1

LOCAL STRUCTURE IN SOCIAL NETWORKS

Paul W. Holland
EDUCATIONAL TESTING SERVICE

Samuel Leinhardt
CARNEGIE-MELLON UNIVERSITY

Social scientists commonly use graph theory and network concepts to operationalize theoretical statements about structural regularities in social systems. They are especially appropriate for models of the fine structure of interpersonal relations

This chapter is part of a continuing research series and reports work that is collaborative in every respect. The order of our names on this and our previous reports is alphabetical. National Science Foundation Grants GS-39778 to Carnegie-Mellon University and GJ-1154X2 to the National Bureau of Economic Research, Inc., provided financial support. We are grateful to James A. Davis, J. Richard Dietrich, and Christopher Winship for aid in conducting this research and to Richard Hill for computer programing. This chapter was written when Paul Holland was with the Computer Research Center for Economics and Management Science of the National Bureau of Economic Research, Inc.

where the identification of individuals as nodes and interpersonal relations as edges in a graph is immediate. These concepts have been used to represent both the social structure itself and data on social structure. A basic reference detailing the use of graph theory in modeling social structure is Harