denote the number of pairs $\{i, j\}$ for which $X_{ij} = X_{ji} = 1$. Then M may be computed as

$$M = \sum_{i<j} X_{ij}X_{ji} . \tag{1}$$

Thus M is the number of reciprocated or symmetric or mutual relationships in X.[2] The *in-degree* of node j is

$$X_{+j} = \sum_{i=1}^{g} X_{ij} , \tag{2}$$

[2] Moreno (1934) tends to use the term *mutuality*. Davis (1977) employs *symmetry*. In Katz and Powell (1955) and Katz, Tagiuri, and Wilson (1958), they are used interchangeably. We use *reciprocity*.

so that X_{+j} is the number of nodes i for which $X_{ij} = 1$. The in-degrees $\{X_{+j}\}$ form a set of numbers with mean, $\bar{X}$, and variance, $V(\text{in})$, defined by

$$\bar{X} = \left(\sum_{j=1}^{g} X_{+j}\right) \Big/ g = X_{++}/g \tag{3}$$

and

$$V(\text{in}) = \left(\sum_{j=1}^{g} (X_{+j} - \bar{X})^2\right) \Big/ g , \tag{4}$$

respectively. The *out-degree* of node i is

$$X_{i+} = \sum_{j=1}^{g} X_{ij} . \tag{5}$$

The mean of the out-degrees is also $\bar{X}$, and their variance, $V(\text{out})$, is defined in obvious analogy to (4).

In the earliest sociometric studies, Moreno (1934) found that M and $V(\text{in})$ usually exceeded their "chance" expected values. To Moreno, empirical sociometric data always seemed to exhibit a "surplus" of mutual relationships, while some individual group members always managed to attract a "surplus" of choices (Moreno and Jennings 1938). Moreno posited a simple null model for X in which all adjacency matrices with out-degrees agreeing with those in the data are equally likely. We denote this probability distribution by conditioning on $\{X_{i+}\}$. The chance expectations of M and $V(\text{in})$ under this null distribution may be shown to be

$$\begin{aligned} E(M \mid \{X_{i+}\}) = (g\bar{X}^2/(2(g-1))) \\ - (gV(\text{out})/(2(g-1)^2)) \end{aligned} \tag{6}$$

and

$$\begin{aligned} E(V(\text{in}) \mid \{X_{i+}\}) = \bar{X} - (\bar{X}^2/(g-1)) \\ - ((g-2)V(\text{out})/(g-1)^2). \end{aligned} \tag{7}$$

The purpose of comparing M and $V(\text{in})$ to (6) and (7) (and other similar types of comparisons) is to show that the digraph's observed edges are *not* distributed randomly and that, in fact, they exhibit expected nonrandom behavior. From intuition and substantive theoretical consideration, many social relationships can be expected to be reciprocated (see Newcomb 1979; Jones and Gerard 1967; Davis 1968), and in these cases we would expect X to exhibit nonrandomness by having a value for M larger than the expected value given by (6). Other types of social relationships, (e.g., "power") can be expected to be nonreciprocated (French 1956; Friedell 1967), and in such cases we would expect M to be smaller than (6). Similarly, from intuition, it should not be surprising that nodes are differentially attractive and that some are involved in more relational ties than are others (Hopkins 1964). This leads to an expectation about the distribution of in-degrees—in Moreno's (1934) terms there will be "stars" (nodes that attract many relations) and "isolates'' (nodes that attract no relations).[3] This will result in large values of $V(\text{in})$ that are larger than the expected value given in (7).

From these empirical observations and substantive theoretical predictions, we wanted to construct a family of distributions for X with parameters that allow us to control the probability of observing different values of M and $\{X_{+j}\}$. Exponential families of distributions are natural choices to consider for this purpose, since they explicitly tie sufficient statistics to parameters. To be more precise, let G denote the set of all g-by-g adjacency matrices so that X may be thought of as a random matrix taking values in G (see Katz and Powell 1955). Let x denote a generic point of G; then let $p_1(x)$ be the probability function[4] on G given by

$$\begin{aligned} p_1(x) &= P(X = x) \\ &= \exp\{\rho m + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_j \beta_j x_{+j}\} \\ &\quad \times K(\rho,\theta,\{\alpha_i\},\{\beta_j\}) \end{aligned} \tag{8}$$

where m, x_{++}, x_{i+}, and x_{+j} are the values of M, X_{++}, X_{i+}, and X_{+j} computed from x. In (8), ρ, θ, α_i, and β_j are parameters with α_i and β_j subject to the identifying constraint $\alpha_+ = \beta_+ = 0$. These parameters control the probability of observing X with specific values of M, X_{+j}, and X_{i+}. The function $K(\rho, \theta, \{\alpha_i\}, \{\beta_j\})$ in (8) is a normalizing constant that insures that $p_1(x)$ sums to 1 over all x in G. Generally speaking, one can always get this far with an exponential family, but unless K can be computed explicitly, little more can be done. Fortunately, there is a simple derivation of (8) from basic assumptions that leads to a formula for K as well as to a deeper understanding of the model. We shall develop (8) from this alternative point of view before we proceed further.

2.1 Derivation of the p_1 Distribution

We first decompose X into its $\binom{g}{2}$ dyads or pairs, $D_{ij} = (X_{ij}, X_{ji})$ for $i<j$. The distribution of X may be specified by giving the joint distribution of the pairs, D_{12}, D_{13}, and so on. To describe the joint distribution of the $\{D_{ij}\}$, we first assume that the D_{ij} are all statistically independent. This independence assumption means that p_1 *cannot* express tendencies toward transitivity, cliquing, hierarchy, and so on, other than those already implied by tendencies toward reciprocation and differential attraction. In this sense, p_1 is essentially a null model that is more realistic than models that do not express tendencies toward reciprocation and differential attraction. However, in Holland and Leinhardt (1978) we present empirical evidence that the assumption of dyad independence may be satisfied in a substantial number of groups studied by social network analysts. Thus in addition to providing a null model, the p_1 family of distributions may also provide adequate models for representing certain types of empirical data. Finally, we point out that it appears to be difficult to relax the independence assumption and to retain the tractability of the model.

Having assumed that the $\{D_{ij}\}$ are independent, we need only specify the distribution of each D_{ij}, $i < j$, in order to completely specify the distribution of X. This is done by specifying values of m_{ij}, a_{ij}, and n_{ij} where

$$m_{ij} = P(D_{ij} = (1,1)) \quad i<j , \tag{9}$$

$$a_{ij} = P(D_{ij} = (1,0)) \quad i \neq j , \tag{10}$$

$$n_{ij} = P(D_{ij} = (0,0)) \quad i<j , \tag{11}$$

and

$$m_{ij} + a_{ij} + a_{ji} + n_{ij} = 1, \quad \text{for all} \quad i<j . \tag{12}$$

In (9) m_{ij} is the probability that the dyad i, j is a mutual or reciprocated pair; in (10) a_{ij} is the probability that the dyad i, j is an asymmetric or nonreciprocated pair; in (11) n_{ij} is the probability that the dyad i, j is a null pair. The probability distribution of X may be expressed in the following way:

$$\begin{aligned} &P(X = x) \\ &= \prod_{i<j} m_{ij}^{x_{ij}x_{ji}} \prod_{i\neq j} a_{ij}^{x_{ij}(1-x_{ji})} \prod_{i<j} n_{ij}^{(1-x_{ij})(1-x_{ji})} . \end{aligned} \tag{13}$$

This may be reexpressed as follows to emphasize the exponential form of (13):

$$P(X = x) = \exp\{\sum_{i<j} \rho_{ij} x_{ij} x_{ji} + \sum_{i\neq j} \theta_{ij} x_{ij}\} \prod_{i<j} n_{ij} , \tag{14}$$

where

$$\rho_{ij} = \log_e((m_{ij} n_{ij})/(a_{ij} a_{ji})) \quad i<j \tag{15}$$

and

$$\theta_{ij} = \log_e(a_{ij}/n_{ij}) \; ; \quad i \neq j \tag{16}$$

and in (16) we interpret $n_{ji} = n_{ij}$ for $i>j$. The exponential or ''natural'' parameters, ρ_{ij} and θ_{ij}, are equivalent to the original set of parameters m_{ij}, a_{ij}, and n_{ij} when these are subjected to the constraint specified by (12).

The parameter, ρ_{ij}, is a log-odds ratio, and a little algebra reveals that it gives the log of the increase in the odds that $X_{ij} = 1$ due to $X_{ji} = 1$, that is,

$$\exp(\rho_{ij}) = \frac{P(X_{ij}=1 \mid X_{ji}=1)}{P(X_{ij}=0 \mid X_{ji}=1)} \Big/ \frac{P(X_{ij}=1 \mid X_{ji}=0)}{P(X_{ij}=0 \mid X_{ji}=0)} . \tag{17}$$

Thus ρ_{ij} measures the ''force of reciprocation'' in the sense that if ρ_{ij} is positive and if $X_{ji} = 1$, then we are more likely to also observe that $X_{ij} = 1$.

The parameter, θ_{ij}, is a log-odds. Again, a little algebra shows that

$$\exp(\theta_{ij}) = P(X_{ij}=1 \mid X_{ji}=0)/P(X_{ij}=0 \mid X_{ji}=0) . \tag{18}$$

---

[3] Strictly speaking, one can only expect literal isolates when choice volume is small.

[4] We denote the probability function in (8) by $p_1(x)$ to emphasize our view that it is the first or simplest family of distributions on digraphs that might be considered for social network data. This is because it expresses the two elementary social tendencies of reciprocation and differential attraction.

Thus θ_{ij} measures the probability of an asymmetric dyad between i and j, given that $X_{ji} = 0$.

As it stands, the distribution in (14) is not the same as that given as $p_1(\cdot)$ in (8). In fact, (14) is a more general family of distributions for X than (8) is, but it has too many parameters to be useful for many statistical purposes. To obtain (8) from (14), we impose restrictions on ρ_{ij} and θ_{ij}. We set

$$\rho_{ij} = \rho \quad \text{for all} \quad i < j \,, \tag{19}$$

and

$$\theta_{ij} = \theta + \alpha_i + \beta_j \quad \text{for all} \quad i \neq j \,, \tag{20}$$

where

$$\alpha_+ = \beta_+ = 0.$$

Assumptions (19) and (20) lead to the following formula for $p_1(x)$:

$$p_1(x) = \exp\{\rho m + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_j \beta_j x_{+j}\} \times \prod_{i<j} n_{ij} \,, \tag{21}$$

where the n_{ij} are functions of the parameters ρ, θ, $\{\alpha_i\}$, and $\{\beta_j\}$, given in (24) and (25).

The restriction (19) has the interpretation that the "force of the reciprocation" is independent of the nodes involved. With this restriction, ρ may be interpreted as the average tendency toward reciprocation for all pairs of nodes. It is natural to consider weakening restriction (19) when generalizing $p_1(\cdot)$. We will not pursue this here. Restriction (20) implies that the probability that $X_{ij} = 1$ given $X_{ji} = 0$ (as measured by the odds in (18)) is the product of a factor for node i and another factor for node j. It is analogous to logit models in more standard statistical problems involving binary data (Cox 1970).

It is also useful to solve for m_{ij}, a_{ij}, and n_{ij} in terms of the exponential parameters. This yields

$$m_{ij} = \exp\{\rho + 2\theta + \alpha_i + \alpha_j + \beta_i + \beta_j\}/k_{ij} \,, \tag{22}$$

$$a_{ij} = \exp\{\theta + \alpha_i + \beta_j\}/k_{ij} \,, \tag{23}$$

$$n_{ij} = 1/k_{ij} \,, \tag{24}$$

where

$$k_{ij} = 1 + e^{\theta+\alpha_i+\beta_j} + e^{\theta+\alpha_j+\beta_i} + e^{\rho+2\theta+\alpha_i+\alpha_j+\beta_i+\beta_j} \,. \tag{25}$$

Equations (24) and (25) express n_{ij} explicitly as a function of the exponential parameters and may be used to derive a formula for $K(\cdot)$ in (8).

We have already discussed one possible interpretation of the parameter ρ in $p_1(\cdot)$. Possible interpretations of the other parameters, θ, α_i, and β_j, follow. If we set ρ, $\{\alpha_i\}$, and $\{\beta_j\}$ all equal to zero, then the resulting distribution on G is equivalent to assuming that the X_{ij} are all independent and identically distributed (iid) indicator variables with $p = P(X_{ij} = 1)$ and $\theta = \log_e(p/(1 - p))$. In this sense, θ governs the density of ones in X or edges in the digraph. Therefore, we refer to θ as a *density parameter*.[5] If we let θ and $\{\alpha_i\}$ be nonzero (but keep ρ and $\{\beta_j\}$ zero), then the resulting distribution on G is equivalent to assuming that the X_{ij} are iid in each row of X with a common $p_i = P(X_{ij} = 1)$. In this case, $\theta + \alpha_i = \log_e(p_i/(1 - p_i)$, so that α_i governs differences in the distributions of the out-degrees of X. Hence we may call α_i a *productivity parameter* since, if α_i is large and positive, node i will tend to have a relatively large out-degree or will appear to "produce" relational ties (see Duck 1977). If α_i is large and negative, then node i will produce relatively few ties and X_{i+} will tend to be zero or small. If we allow θ and β_j to be nonzero (but keep ρ and α_i zero), then the resulting distribution on G is equivalent to assuming that the X_{ij} are iid in each column of X with common $p_j = P(X_{ij} = 1)$. In this case, $\theta + \beta_j = \log_e(p_j/(1 - p_j))$, so that β_j governs differences in the distributions of the in-degrees, X_{+j}. Hence we may call β_j an *attractiveness parameter* since, if β_j is large and positive, node j will tend to have a large in-degree or will appear to "attract" relational ties (see Berscheid and Walster 1977; Huston 1974). If β_j is large and negative, then node j will attract few ties and X_{+j} will tend to be zero or small.

In the preceding discussion we indicated how setting various exponential parameters of p_1 to zero corresponds to easily interpreted distributions on G. We summarize these and other special cases of p_1 in Table 1.

Table 1: Selected Special Cases of p_1

Parameter Values	Interpretation
$\rho = \theta = \alpha_i = \beta_j = 0$	: The uniform distribution over G in which all digraphs are equally likely.
$\rho = \alpha_i = \beta_j = 0$	: X_{ij} are iid, $\theta = \log_e(p/(1 - p))$.
$\alpha_i = \beta_j = 0$	: D_{ij} are iid; $m_{ij} = m$, $a_{ij} = a_{ji} = a$, $n_{ij} = n$ and $m + 2a + n = 1$.
$\rho = \beta_j = 0$	: X_{ij} are iid in each row of X; $\theta + \alpha_i = \log_e(p_i/(1 - p_i))$,
$\rho = \alpha_i = 0$	: X_{ij} are iid in each column of X; $\theta + \beta_j = \log_e(p_j/(1 - p_j))$.
$\rho = 0$	: X_{ij} are independent; logit of p_{ij} is additive.
$\rho = \infty$	: $X_{ij} = X_{ji}$ and the graph is symmetric.
$\rho = -\infty$	: $X_{ij}X_{ji} = 0$ and the digraph is asymmetric.

2.2 Simulated Digraphs From the p_1 Distribution

To provide an informal feeling for the types of digraphs that the p_1 distribution will generate, we present in Figure 3 four simulated adjacency matrices for four different sets of parameter combinations. It is easy to simulate random digraphs from $p_1(x)$ because the $D_{ij} = (X_{ij}, X_{ji})$ are independent. We used the following procedure. For specified values of ρ, θ, $\{\alpha_i\}$, and $\{\beta_j\}$, calculate m_{ij}, a_{ij}, and n_{ij} from (22), (23), and (24). A pseudorandom number

[5] This is similar to Loomis and Proctor's (1951) notion of "gross expansiveness" in the affective sociometric context. The term *density* seems more context free.

a. *Matrices*

Case 1

	X_{ij}										X_{+j}
	0	1	0	1	0	0	1	0	0	0	3
	0	0	1	0	1	0	0	0	0	1	3
	0	0	0	0	0	0	1	0	1	0	2
	0	0	0	0	0	0	1	0	0	0	1
	1	0	0	1	0	1	1	1	0	0	5
	0	0	1	1	1	0	0	1	0	0	4
	0	1	0	0	0	0	0	0	0	0	1
	1	0	0	1	1	1	1	0	0	0	5
	0	1	1	0	0	0	1	0	0	0	3
	0	0	0	1	1	0	0	1	1	0	4
X_{i+}	2	3	3	5	4	2	6	3	2	1	$31 = X_{++}$

Case 2

	X_{ij}										X_{+j}
	0	1	0	0	0	0	1	0	0	0	2
	1	0	1	0	1	0	0	0	0	0	3
	0	1	0	0	0	0	1	0	1	0	3
	1	0	1	0	0	0	1	0	0	0	3
	1	1	1	1	0	1	1	1	0	0	7
	1	1	1	1	1	0	0	0	0	0	5
	0	1	1	0	0	0	0	0	0	0	2
	1	0	0	0	1	1	0	0	0	0	3
	0	1	0	0	0	0	0	0	0	0	1
	0	1	0	1	0	0	0	0	0	0	2
X_{i+}	5	7	5	3	3	2	4	1	1	0	$31 = X_{++}$

Case 3

	X_{ij}										X_{+j}
	0	1	0	1	1	0	1	0	0	0	4
	1	0	1	0	1	0	1	0	0	1	5
	0	1	0	0	0	1	1	0	1	0	4
	1	0	1	0	0	0	1	0	0	0	3
	0	1	1	0	0	1	1	1	0	0	5
	1	0	0	0	1	0	0	0	0	0	2
	1	1	1	0	1	0	0	0	0	0	4
	1	0	0	0	1	1	0	0	0	0	3
	0	1	1	0	0	0	0	0	0	0	2
	0	0	0	0	0	0	0	0	0	0	0
X_{i+}	5	5	5	1	5	3	5	1	1	1	$32 = X_{++}$

Case 4

	X_{ij}										X_{+j}
	0	0	1	1	1	0	0	0	0	0	3
	1	0	1	0	0	0	1	0	0	0	3
	0	1	0	0	0	0	0	0	0	0	1
	0	1	0	0	0	0	0	0	0	0	1
	1	1	1	1	0	0	1	0	0	0	5
	0	0	1	1	1	0	0	0	0	0	3
	1	0	0	0	0	0	0	0	1	1	3
	0	1	0	0	1	0	1	0	0	0	3
	1	1	1	1	0	0	0	0	0	0	4
	1	1	0	1	1	0	0	0	0	0	4
X_{i+}	5	6	5	5	4	0	3	0	1	1	$30 = X_{++}$

b. *Parameter values and summary statistics*

Case	θ	ρ	α_j	β_j									
1	−.69	0	0	0	0	0	0	0	0	0	0	0	0
2	−.90	0	0	1.5,	1.5,	1.5,	0,	0,	0,	0,	−1.5,	−1.5,	−1.5
3	−1.67	2	0	1.5,	1.5,	1.5,	0,	0,	0,	0,	−1.5,	−1.5,	−1.5
4	−0.42	−2	0	1.5,	1.5,	1.5,	0,	0,	0,	0,	−1.5,	−1.5,	−1.5

Case	$\bar{X}$	M	$E(M \mid X_{i+})$	V(in)	$E(V(in) \mid X_{i+})$
1	3.1	4	5.20	2.09	1.85
2	3.1	6	5.17	4.29	1.78
3	3.2	11	5.56	3.56	1.85
4	3.0	2	4.91	4.80	1.86

Figure 3. Four Examples of Digraphs Simulated From the p_1 Distribution

from the uniform distribution is then used to simulate one of the four events: $D_{ij} = (1,1), (1,0), 0,1)$, or $(0,0)$. Repeat this operation independently for all $g(g-1)/2$ dyads, D_{ij}, and X is thereby simulated.

In Figure 3 all four cases have been chosen to make $E(\bar{X}) = 3.0$, a commonly observed value in empirical data (see, e.g., Bjerstedt 1956). The parameter combinations are chosen to emphasize different features of the p_1 distribution.

In case 1 all parameters except θ are set to zero. The observed values of M and $V(\text{in})$ for this case are near their expected values, given X_{i+}, because $\rho = 0$ and $\beta_j = 0$.

In case 2 there is differential attraction, with nodes 1, 2, and 3 being the most highly attractive, and nodes 8, 9, and 10 the least. Here $\rho = 0$, and therefore M is near its null expected value. Since $\beta_j \neq 0$ the digraph exhibits differential attraction, and therefore $V(\text{in})$ exceeds its null expectation by more than a factor of two.

Case 3 has $\rho = 2$ and the same set of nonzero β_j as in case 2. Thus both reciprocity and differential attraction are present. Here both M and $V(\text{in})$ exceed their null expected values by nearly factors of two.

Case 4 is the same as case 3 except that $\rho = -2$, so there is a tendency away from reciprocation. Here M is less than half its null expected value, while $V(\text{in})$ exceeds its null expected value by more than a factor of two.

Although these four examples do not exhaust the possibilities, they illustrate how, by varying the values of the parameters of the p_1 distribution, we are able to independently vary tendencies toward reciprocation and differential attraction. Any structural features that may be detected in these simulated digraphs are either implied by tendencies towards reciprocation and differential attraction or are accidents of chance. Similar remarks are in order when structural features are observed in empirical data that fit a p_1 model.

3. ESTIMATION AND TESTING USING THE p_1 DISTRIBUTION

In order to use $p_1(x)$ for data analysis, we need to be able to estimate the parameters of $p_1(x)$, that is, the vector

$$\pi = (\rho, \theta, \alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g)\,. \qquad (26)$$

Since $\alpha_+ = \beta_+ = 0$, π ranges over a $(2g)$-dimensional space. We shall denote the maximum likelihood estimates (MLE) of these parameters by $\hat{\pi}$ or $\hat{\rho}$, $\hat{\theta}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$.

If X is the observed adjacency matrix, then the likelihood function is

$$p_1(X) = \exp\{\rho M + \theta X_{++} + \sum_i \alpha_i X_{i+} + \sum_j \beta_j X_{+j}\} \prod_{i<j} c_{ij} \qquad (27)$$

where

$$c_{ij} = (1 + e^{\theta+\alpha_i+\beta_j} + e^{\theta+\alpha_j+\beta_i} + e^{\rho+2\theta+\alpha_i+\alpha_j+\beta_i+\beta_j})^{-1}\,.$$

Since p_1 is an exponential family, the likelihood equations, found by differentiating (27) with respect to the parameters and setting the resulting system equal to zero, must have the form "sufficient statistics equal their expected values." Thus the likelihood equations that are needed to find the MLE of π are

$$M = E_\pi(M) = \sum_{i<j} m_{ij}\,, \qquad (28)$$

$$X_{i+} = E_\pi(X_{i+}) = \sum_j (m_{ij} + a_{ij}),\ i = 1, \ldots, g\,, \qquad (29)$$

$$X_{+j} = E_\pi(X_{+j}) = \sum_i (m_{ij} + a_{ij}),\ j = 1, \ldots, g\,. \qquad (30)$$

Note that in (29) and (30) we have expanded the definitions of m_{ij} and a_{ij}. For $i > j$ we set $m_{ij} = m_{ji}$ and let $m_{ii} = 0$. We also set $a_{ii} = 0$ and expand n_{ij} to a full g-by-g matrix by setting $n_{ij} = n_{ji}$ for $i < j$ and $n_{ii} = 0$. Thus (m_{ij}), (a_{ij}), and (n_{ij}) are all g-by-g matrices with zero main diagonals; (m_{ij}) and (n_{ij}) are also symmetric. These conventions simplify the subsequent discussion.

The MLE of π is the solution to the system (28), (29), and (30).

There are two approaches that are commonly used to solve such systems—one *direct* and one *indirect*. The direct approach, exemplified by Newton-method iterations, Fisher's method of scoring, and various weighted least squares iterations, sets up an iterative system of approximations to $\hat{\pi}$. The indirect approach sets up an iterative system of approximations to $\hat{m}_{ij}$, $\hat{a}_{ij}$, and $\hat{n}_{ij}$ (defined by (22), (23), (24), with $\hat{\pi}$ substituted for π). For this problem, generalized iterative scaling, described and analyzed by Darroch and Ratcliff (1972), is the natural candidate for the indirect approach.

There are two main drawbacks to the direct approach here. First, there are, potentially, a large number of parameters—$2g$—and this will result in large matrices ($2g$ by $2g$) and the need for careful numerical methods in the iterations. Second, it is easy to have cases in which one or more of the $\hat{\beta}_j = -\infty$ (e.g., if $X_{+j} = 0$). This situation causes nonconvergence in Newton-method and related approaches unless they have special adjustments to deal with it.

The indirect approach—generalized iterative scaling—suffers from neither of these two drawbacks. The largest matrices that arise are g by g, and the computations done on them are simple row and column multiplications. When $\hat{\beta}_j = -\infty$, the corresponding $\hat{m}_{ij}$ or $\hat{a}_{ij}$ are zero, and iterative scaling automatically adjusts for this. We used a version of iterative scaling to estimate $\hat{\pi}$ and will describe the algorithm here.

3.1 An Iterative Scaling Algorithm[6]

We have tried several variants of iterative scaling for fitting the p_1 distribution to data. The following algorithm

[6] FORTRAN code for these algorithms is available from the authors. The code was developed on a DEC-20 machine. Other algorithms have been developed since this work was initiated (e.g., Fienberg and Wasserman 1979).

was suggested to us by Y. Wang. It is quite simple and can be shown to converge to the MLE by using the methods described by Darroch and Ratcliff (1972).

Let $(m_{ij}^{(n)})$, $(a_{ij}^{(n)})$, and $(n_{ij}^{(n)})$ be the nth iterates in a sequence of approximations to the MLE's, $(\hat{m}_{ij})$, $(\hat{a}_{ij})$, and $(\hat{n}_{ij})$. We begin with initial values $(m_{ij}^{(0)})$, $(a_{ij}^{(0)})$, and $(n_{ij}^{(0)})$, which satisfy (22), (23), and (24) for some set of values $\theta^{(0)}$, $\rho^{(0)}$, $\{\alpha_i^{(0)}\}$, $\{\beta_j^{(0)}\}$. For example, if we set $m_{ij}^{(0)} = a_{ij}^{(0)} = n_{ij}^{(0)} = .25$ for all $i \neq j$, and $m_{ii}^{(0)} = a_{ii}^{(0)} = n_{ii}^{(0)} = 0$, then these initial values satisfy (22), (23), and (24), with $\theta^{(0)} = \rho^{(0)} = \alpha_i^{(0)} = \beta_j^{(0)} = 0$. The iterations proceed in cycles of four steps, which we call the row step, the column step, the mutual step, and the normalizing step, respectively.

The row step: For all $i \neq j$,

$$\begin{aligned} m_{ij}^{(n+1)} &= m_{ij}^{(n)} (F_i^{(n)} F_j^{(n)})^{1/2} \\ a_{ij}^{(n+1)} &= a_{ij}^{(n)} (F_i^{(n)} K^{(n)})^{1/2} \\ n_{ij}^{(n+1)} &= n_{ij}^{(n)} K^{(n)}, \end{aligned} \tag{31}$$

where

$$F_i^{(n)} = X_{i+}/(m_{i+}^{(n)} + a_{i+}^{(n)}) \tag{32}$$

$$K^{(n)} = (g(g-1) - X_{++})/(a_{++}^{(n)} + n_{++}^{(n)}). \tag{33}$$

The column step: For all $i \neq j$,

$$\begin{aligned} m_{ij}^{(n+2)} &= m_{ij}^{(n+1)} (G_i^{(n+1)} G_j^{(n+1)})^{1/2} \\ a_{ij}^{(n+2)} &= a_{ij}^{(n+1)} (G_j^{(n+1)} K^{(n+1)})^{1/2} \\ n_{ij}^{(n+2)} &= n_{ij}^{(n+1)} K^{(n+1)}, \end{aligned} \tag{34}$$

where

$$G_j^{(n+1)} = X_{+j}/(m_{+j}^{(n+1)} + a_{+j}^{(n+1)}) \tag{35}$$

and

$$K^{(n+1)} = (g(g-1) - X_{++})/(a_{++}^{(n+1)} + n_{++}^{(n+1)}). \tag{36}$$

The mutual step: For all $i \neq j$,

$$\begin{aligned} m_{ij}^{(n+3)} &= m_{ij}^{(n+2)} H^{(n+2)} \\ a_{ij}^{(n+3)} &= a_{ij}^{(n+2)} L^{(n+2)} \\ n_{ij}^{(n+3)} &= n_{ij}^{(n+2)} L^{(n+2)}, \end{aligned} \tag{37}$$

where

$$H^{(n+2)} = M/(\tfrac{1}{2} m_{++}^{(n+2)}) \tag{38}$$

and

$$L^{(n+2)} = \left[\binom{g}{2} - M\right] \Big/ \left[\binom{g}{2} - (\tfrac{1}{2} m_{++}^{(n+2)})\right]. \tag{39}$$

The normalizing step: For all $i \neq j$,

$$\begin{aligned} m_{ij}^{(n+4)} &= m_{ij}^{(n+3)}/R_{ij}^{(n+3)} \\ a_{ij}^{(n+4)} &= a_{ij}^{(n+3)}/R_{ij}^{(n+3)} \\ n_{ij}^{(n+4)} &= n_{ij}^{(n+3)}/R_{ij}^{(n+3)}, \end{aligned} \tag{40}$$

where

$$R_{ij}^{(n+3)} = m_{ij}^{(n+3)} + a_{ij}^{(n+3)} + a_{ji}^{(n+3)} + n_{ij}^{(n+3)}. \tag{41}$$

The full algorithm consists of chaining together these four steps into a single cycle and repeating the cycle until convergence. The output of the normalizing step is used as the initial values for the row step in the next cycle. Our experience with the algorithm suggests that it is a practical way to fit the p_1 distribution to adjacency matrices for which g is as large as 60. We have had no experience with fitting larger matrices, but we do not expect that they would create problems beyond the obvious ones involved with storage and machine time.

The algorithm just described fits the full p_1 distribution. However, we are also interested in fitting submodels of p_1 (e.g., those described in Sec. 2) to data. For example, the submodel of p_1 for which $\rho = 0$ is important for testing hypotheses about ρ. It may be estimated by maximum likelihood in a number of ways. One way is to leave out the mutual step in the algorithm just described. We can then obtain the MLE's of m_{ij}, a_{ij}, and n_{ij} for p_1 with $\rho = 0$. We have also used the following algorithm for fitting the $\rho = 0$ case. Let.

$$p_{ij} = m_{ij} + a_{ij} = P(X_{ij} = 1) \tag{42}$$

and

$$q_{ij} = 1 - p_{ij} = P(X_{ij} = 0). \tag{43}$$

When $\rho = 0$, the X_{ij} are independent and p_{ij} satisfies

$$\log_e(p_{ij}/q_{ij}) = \theta + \alpha_i + \beta_j, \quad i \neq j. \tag{44}$$

The likelihood equations for this submodel of p_1 may be expressed in terms of the p_{ij}. They are

$$X_{i+} = p_{i+}, \quad i = 1, \ldots, g \tag{45}$$

$$X_{+j} = p_{+j}, \quad j = 1, \ldots, g. \tag{46}$$

The algorithm creates a sequence of iterates, $p_{ij}^{(n)}$ and $q_{ij}^{(n)}$, that converge to $\hat{p}_{ij}$ and $\hat{q}_{ij}$. There are three steps to each cycle of this algorithm: a row step, a column step and a normalizing step.

The row step: For all $i \neq j$,

$$p_{ij}^{(n+1)} = p_{ij}^{(n)} (X_{i+}/p_{i+}^{(n)}) \tag{47}$$

$$q_{ij}^{(n+1)} = q_{ij}^{(n)} ((g - 1 - X_{i+})/q_{i+}^{(n)}). \tag{48}$$

The column step: For all $i \neq j$,

$$p_{ij}^{(n+2)} = p_{ij}^{(n+1)} (X_{+j}/p_{+j}^{(n+1)}) \tag{49}$$

$$q_{ij}^{(n+2)} = q_{ij}^{(n+1)} ((g - 1 - X_{+j})/q_{+j}^{(n+1)}). \tag{50}$$

The normalizing step: For all $i \neq j$,

$$p_{ij}^{(n+3)} = p_{ij}^{(n+2)}/R_{ij}^{(n+2)} \tag{51}$$

$$q_{ij}^{(n+3)} = q_{ij}^{(n+2)}/R_{ij}^{(n+2)}, \tag{52}$$

where

$$R_{ij}^{(n+2)} = p_{ij}^{(n+2)} + q_{ij}^{(n+2)}. \tag{53}$$

This algorithm is related to the usual iterative scaling algorithm used to fit the model of "no three-factor interaction" to a three-way contingency table. The initial values for this algorithm for fitting p_1 with $\rho = 0$ are $p_{ij}^{(0)} = q_{ij}^{(0)} = .50$.

Another important submodel of p_1 requiring iterative methods for its estimation by maximum likelihood sets $\beta_j = 0, j = 1, \ldots, g$. This submodel can be estimated by dropping the column step from the first algorithm described before, and using $m_{ij}^{(0)} = a_{ij}^{(0)} = n_{ij}^{(0)} = .25$ as the initial values.

Those submodels of $p_1(x)$ described in Section 2, in which $\rho = 0$ *and* either $\alpha_i = 0$ or $\beta_j = 0$, do not need iterative methods for their estimation. For example, if $\rho = 0$ and $\beta_j = 0$ but θ and α are free to vary, then the MLE of $p_i = P(X_{ij} = 1)$ is

$$\hat{p}_i = X_{i+}/(g - 1). \tag{54}$$

Under this model, $\hat{m}_{ij}$, $\hat{a}_{ij}$, and $\hat{n}_{ij}$ may be computed directly from (54).

3.2 Testing Hypotheses Within the p_1 Distribution

The algorithms just described may be used to obtain MLE's of m_{ij}, a_{ij}, and n_{ij} for the full p_1 distribution and for various submodels that we defined by setting certain parameter values to zero. These submodels correspond to hypotheses within the p_1 distribution that have interesting substantive interpretations. We will consider these hypotheses

$$H_0\colon \rho = 0, \{\beta_j = 0\}; \theta, \{\alpha_i\} \text{ unconstrained}, \tag{55}$$

$$H_1\colon \rho = 0; \theta, \{\alpha_i\}, \{\beta_j\} \text{ unconstrained}, \tag{56}$$

$$H_2\colon \{\beta_j = 0\}; \theta, \rho, \{\alpha_i\} \text{ unconstrained}, \tag{57}$$

$$H_3\colon \theta, \rho, \{\alpha_i\}, \{\beta_j\} \text{ all unconstrained}. \tag{58}$$

The hypothesis H_3 corresponds to the full p_1 distribution with no constraints on the parameter values. In H_0, only θ and the $\{\alpha_i\}$ are unconstrained. H_0 corresponds to the assumption that each node produces directed edges at random and that there are no tendencies for reciprocation ($\rho = 0$), nor is any node more attractive than any other ($\beta_j = 0$). H_1 extends H_0 to allow the nodes to be differentially attractive, while H_2 extends H_0 to allow a tendency toward (or away from) reciprocation.

Tests for reciprocation have previously been based on testing H_2 against H_0. This is implicit, for example, in the work of Katz and Wilson (1956). This approach assumes that the attractiveness parameters, β_j, are all zero. For example, the procedure of comparing M to $E(M \mid \{X_{i+}\})$ from (6) may be justified by the fact that the uniformly most powerful unbiased test of H_0 against H_2 may be shown to be based on the conditional distribution of M given $\{X_{i+}\}$ under H_0 (see Lehmann 1959, p. 134). Assuming the relevant normal approximations hold, this is the same as comparing M to $E(M \mid \{X_{i+}\})$ and rejecting H_0 for H_2 if this difference is large compared to the conditional standard deviation of M given $\{X_{i+}\}$ under H_0. See Katz and Wilson (1956) for this variance calculation.

The problem with the procedure just outlined is that in many applications there is often evidence that $\beta_j \neq 0$. Thus to test for reciprocation within the p_1 distribution, it is more natural to test H_1 against H_3. An important use of the p_1 distribution is to allow us to form the likelihood ratio test for H_1 against H_3. The MLE's of m_{ij}, a_{ij}, and n_{ij} may be easily computed under either H_1 or H_3 by the algorithms just described. If $\hat{m}_{ij}$, $\hat{a}_{ij}$, and $\hat{n}_{ij}$ are the MLE's under H_3 and m_{ij}^*, a_{ij}^*, n_{in}^* are MLE's under H_1, then the usual log-likelihood ratio (LLR) statistic for this problem is

$$\text{LLR} = -2 \log_e(\lambda) = L_m + L_a + L_n, \tag{59}$$

where λ denotes the likelihood ratio and,

$$\begin{aligned} L_m &= 2 \sum_{i<j} X_{ij}X_{ji} \log_e \left(\frac{\hat{m}_{ij}}{m_{ij}^*}\right), \\ L_a &= 2 \sum_{i\neq j} X_{ij}(1 - X_{ji}) \log_e \left(\frac{\hat{a}_{ij}}{a_{ij}^*}\right), \\ L_n &= 2 \sum_{i<j} (1 - X_{ij})(1 - X_{ji}) \log_e \left(\frac{\hat{n}_{ij}}{n_{ij}^*}\right). \end{aligned} \tag{60}$$

The reference distribution of LLR from (59) might be expected to be chi squared on one degree of freedom, but the standard theory does not apply in this case. The relevant "sample size" is $g(g - 1)$, which will be large in many applications, but there are $2g - 1$ nuisance parameters, θ, $\{\alpha_i\}$, $\{\beta_j\}$, that are being estimated and they may affect the null distribution of LLR under H_1. We have not explored the theoretical analysis of this distribution problem but have performed a small pilot simulation study[7] (1,000 replications per case) to see if the chi-squared distribution on one degree of freedom is plausible. Table 2 summarizes these preliminary simulation results. There are eight sets of parameter values used in the simulation; these are described in Table 2a. In cases A–10 and B–10, $g = 10$, in A–20 and B–20, $g = 20$, in A–30 and B–30, $g = 30$, in A–40 and B–40, $g = 40$. In the A cases all parameters except θ are zero. In order to examine the effect of differential attractions on the distribution of LLR for the B cases, ρ and α_i were set to zero, but the β_j are not all zero.

Although this pilot study is too small to give definitive results (because of the small number of parameter sets studied, i.e., 8), some useful conclusions and conjectures can be drawn from it. First of all, there is consistent evidence across the eight cases for a modest bias in the use of the one-degree-of-freedom chi-squared distribution for the likelihood ratio test of H_1 against H_3. All of the

[7] Our procedure employed a FORTRAN coded algorithm, RANGEN, to generate random graphs from p_1 as described in Section 2. The code uses the FORTRAN–20 uniform random number generator (RAN) as implemented on the DEC–20 computer.

means exceed one, with an average mean of 1.17. All of the eight variances exceed two, with an average variance of 2.87. All of the eight medians equal or exceed the chi-squared values of .45, with average median value .52. Fifteen of the sixteen obtained percentage values equal or exceed the corresponding chi-squared values. The average of the obtained percentages for the nominal 5 percent chi-squared point is 7.25 percent. The average for the 10 percent point is 12.50 percent. All of this is consistent with the hypothesis that the correct reference distribution for this likelihood ratio test has a slightly heavier upper tail than the chi-squared distribution with one degree of freedom. The degree of this bias, however, appears to vary somewhat across the eight cases studied. It is never so large as to make the use of the chi-squared distribution seriously suspect, and it appears to become smaller as g increases. In Section 4 we investigate an empirical digraph and obtain a value for the LLR statistic of 30.4 for $g = 18$. These pilot simulation results suggest that this value is highly significant.

Two important dimensions are varied in this simulation study—the number of nodes, g, and the values of the "nuisance" parameters β_j. We expected that as g increased, the agreement with the chi-squared distribution would improve. As mentioned previously, this seems to be true. If the means for the A-g and B-g cases are averaged, we obtain values of 1.33, 1.18, 1.15, and 1.03 as g varies from 10 to 40—an apparent tendency to approach the chi-squared values of one as g increases. A similar trend can be observed in the other A, B-g averages displayed in Table 2b. To study the effect of varying values of the nuisance parameters, we averaged the entries in Table 2b separately from the A cases and the B cases. These values are also given in Table 2b. There is a slight tendency for the A cases to be more in agreement with the chi-squared results than are the B cases. This suggests that the values of the nuisance parameters do have a modest effect on the distribution of the likelihood ratio statistic.

These results, while based on a pilot study, are reassuring, and although more detailed simulation studies and theoretical analyses need to be carried out, we do not anticipate any surprises. We intend to report the results of a more extensive simulation study elsewhere. Our main conclusion from this pilot study is that the chi-squared distribution on one degree of freedom is adequate for crude evaluations of the significance levels of the likelihood ratio test of H_1 against H_3.

Other likelihood ratio tests of hypotheses within p_1 may be constructed: for example, a test of $\{\beta_j = 0\}$ that does not also assume that $\rho = 0$ is obtained by forming the likelihood ratio statistic for testing H_2 against H_3. We have not explored the behavior of this test statistic to see if the chi-squared distribution on $g - 1$ degrees of freedom is a reasonable approximation to its distribution under H_2. More research is needed to clarify the use of likelihood tests in these circumstances.

Table 2. Summary of Pilot Simulations of the Log-likelihood Ratio Statistic (LLR) for Testing H_1 Versus H_3 in 8 Cases of the Null Hypothesis (1,000 replications for each case)

a. Parameter Values for Simulation Cases[a]

Case	g	θ	ρ	α	β
A-10	10	−.693	0	0	0
A-20	20	−1.674	0	0	0
A-30	30	−2.159	0	0	0
A-40	40	−2.485	0	0	0
B-10	10	−.906	0	0	$\begin{cases} 1.5 & \text{for } 1 \le j \le 3 \\ 0 & \text{for } 3 < j < 8 \\ -1.5 & \text{for } 8 \le j \le 10 \end{cases}$
B-20	20	−2.100	0	0	$\begin{cases} 1.5 & \text{for } 1 \le j \le 6 \\ 0 & \text{for } 6 < j < 15 \\ -1.5 & \text{for } 15 \le j \le 20 \end{cases}$
B-30	30	−2.647	0	0	$\begin{cases} 1.5 & \text{for } 1 \le j \le 9 \\ 0 & \text{for } 9 < j < 22 \\ -1.5 & \text{for } 22 \le j \le 30 \end{cases}$
B-40	40	−3.001	0	0	$\begin{cases} 1.5 & \text{for } 1 \le j \le 12 \\ 0 & \text{for } 12 < j < 29 \\ -1.5 & \text{for } 29 \le j \le 40 \end{cases}$

b. Summary Statistics for Simulated Values of the Likelihood Ratio Statistic

Case	Mean	Variance	Median	% ≥ 3.84	% ≥ 2.71
Chi-squared 1 df	1	2	.45	5	10
A-10	1.26	3.06	.57	10	14
A-20	1.15	2.42	.59	6	11
A-30	1.14	2.78	.51	7	12
A-40	1.04	2.17	.45	6	11
B-10	1.39	4.65	.56	11	16
B-20	1.21	3.01	.51	8	14
B-30	1.16	2.69	.53	6	12
B-40	1.01	2.14	.45	4	10
Overall Average	1.17	2.87	.52	7.25	12.50
A-average	1.15	2.61	.53	7.25	12.00
B-average	1.19	3.12	.51	7.25	13.00
A, B-10 average	1.33	3.86	.57	10.5	15.0
A, B-20 average	1.18	2.72	.55	7.0	12.5
A, B-30 average	1.15	2.74	.52	6.5	12.0
A, B-40 average	1.03	2.16	.45	5.0	10.5

[a] Values of θ are chosen so that the expected value of $\bar{X}$ is three.

3.3 Testing the Fit of the p_1 Distribution

We have two suggestions for ascertaining whether an observed adjacency matrix X is well represented by the p_1 distribution. The first is the time-honored study of residuals, while the second uses approximate test statistics that we developed elsewhere (Holland and Leinhardt 1975a,b and 1978). In the example in Section 4 we will illustrate how the fitted p_1 distribution can be used for residual analysis. In the remainder of this section we discuss how the tests proposed in Holland and Leinhardt (1975 and 1978) can be used to provide goodness-of-fit tests of p_1.

The natural way to test the fit of an exponential family of distributions is to embed it in a larger family of distributions and perform the corresponding tests. For example, let $p_2(x)$ be a new probability distribution over G having the form

$$\begin{aligned} p_2(x) = \exp\{&\delta Z(x) + \rho m + \theta x_{++} \\ &+ \sum_i \alpha_i x_{i+} + \sum_j \beta_j x_{+j}\} \\ &\times k_2(\delta, \rho, \theta, \{\alpha_i\}, \{\beta_j\}) \,, \end{aligned} \tag{61}$$

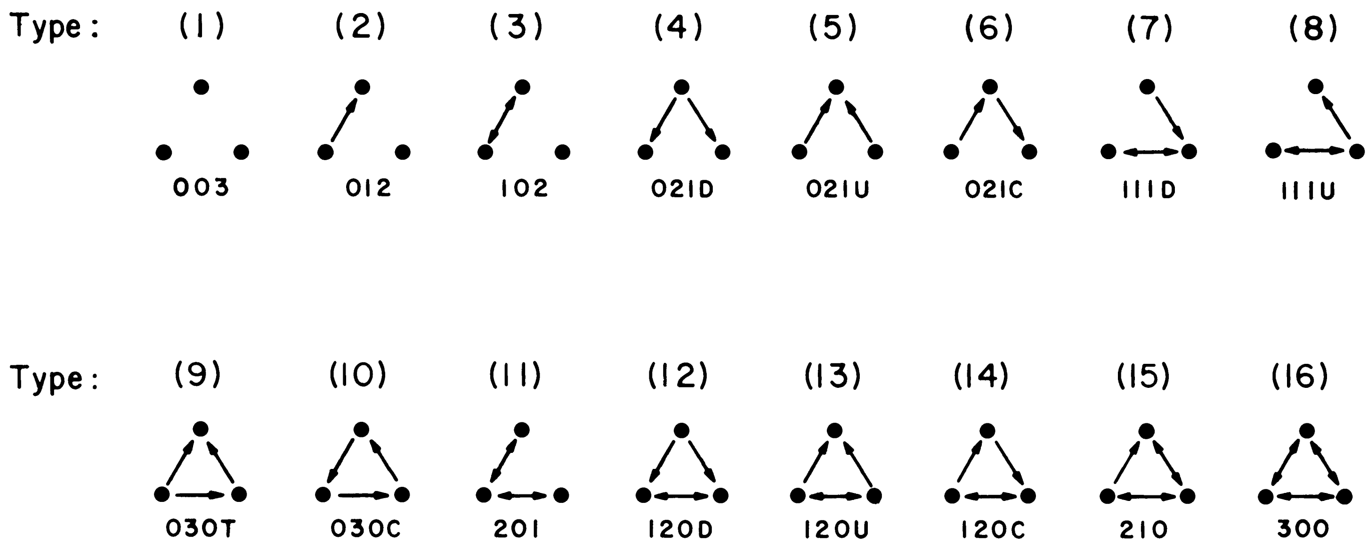
where everything in (61) is just like p_1 in (8), except that δ is a new parameter and $Z(x)$ is a new sufficient statistic (or real-valued function of the matrix x). The p_2 distribution contains the p_1 distribution as a special case, and thus the natural goodness-of-fit test of p_1 is the test of $\delta = 0$ in p_2 (where all the other parameters in p_2 are allowed to vary freely).

The form of p_2 depends on the function $Z(x)$. The choice of $Z(x)$ depends on the type of departure from p_1 that the analyst wishes to be able to detect. There are several considerations in choosing $Z(x)$. For example, any choice of $Z(x)$ that can be expressed as a linear combination of m, $\{x_{i+}\}$, and $\{x_{+j}\}$ will yield a family, p_2, that would be identical to p_1. Another consideration in the choice of $Z(x)$ is the tractability of the resulting p_2 distribution. We have not succeeded in finding a function $Z(x)$ that leads to a tractable p_2 and that uses information from x that is more complicated than the dyads, D_{ij}. Nonetheless, formulating goodness-of-fit tests of p_1 in terms of embedding p_1 in a larger family is useful because, regardless of what $Z(x)$ is, the form of the uniformly most powerful unbiased (UMPU) test of $\delta = 0$ is easy to describe. From Lehmann (1959, p. 134) it follows that the UMPU test of $\delta = 0$ against $\delta \neq 0$ is based on the conditional distribution of $Z(x)$ given M, $\{X_{i+}\}$, and $\{X_{+j}\}$, under the uniform distribution over G. Thus if $Z(x)$ has an approximate normal conditional distribution given M, $\{X_{i+}\}$, and $\{X_{+j}\}$, then the UMPU test of $\delta = 0$ will reject if

$$\tau = \frac{Z(x) - e}{s} \tag{62}$$

is extreme, where e is the conditional mean and s is the conditional standard deviation of $Z(x)$, given M, $\{X_{i+}\}$, and $\{X_{+j}\}$. Thus in order to obtain a goodness-of-fit test of p_1, we need to find a $Z(x)$ that reflects the types of departures from p_1 that interest us and for which the conditional distribution of $Z(x)$, given M, $\{X_{i+}\}$, and $\{X_{+j}\}$, is adequately represented by a normal approximation.

In Holland and Leinhardt (1975 and 1978) we have discussed tests of the form given in (62) where $Z(x)$ is a function of the "triad census" of x. The triad census of x is defined as follows. Each of the $\binom{g}{3}$ distinct triples of nodes defines a triad of the original digraph. There are 16 possible nonisomorphic triads of a digraph. These are displayed in Figure 4. The triad census of x is the 16 vector whose ith entry gives the number of triples of nodes of x of the ith triad type. We have suggested using linear combinations of the counts in a triad census as possible choices for $Z(x)$ because they reflect information



*Triad isomorphism classes are coded by three digits. The first digit indicates the number of reciprocated or mutual pairs (M), the second the number of asymmetric pairs (A), and the third the number of null pairs or pairs without ties (N). Trailing letters distinguish among classes that differ because of orientations of asymmetric pairs. See Holland and Leinhardt (1970).

*Figure 4. The 16 Triad Isomorphism Classes for a Digraph**

in x that goes beyond the dyads. Furthermore, because linear combinations of triad counts are sums, it is relatively easy to calculate means, variances, and covariances. Finally, normal approximations are plausible when the triad frequencies are large. In Holland and Leinhardt (1970) we considered test statistics of the form

$$\tau(w) = \frac{w'T - w'\mu}{\sqrt{w'\Sigma w}} \tag{63}$$

where T is the 16 vector of triad counts of x and w is a weighting vector. In earlier reports we used weighting vectors that yielded $w'T$ equal to the number of intransitivities in x (i.e., i, j, k form an *intransitivity* if $X_{ij} = 1$, $X_{jk} = 1$ but $X_{ik} = 0$). We call this test statistic τ(intran) (Holland and Leinhardt 1971, 1972). In Holland and Leinhardt (1970) μ and Σ were the conditional mean vector and covariance matrix of T given M and $\bar{X}$. In Holland and Leinhardt (1975) we proposed a method that can be used to obtain approximate values for the conditional mean vector and covariance matrix of T given M, $\bar{X}$, V(in), V(out), and the correlation of (X_{i+}, X_{+i}) or COR(out,in). While this method is approximate and does not go all the way to the full conditioning of T on M, $\{X_{i+}\}$, and $\{X_{+j}\}$, it appears to be a useful step in the right direction.

In Holland and Leinhardt (1978) we proposed the test statistic τ^2(max) defined by

$$\tau^2(\max) = \max_w \tau^2(w) \tag{64}$$

where $\tau(w)$ is defined in (63). This test statistic may also be used to test p_1. Instead of loading all the discriminating power of the test in one direction, as $\tau(w)$ does (i.e., that defined by the weighting vector), τ^2(max) is able to detect any sufficiently large departure from p_1 that may be expressed as linear combinations of triad counts. The null distribution for τ^2(max) is chi-squared distributed if the conditional asymptotic normality of T holds. The degrees of freedom for τ^2(max) depend on the level of conditioning. If μ and Σ used in (63) are the conditional moments of T given M, X, V(in), V(out), and COR(out,in), then the degrees of freedom for τ^2(max) are $16 - 1 - 5 = 10$.

At present, $\tau(w)$ and τ^2(max) are the only tools we know of for formally testing the goodness of fit of the unrestricted p_1 distribution to an observed adjacency matrix. Further research is necessary to substantiate and refine our suggestion to use $\tau(w)$ and τ^2(max) in this way. In Section 4 we illustrate the use of $\tau(w)$ and τ^2(max) to test the fit of p_1 to an empirical example.

4. AN EMPIRICAL EXAMPLE

Figure 5 gives the adjacency matrix for friendship data originally gathered by Sampson (1969) in a study of the interpersonal ties among 18 members of a monastery. The matrix in Figure 5 is taken from White, Boorman, and Breiger (1976), who rearranged the rows and columns of X to emphasize blocks of high and low edge density. The out-degrees in Figure 5 are all three or four because White, Boorman, and Breiger used only the top three friendship choices of a complete ranking in which ties were allowed. We have not investigated whether varying the number of top choices alters the conclusions of the analysis.

Table 3 gives some summary information computed from the matrix in Figure 5. It is evident that reciprocation is high—M is nearly three times its null expected value. The in-degree distribution does not seem to be markedly different from the chance prediction because V(in) is only 1.2 times its null expected value. The test statistic τ(intran) is significantly negative (-4.92), indicating a tendency toward a transitive structure. Since there is statistical evidence for both reciprocation and transitivity, we would expect to see a reasonable level of cliquing in these data. This supports the division of these data into blocks found by White, Boorman, and Breiger. The large value of τ^2(max), 38.37, suggests that p_1 does not fit this set of data well. The ratio $(\tau(\text{intrans}))^2/\tau^2(\max)$ is .63, indicating that the tendency towards transitivity and reciprocation accounts for most of the structure detected by τ^2(max). (See Holland and Leinhardt 1971; Davis and Leinhardt 1972; Leinhardt 1972 for discussions of transitivity and cliquing.)

Table 4 gives the fitted expected values of X_{ij} under the p_1 distribution, that is, $\hat{p}_{ij} = \hat{m}_{ij} + \hat{a}_{ij}$, for the adjacency matrix in Figure 5.

Table 5 gives the residuals, $r_{ij} = X_{ij} - \hat{p}_{ij}$, $i \neq j$. In Table 6 we have formed the distribution of the nonzero residuals from Table 5. The left-most column of Table 6 gives the tenths digit for the residuals. These range from 9 down to -4 since the residuals range from .94 down to $-.47$. The second column of Table 6 gives the number of residuals with the specified tenths digit, and the third column gives the corresponding percentage. The bulk of the residuals are negative, corresponding to the fact that X has more zeros than positive values. Six of the positive residuals are .90 or larger. Figure 6 contains coded residuals in which all those .70 or larger are coded "+," those $-.70$ or smaller are coded "$-$" (there are none in this example), and all those between .69 and $-.69$ are left blank. These coded residuals may be interpreted as "unexpected" ties (for +) and "unexpected" nonties (for $-$). They are unexpected in the sense that they are not what the p_1 distribution would predict based on the observed in- and out-degrees and reciprocation. Most of the relational ties in this example are "unexpected" because of the clean-cut pattern of cliquing. The six most unexpected ties (i.e., residuals .90 or greater) are (6,7), (6,11), (7,4), (11,12), (13,14), and (14,12). Except for

Table 3. Some Summary Information for Adjacency Matrix in Figure 5

g	$\bar{X}$	V (out)	V (in)	M	$E(V(in)\mid X_{i+})$	$E(M\mid X_{i+})$	τ(intran)	$\tau^2(max)$
18	3.11	.099	2.99	15	2.54	5.12	−4.92	38.37

	X_{ij}																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	X_{i+}
1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	4
2	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	3
3	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	3
4	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3
5	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3
6	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	3
7	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3
8	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3
9	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	3
10	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	3
11	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	3
12	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	3
13	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	3
14	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	3
15	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	3
16	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	4
17	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	3
18	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	3
X_{+j}	0	6	4	2	4	2	2	6	4	6	2	2	5	1	2	3	2	3	$56 = X_{++}$

Figure 5. Adjacency Matrix From Sampson (1969) As Presented in White, Boorman, and Breiger (1976). Dashed Lines Indicate High and Low Tie-Density Blocks Found by White et al. Left-most Column and Uppermost Row Are the Indices of i and j, Respectively

(6,11), these are all "within-block ties" from the point of view of the blocks identified by White, Boorman, and Breiger (1976), and except for (13, 14) these are all nonreciprocated ties. These support the block structure found by White et al.

Table 7 gives the parameter estimates of ρ, α_i, and β_j for these data. The value of $\hat{\rho} = 3.10$ means that the odds ratio in (17) is 22.2, indicating a 22-fold increase in the odds that $X_{ij} = 1$ when $X_{ji} = 1$ over the value obtained when $X_{ji} = 0$.

Table 4. Fitted Expected Values of X_{ij} From Figure 5 for the p_1 Distribution[a]

	$\hat{p}_{ij} \times 100$																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	$\hat{p}_{i+}$
1	0	47	29	12	29	12	12	47	29	47	12	12	38	05	12	16	12	20	4
2	0	0	23	12	23	12	12	33	23	33	12	12	28	06	12	19	12	18	3
3	0	33	0	12	22	12	12	33	22	33	12	12	28	06	12	18	12	17	3
4	0	34	22	0	22	10	10	34	22	34	10	10	28	05	10	16	10	16	3
5	0	33	22	12	0	12	12	33	22	33	12	12	28	06	12	18	12	17	3
6	0	34	22	10	22	0	10	34	22	34	10	10	28	05	10	16	10	16	3
7	0	34	22	10	22	10	0	34	22	34	10	10	28	05	10	16	10	16	3
8	0	33	23	12	23	12	12	0	23	33	12	12	28	06	12	19	12	18	3
9	0	33	22	12	22	12	12	33	0	33	12	12	28	06	12	18	12	17	3
10	0	33	23	12	23	12	12	33	23	0	12	12	28	06	12	19	12	18	3
11	0	34	22	10	22	10	10	34	22	34	0	10	28	05	10	16	10	16	3
12	0	34	22	10	22	10	10	34	22	34	10	0	28	05	10	16	10	16	3
13	0	33	23	12	23	12	12	33	23	33	12	12	0	06	12	19	12	18	3
14	0	35	21	09	21	09	09	35	21	35	09	09	28	0	09	14	09	15	3
15	0	34	22	10	22	10	10	34	22	34	10	10	28	05	0	16	10	16	3
16	0	43	30	15	30	15	15	43	30	43	15	15	36	08	15	0	15	23	4
17	0	34	22	10	22	10	10	34	22	34	10	10	28	05	10	16	0	16	3
18	0	33	22	11	22	11	11	33	22	33	11	11	27	05	11	18	11	0	3
$\hat{p}_{+j}$	0	6	4	2	4	2	2	6	4	6	2	2	5	1	2	3	2	3	$56 = \hat{p}_{++}$

[a] The decimal points have been left off the entries in the body of the table. The marginal totals have not been so altered.

Table 5. Residuals, $r_{ij} = X_{ij} - \hat{p}_{ij}$, From Figure 5 and Table 4, Multiplied by 100

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	53	71	−12	71	−12	−12	−47	−29	−47	−12	−12	−38	−5	88	−16	−12	−20
2	0	0	77	−12	77	88	−12	−33	−23	−33	−12	−12	−28	−6	−12	−19	−12	−18
3	0	67	0	−12	−22	−12	88	67	−22	−33	−12	−12	−28	−6	−12	−18	−12	−17
4	0	66	78	0	78	−10	−10	−34	−22	−34	−10	−10	−28	−5	−10	−16	−10	−16
5	0	67	−22	88	0	88	−12	−33	−22	−33	−12	−12	−28	−6	−12	−18	−12	−17
6	0	66	−22	−10	−22	0	90	−34	−22	−34	90	−10	−28	−5	−10	−16	−10	−16
7	0	−34	78	90	78	−10	0	−34	−22	−34	−10	−10	−28	−5	−10	−16	−10	−16
8	0	−33	−23	−12	−23	−12	−12	0	77	67	−12	−12	72	−6	−12	−19	−12	−18
9	0	−33	−22	−12	−22	−12	−12	67	0	−33	88	−12	−28	−6	−12	82	−12	−17
10	0	−33	−23	−12	−23	−12	−12	67	77	0	−12	−12	72	−6	−12	−19	−12	−18
11	0	−34	−22	−10	−22	−10	−10	66	78	−34	0	90	−28	−5	−10	−16	−10	−16
12	0	−34	−22	−10	−22	−10	−10	66	−22	66	−10	0	72	−5	−10	−16	−10	−16
13	0	−33	−23	−12	−23	−12	−12	67	−23	67	−12	−12	0	94	−12	−19	−12	−18
14	0	−35	−21	−9	−21	−9	−9	−35	−21	65	−9	91	72	0	−9	−14	−9	−15
15	0	66	−22	−10	−22	−10	−10	−34	−22	−34	−10	−10	72	−5	0	−16	−10	84
16	0	−43	−30	−15	−30	−15	−15	−43	70	−43	−15	−15	−36	−8	85	0	85	77
17	0	−34	−22	−10	−22	−10	−10	−34	−22	66	−10	−10	−28	−5	−10	84	0	84
18	0	−33	−22	−11	−22	−11	−11	−33	−22	67	−11	−11	−27	−5	−11	82	89	0

This estimate of ρ assumes the $\{\alpha_i\}$ and the $\{\beta_j\}$ are not equal to zero. It can be contrasted with an alternative estimate, which assumes the $\{\alpha_i\}$ and $\{\beta_j\}$ equal zero. This alternative estimate derives from Davis's (1977) suggestion that one could summarize an adjacency matrix X by forming the 2-by-2 table of the frequencies of the pairs of nodes $D_{ij} = (X_{ij}, X_{ji})$ for $i \neq j$. His table has the form

		X_{ji}: 1	X_{ji}: 0	Total
X_{ij}	1	$2M$	A	X_{++}
	0	A	$2N$	$g(g-1)-X_{++}$
	Total	X_{++}	$g(g-1)-X_{++}$	$g(g-1)$

(65)

where M, as before, is the number of reciprocated paris, A is the number of nonreciprocated pairs, and $N = \binom{g}{2} - M - A$ is the number of null pairs. Davis prousing contingency table measures of association to obtain measures of particular structural effects for digraph data. His proposed measure of reciprocation is a monotonic function of the log-cross-product ratio in the 2-by-2 table in (65), that is,

$$\hat{\hat{\rho}} = \log_e(4MN/A^2). \tag{66}$$

We have denoted this by $\hat{\hat{\rho}}$ because, under the submodel of p_1 that assumes that $\{\alpha_i\} = \{\beta_j\} = 0$, $\hat{\hat{\rho}}$ is the MLE of

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0		+		+										+			
2		0	+		+	+												
3			0				+											
4			+	0	+													
5				+	0	+												
6						0	+				+							
7			+	+	+		0											
8								0	+				+					
9									0		+					+		
10									+	0			+					
11									+		0	+						
12												0	+					
13													0	+				
14												+	+	0				
15													+		0			+
16									+						+	0	+	+
17																+	0	+
18																+	+	0

Codes

+ = $r_{ij} \geq .70$.
0 = main diagonal

Figure 6. Coded Residuals From Table 5 (X_{ii} = 0 by convention)

Table 6. Distribution of Residuals in Table 5

Tenths Digit	Frequency	Percent
9	6	2%
8	14	5
7	18	6
6	17	6
5	1	0
4	—	—
3	—	—
2	—	—
1	—	—
0	—	—
−0	22	8
−1	126	44
−2	48	17
−3	32	11
−4	5	2
	289	101%

ρ. When the α_i and β_j differ from zero, $\hat{\hat{\rho}}$ may be a misleading measure of reciprocation because it ignores differential attraction among the nodes. Indeed, in the case at hand $\hat{\hat{\rho}} = 2.30$, corresponding to an odds ratio of 10.0. Thus, while these two estimates of ρ both indicate a large value for this parameter, the difference between them illustrates the effect of simultaneously estimating ρ, $\{\alpha_i\}$, and $\{\beta_j\}$.

In Table 7 we have estimated β_1 as $-\infty$ because this is the required value for $p_{+1} = X_{+1} = 0$. The other $\hat{\beta}_j$ have been parameterized so that they sum to zero. In this example, there is a near-monotonic relationship between $\hat{\beta}_j$ and X_{+j}. The exception to monotonicity occurs in the estimate of β for node 16. Although node 16 and node 18 both receive three choices, $\hat{\beta}_{16} = -.25$, while $\hat{\beta}_{18} = 0$. Thus an estimate of the relative attractiveness of nodes 16 and 18 assuming $\rho = 0$ would imply no difference between the two individuals, while an estimate assuming $\rho \neq 0$ implies that node 18 is more attractive than node 16. In other empirical situations the divergence from monotonicity may be more extreme. Comparison of the α_i with the X_{i+} indicates that these are definitely not monotonically related.

These results and those for $\hat{\rho}$ and $\hat{\hat{\rho}}$ illustrate the importance of p_1. Fitting p_1 to data involves the simultaneous estimation of ρ, θ, $\{\alpha_i\}$, and $\{\beta_j\}$. Earlier analyses have at best estimated analogs of ρ and $\{\beta_j\}$ separately. This implicitly assumes that the nonestimated parameters are all zero. The differences we have observed in this empirical example suggest that assuming that parameter values equal zero may be misleading.

The value of the log-likelihood ratio statistic for the test of $\rho = 0$ (i.e., H_1 versus H_3 in Section 3, Eqs. 56 and 58) is LLR = 30.41. When referred to the chi-squared distribution with one degree of freedom, this is statistically significant at the usual levels, as mentioned earlier in the simulation study of Section 3. This supports rejection of H_1 and gives inferential support to the previously indicated evidence of high reciprocation in these data.

Table 8 gives the triad census for this example along with the approximate expected values of the triad census conditional on $\bar{X}$, M, V(out), V(in), and COR(out,in). The value of τ^2(max) is 38.37. This value exceeds the .005 cut-off level of the chi-squared distribution on 10 degrees of freedom. Thus, although the agreement between the observed and expected triad frequencies looks remarkably good, the discrepancy is statistically significant. This demonstrates the difficulty encountered in performing "eyeball" analyses of surpluses and deficits of triads. Given the large value of τ^2(intrans)—that is, 24.21 out of a maximum of 38.37—most of the discrepancies between the observed and expected values in Table 8 are associated with the single degree of freedom given by intransitivity. This is especially hard to "eyeball" because it corresponds to a specific linear combination of the triad frequencies.

Table 7. Estimates of ρ, θ, α_i, and β_j for Data in Figure 4 ($\hat{\rho} = 3.14$, $\hat{\theta} = -2.50$)

i	1	2	3	4	5	6	7	8	9
$\hat{\alpha}$	1.15	−.73	−.30	.22	−.30	.22	.22	−.73	−.30
$\hat{\beta}$	−∞	1.25	.49	−.62	.49	−.62	−.62	1.25	.49
i	10	11	12	13	14	15	16	17	18
$\hat{\alpha}$	−.73	.22	.22	−.53	.49	.22	.48	.22	−.05
$\hat{\beta}$	1.25	−.62	−.62	.89	−1.53	−.62	−.25	−.62	.00

Table 8. Triad Census for Data in Figure 5 With Expected Values Conditional on g, $\bar{X}$, M, V(in), V(out), and COR(out, in)

Triad Type[a]	*Triad Census* (0)	*Expected Value* (*E*)	*O-E*
003	293	307.82	−14.82
012	257	231.16	25.84
102	155	140.53	14.47
021D	7	9.99	−2.99
021U	13	17.35	−4.35
021C	20	23.64	−3.64
111D	27	38.71	−11.71
111U	13	21.05	−8.05
030T	3	2.66	.34
030C	1	.62	.36
201	9	14.42	−5.42
120D	7	1.98	5.02
120U	1	1.00	.00
120C	3	2.48	.52
210	5	2.37	2.63
300	2	.23	1.77

[a] See Figure 4 for this code.

5. A GENERALIZATION

We have developed the p_1 distribution for data on a single relationship observed at one point in time, case (a) of Section 1. Although such data represent by far the most common kind of data studied by social network analysts (Davis and Leinhardt 1972), probably the next most common form consists of one-time observations of multiple relationships, case (e) of Section 1. Together with the increase in theoretical richness that multiple relationship data provide (see, e.g., White, Boorman, and Breiger 1976; and Boorman and White 1976), the frequency with which these data are collected would indicate that extending p_1 to the multiple relationship case is a natural next step. Because of its importance, the straightforwardness of its development, and the way in which it illustrates the utility of p_1, we briefly discuss such an extension here. Of course, extensions of p_1 that incorporate other complications, such as those represented by cases (b) through (d) of Section 1, are also important for social network analysis, but their development here is precluded by space limitations.

5.1 Multiple Relationship Digraph Data

Suppose that two different adjacency matrices are observed for the same set of nodes; denote them x and y, respectively. Then x and y are both g-by-g zero-one matrices and their rows and columns are in correspondence (i.e., x_{ij} and y_{ij} refer to two different relationships between the same two nodes, i and j). There are many examples of multiple relationship (or multiple generator) data in social network research. For example, in studies of friendship it is common to collect information on "dislike" as well as "like." In Sampson's (1969) study of the social relationships among members of a monastery, he reports data on no less than eight different types of relationships. In this section we shall be content to generalize $p_1(x)$ to the "bivariate" case of two adjacency matrices, x and y, since this illustrates the essential features of the general case and allows us to address an important substantive issue, the correlation of x and y.

Suppose, to begin, that X and Y are two random g-by-g adjacency matrices that are statistically independent and that both have p_1 distributions with possibly different parameter values. The joint distribution of X and Y is thus

$$\begin{aligned} P(X = x, Y = y) &= p_1(x)p_1(y) \\ &= \exp\{\rho_1 M(x) + \rho_2 M(y) \\ &\quad + \theta_1 x_{++} + \theta_2 y_{++} \\ &\quad + \sum_i \alpha_{i1} x_{i+} + \sum_i \alpha_{i2} y_{i+} \\ &\quad + \sum_i \beta_{j1} x_{+j} + \sum_j \beta_{j2} y_{+j}\} \times k \end{aligned} \tag{67}$$

where k is the product of the two normalizing constants and $M(x)$ and $M(y)$ are the numbers of mutual or reciprocated pairs in x and y, respectively.[8] Even this simple distribution illustrates an important consideration in the analysis of multivariate digraph data. If we set $\beta_{j1} = \beta_{j2}$ for $j = 1, \ldots, g$ so that the attractiveness parameters are the same for the two random digraphs, then, if the β_j vary widely, the entries in X and Y will exhibit an apparent positive correlation. This is because a node that has a high β_{j1} will have a high β_{j2} and will tend to attract relational ties of both types, X and Y. The apparent correlation between X and Y that is due to similar parameter values is analogous to similar statistical artifacts in other settings—the ecological correlation fallacy, for example.

To introduce a "true" correlation between X and Y, it is convenient to proceed as we did for $p_1(x)$ by considering the pairs (i, j). We may decompose X, Y into $\binom{g}{2}$ vectors

$$D_{ij}^{(2)} = (X_{ij}, X_{ji}, Y_{ij}, Y_{ji}) \tag{68}$$

where $i < j$.

In (68) $D_{ij}^{(2)}$ has a superscript two to remind us that it is the two-relation analog of D_{ij} defined in Section 2. $D_{ij}^{(2)}$ can have any of 16 possible values. It may happen that the relationships X and Y are mutually exclusive in the sense that $X_{ij} = 1$ implies that $Y_{ij} = 0$ and $X_{ij} = 0$ implies that $Y_{ij} = 1$. For example, *like* and *dislike* will often have this property in classroom sociometric studies of friendship. In the case of mutually exclusive relations, $D_{ij}^{(2)}$ has zero probability of taking on certain values (such as (1, 1, 1, 1)).

To generalize $p_1(x)$ to the case of two relationships, we first assume that the vectors $D_{ij}^{(2)}$ are all independent, as we did for D_{ij} in the theoretical development of $p_1(x)$. Then, to specify the joint distribution of X and Y, we need only specify the values of the 16 probabilities that $D_{ij}^{(2)}$ takes on its 16 possible values. We let

$$q_{ij}(t, u, v, w) = P(D_{ij}^{(2)} = (t, u, v, w)) \tag{69}$$

for $i < j$ and $t, u, v, w = 0,1$. We also set

$$I_{ij}(x, y; t, u, v, w) = \begin{cases} 1 & \text{if } x_{ij} = t, x_{ji} = u, y_{ij} = v, y_{ji} = w \\ 0 & \text{otherwise} \end{cases} \tag{70}$$

for $i < j$ and $t, u, v, w = 0, 1$, x and $y \in G$. The $I_{ij}(.)$ functions may be expressed in terms of products of x_{ij}, $1 - x_{ij}$, x_{ji}, and so on. For example,

$$I_{ij}(x, y; 1, 1, 1, 1) = x_{ij}x_{ji}y_{ij}y_{ji} . \tag{71}$$

The joint probability distribution of (X, Y) is given by

$$P(X = x, Y = y) = \prod_{i<j} \prod_{t,u,v,w=0,1} q_{ij}(t, u, v, w)^{I_{ij}(x,y;t,u,v,w)}. \tag{72}$$

Equation (72) is the bivariate version of (13) in the development of the univariate case in Section 2. We may reexpress (72) in the following way that emphasizes its exponential form:

$$\begin{aligned} P(X = x, Y = x) &= \exp\{\sum_{i\neq j} \theta_{1ij}x_{ij} + \sum_{i\neq j} \theta_{2ij}y_{ij} \\ &\quad + \sum_{i<j} \rho_{1ij}x_{ij}x_{ji} + \sum_{i<j} \rho_{2ij}y_{ij}y_{ji} \\ &\quad + \sum_{i\neq 1} \theta_{12ij}x_{ij}y_{ij} + \sum_{i\neq j} \rho_{12ij}x_{ij}y_{ji} \\ &\quad + \sum_{i\neq j} \psi_{1ij}x_{ij}x_{ji}y_{ij} + \sum_{i\neq} \psi_{2ij}x_{ij}y_{ij}y_{ji} \\ &\quad + \sum_{i<j} \psi_{12}x_{ij}x_{ji}y_{ij}y_{ji}\} \\ &\quad \times \prod_{i<j} q_{ij}(0, 0, 0, 0) . \end{aligned} \tag{73}$$

In (73) the θ's, ρ's, and ψ's are the logs of products and ratios of the q_{ij}'s.

Just as we simplified (14) to (21) by placing restrictions on ρ_{ij} and θ_{ij}, so too can we simplify (73) by placing restrictions on the θ's, ρ's, and ψ's. For example, a potentially useful model for (X, Y) that introduces true cor-

[8] Producing random matrices from a bivariate p_1 distribution proceeds along lines similar to those described earlier for the univariate case. We have developed a FORTRAN routine on a DEC-20 computer that produces pairs of random adjacency matrices from a bivariate p_1 distribution with specified parameter values.

relation between X and Y (in 67) without going to the full parameterization in (73) is obtained by making the following restrictions:

$$\theta_{1ij} = \theta_1 + \alpha_{i1} + \beta_{j1} \tag{74}$$

$$\theta_{2ij} = \theta_2 + \alpha_{i2} + \beta_{j2} \tag{75}$$

$$\rho_{1ij} = \rho_1 ; \rho_{2ij} = \rho_2 \tag{76}$$

$$\rho_{12ij} = \rho_{12} ; \theta_{12ij} = \theta_{12} \tag{77}$$

$$\psi_{1ij} = \psi_{2ij} = \psi_{12ij} = 0 . \tag{78}$$

The resulting bivariate distribution for (X, Y) has the following exponential form:

$$\begin{aligned} P(X = x, Y = y) &= p_1(x, y) \\ &= \exp\{\rho_1 M(x) + \rho_2 M(y) + \theta_1 x_{++} + \theta_2 y_{++} \\ &\quad + \sum_i \alpha_{i1} x_{i+} + \sum_i \alpha_{i2} y_{i+} \\ &\quad + \sum_i \beta_{j1} x_{+j} + \sum_j \beta_{j2} y_{+j} \\ &\quad + \rho_{12} R(x, y) + \theta_{12} C(x, y)\} \times K , \end{aligned} \tag{79}$$

where $M(x) = \sum_{i<j} x_{ij} x_{ji}$, $M(y) = \sum_{i<j} y_{ij} y_{ji}$ and $R(x, y) = \sum_{i<j} x_{ij} y_{ji}$, $C(x, y) = \sum_{i \neq j} x_{ij} y_{ij}$, and K in (79) is the normalizing constant. We denoted the distribution in (79) by $p_1(x, y)$ because it is a bivariate version of p_1 in the sense that if $\rho_{12} = \theta_{12} = 0$, then $p_1(x, y) = p_1(x)\, p_1(y)$ as given in (67). Thus a submodel of $p_1(x, y)$ is the case of independent X and Y where each follows the p_1 distribution. This leads naturally to tests of correlation between X and Y that are not confounded by the artifactual correlation introduced by the β's that was described earlier (see Hubert and Baker 1978; and Katz and Powell 1953).

The substantive importance of tests of correlation between digraphs rests on the fact that social network analysts typically assume or hypothesize that the structural properties of one social relationship have implications for the properties of another. Thus Homans (1950), for example, argues that affective ties and interactional ties are positively associated. An investigator of this proposition could proceed by studying the correlation of a group's digraph of friendship relations with its digraph of interaction. Similarly, one might study the proposition that liking and influence are inversely related (French 1956; and Hopkins 1964) by studying the correlation between a group's digraph of friendship relations and its digraph of influence relations.

6. CONCLUDING REMARKS

We believe that the study of statistical models for digraph data is an important area for future statistical research. This paper has concentrated on introducing an approach that is useful for applications in the study of *social* networks. With these beginnings, it is likely that related problems can be identified in other fields of application and that eventually a consistent statistical methodology for analyzing relationship data will be developed, one that possesses the flexibility and completeness of methods that currently exist for analyzing attribute data.

[*Received April 1979. Revised May 1980.*]

REFERENCES

ALBA, R.D. (1973), "A Graph-Theoretic Definition of a Sociometric Clique," *Journal of Mathematical Sociology*, 3, 113–126.

ALBA, R.D., and MOORE, G. (1978), "Elite Social Circles," *Sociological Methods and Research* 7, 167–188.

ALDRICH, H. (1977), "Organization Sets, Action Sets, and Networks: Making the Most of Simplicity," in *Handbook of Organizational Design* (Vol. 1), eds. P.G. Nystrom and W. Starback, New York: Elsevier.

ARABIE, P., BOORMAN, S.A., and LEVITT, P.R. (1978), "Constructing Blockmodels: How and Why," *Journal of Mathematical Psychology*, 17, 21–63.

BALLARD, M., CORMAN, L., GOTTLIEB, J., and KAUFMAN, M.J. (1977), "Improving the Social Status of Mainstreamed Retarded Children," *Journal of Educational Psychology*, 69, October, 605–611.

BARNES, J.A. (1966), "Graph Theory and Social Networks: A Technical Comment on Connectedness and Connectivity," *Sociology*, 3, 215–232.

——— (1969), "Networks and Political Processes," in *Social Networks in Urban Situations*, ed. J.C. Mitchell, Manchester, Great Britain: Manchester University Press, 51–76.

BERSCHEID, E., and WALSTER, E. (1977), *Interpersonal Attraction* (2nd ed.), Reading, Mass.: Addison-Wesley.

BJERSTEDT, A. (1956), *Interpretations of Sociometric Choice Status*, Lund, Sweden: Hakan Ohlssons Boktryckeri.

BOORMAN, S.A. (1975), "A Combinatorial Optimization Model for Transmission of Job Information Through Contact Networks," *Bell Journal of Economics*, 6, 216–249.

BOORMAN, S.A., and WHITE, H.C. (1976), "Social Structure From Multiple Networks II: Role Structures," *American Journal of Sociology*, 81, 1384–1446.

BOTT, E. (1971), *Family and Social Network* (2nd ed.), New York: Free Press.

BREIGER, R. (1976), "Career Attributes and Network Structure: A Blockmodel Study of a Biomedical Research Specialty," *American Sociological Review*, 41, 117–135.

BURNS, E. (1974), "Reliability and Transitivity of Pair-Comparison Sociometric Responses of Retarded and Nonretarded Subjects," *American Journal of Mental Deficiency*, 78, 482–484.

BURT, R. (1980), "Models of Network Structure," in *Annual Review of Sociology* (Vol. 6), Palo Alto: Annual Reviews.

COX, D. (1970), *Analysis of Binary Data*, London: Methuen.

DARROCH, J., and RATCLIFF, R. (1972), "Generalized Iterative Scaling for Log-Linear Models," *Annals of Mathematical Statistics*, 43, 1470–1480.

DAVIS, J.A. (1968), "Statistical Analysis of Pair Relationships: Symmetry, Subjective Consistency and Reciprocity," *Sociometry*, 31, 102–119.

——— (1977), "Sociometric Triads As Multi-variate Systems," *Journal of Mathematical Sociology*, 5, 41–59.

DAVIS, J.A., and LEINHARDT, S. (1972), "The Structure of Positive Interpersonal Relations in Small Groups," in *Sociological Theories in Progress* (Vol. 2), ed. J. Berger, Boston: Houghton-Mifflin.

DE SOLA POOL, I., and KOCHEN, M. (1978), "Contacts and Influence," *Social Networks*, 1, 5–52.

DOREIAN, P. (1974), "On the Connectivity of Social Networks," *Journal of Mathematical Sociology*, 3, 245–258.

DUCK, S. (1977), *The Study of Acquaintance*, Westmead, England: Saxon House, Teakfield.

FENNEMA, M., and SCHIJF, H. (1979), "Analyzing Interlocking Directorates: Theory and Methods," *Social Networks*, 1, 297–332.

FIENBERG, S., and WASSERMAN, S.S. (1979), "Methods for the Analysis of Data From Multivariate Directed Graphs," Technical Report #351, University of Minnesota, School of Statistics.

——— (1981), "Categorical Data Analysis of Single Sociometric Relations," in *Sociological Methodology 1981*, ed. S. Leinhardt, San Francisco: Jossey-Bass, 156–192.

FISCHER, C.S. et al. (1977), *Networks and Places*, New York: Free Press.

FRANK, O. (1978), "Sampling and Estimation in Large Social Networks," *Social Networks*, 1, 91–101.

FREEMAN, L.C. (1968), *Patterns of Local Community Leadership*, Indianapolis: Bobbs-Merrill.

——— (1977), "A Set of Measures of Centrality Based on Betweeness," *Sociometry*, 40, 35–41.

——— (1979), "Centrality in Social Networks: Conceptual Clarification," *Social Networks*, 1, 215–239.

FREIDELL, M.F. (1967), "Organizations As Semilattices," *American Sociological Review*, 32, 46–54.

FRENCH, J.R.P. (1956), "A Formal Theory of Social Power," *Psychological Review*, 63, 181–195.

FRIEDKIN, N.E. (1978), "University Social Structure and Social Networks Among Scientists," *American Journal of Sociology*, 83, 1444–1465.

GALASKIEWICZ, J., and MARSDEN, P.V. (1978), "Interorganizational Resource Networks: Formal Patterns of Overlap," *Social Science Research*, 7, 89–107.

GALASKIEWICZ, J., and WASSERMAN, S.S. (1979), "A Dynamic Study of Change in a Regional Corporate Network," Technical Report #347, University of Minnesota, School of Statistics.

GRANOVETTER, M.S. (1974), *Getting a Job*, Cambridge, Mass.: Harvard University Press.

HALLINAN, M.T. (1978), "The Process of Friendship Formation," *Social Networks*, 1, 193–210.

HARARY, F., NORMAN, R.Z. and CARTWRIGHT, D. (1965), *Structural Models*, New York: John Wiley & Son.

HOLLAND, P.W., and LEINHARDT, S. (1970), "A Method for Detecting Structure in Sociometric Data," *American Journal of Sociology*, 76, 492–413.

——— (1971), "Transitivity in Structural Models of Small Groups," *Comparative Group Studies*, 2, 107–124.

——— (1972), "Some Evidence on the Transitivity of Positive Interpersonal Sentiment," *American Journal of Sociology*, 77, 1205–1209.

——— (1975), "Local Structure in Social Networks," *Sociological Methodology, 1976*, ed. D. Heise, San Francisco: Jossey-Bass.

——— (1977a), "Social Structure As a Network Process," *Zeitschrift für Soziologie*, 6, 386–402.

——— (1977b), "A Dyamic Model for Social Networks," *Journal of Mathematical Sociology*, 5, 5–20.

——— (1978), "An Omnibus Test for Social Structure Using Triads," *Sociological Methods and Research*, 7, 227–256.

——— (eds.) (1979a), *Perspectives on Social Network Research*, New York: Academic Press.

——— (1979b), "Structural Sociometry," in *Perspectives on Social Network Research*, eds. P.W. Holland and S. Leinhardt, New York: Academic Press.

HOMANS, G.C. (1950), *The Human Group*, New York: Harcourt, Brace and World.

HOPKINS, T.K. (1964), *The Exercise of Influence in Small Groups*, Totowa, N.J.: Bedminster.

HUBERT, L.J., and BAKER, F.B. (1978), "Evaluating the Conformity of Sociometric Measurements," *Psychometrika*, 43, 31–41.

HUSTON, T. (ed.) (1974), *Foundations of Interpersonal Attraction*, New York: Academic Press.

JONES, E.E., and GERARD, H.B. (1967), *Foundations of Social Psychology*, New York: John Wiley & Sons.

KATZ, L. (1951), "The Distribution of the Number of Isolates in a Social Group," *Annals of Mathematical Statistics*, 23, 271–276.

KATZ, L., and POWELL, J.H. (1953), "A Proposed Index of Conformitory of One Sociometric Measurement to Another," *Psychometrika*, 18, 249–258.

——— (1955), "Measurement of the Tendency Toward Reciprocation of Choice," *Sociometry*, 18, 403–409.

KATZ, L., and TAGIURI, R., and WILSON, T. (1958), "A Note on Estimating the Statistical Significance of Mutuality," *Journal of General Psychology*, 58, 97–103.

KATZ, L., and WILSON, T.R. (1956), "The Variance of the Number of Mutual Choices in Sociometry," *Psychometrika*, 21, 299–304.

KENNY, D.A., and NASBY, W. (1980), "Splitting the Reciprocity Correlation," *Journal of Personality and Social Psychology*, 38, 249–256.

KINDERMAN, R., and SNELL, J.L. (1980), "Markov Random Fields," *Journal of Mathematical Sociology*, 7, 1–14.

LAUMANN, E.O., and PAPPI, F.U. (1976), *Networks of Collective Action*, New York: Academic Press.

LEHMANN, E.L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley & Sons.

LEIK, R.L., and MEEKER, B.F. (1975), *Mathematical Sociology*, Englewood Cliffs, N.J.: Prentice-Hall.

LEINHARDT, S. (1972), "Developmental Change in the Sentiment Structure of Children's Groups," *American Sociological Review*, 37, 202–212.

——— (ed.) (1977), *Social Networks: A Developing Paradigm*, New York: Academic Press.

LEVINE, J.H. (1972), "The Sphere of Influence," *American Sociological Review*, 37, 14–27.

LIGHT, J.M., and MULLINS, N.C. (1979), "A Primer on Blockmodeling Procedure," in *Perspectives on Social Network Research*, eds. P.W. Holland and S. Leinhardt, New York: Academic Press, 85–118.

LOOMIS, C., and PROCTOR, C. (1951), *Research Methods in Social Relations*, New York: Dryden.

LUCE, R.D. (1950), "Connectivity and Generalized Cliques in Sociometric Group Structure," *Psychometrika*, 15, 169–190.

MOORE, G. (1979), "The Structure of a National Elite Network," *American Sociological Review*, 44, 673–691.

MORENO, J.L. (1934), *Who Shall Survive?*, Washington, D.C.: Nervous and Mental Disease Publishing Co.

MORENO, J.L., and JENNINGS, H.H. (1938), "Statistics of Social Configurations," *Sociometry*, 1, 342–374.

MOXLEY, R.L., and MOXLEY, N.F. (1974), "Determining Point-Centrality in Uncontrived Social Networks," *Sociometry*, 37, 122–130.

NEWCOMB, T.M. (1979), "Reciprocity of Interpersonal Attraction: A Nonconfirmation of a Plausible Hypothesis," *Social Psychology Quarterly*, 42, 299–306.

NOBLE, T. (1970), "Family Breakdown and Social Networks," *British Journal of Sociology*, 21, 135–150.

NOSANCHUK, T. (1963), "A Comparison of Several Sociometric Partitioning Techniques," *Sociometry*, 26, 112–124.

ROGERS, E.M. (1979), "Network Analysis of the Diffusion of Innovations," in *Perspectives on Social Network Research*, eds. P.W. Holland and S. Leinhardt, New York: Academic Press, 137–164.

ROISTACHER, R.C. (1974), "A Review of Mathematical Methods in Sociometry," *Sociological Methods and Research*, 3, 123–171.

SAMPSON, F. (1969), "Crisis in a Cloister," PhD dissertation, Cornell University, Dept. of Sociology.

SCHOFIELD, J., and SAGER, H.A. (1977), "Interracial Behavior in a Desegregated School," paper presented at the American Psychological Association Convention, San Francisco.

SØRENSEN, A.B., and HALLINAN, M.T. (1976), "A Stochastic Model for Change in Group Structure," *Social Science Research*, 5, 43–61.

TOLSDORF, C.C. (1976), "Social Networks, Support, and Coping: An Exploratory Study," *Family Processes*, 407–417.

WARNER, R.M., KENNY, D.A., and STOTO, M. (1979), "A New Round Robin Analysis of Variance for Social Interaction Data," *Journal of Personality and Social Psychology*, 37, 1742–1757.

WASSERMAN, S.S. (1977), "Stochastic Models for Directed Graphs," PhD dissertation, Harvard University, Statistics Dept.

——— (1979), "A Stochastic Model for Directed Graphs With Transition Rates Determined by Reciprocity," in *Sociological Methodology 1980*, ed. K. Schuessler, San Francisco: Jossey-Bass.

——— (1980), "Analyzing Social Networks As Stochastic Processes," *Journal of the American Statistical Association*, 75, 280–294.

WHITE, H.C. (1977), "Probabilities of Homomorphic Mappings From Multiple Graphs," *Journal of Mathematical Psychology*, 16, 121–134.

WHITE, H.C., BOORMAN, S.A. and BREIGER, R.L. (1976), "Social Structure From Multiple Networks, I: Blockmodels for Roles and Positions," *American Journal of Sociology*, 81, 730–79.

WOLFE, A.W. (1978), "The Rise of Network Thinking in Anthropology," *Social Networks*, 1, 53–64.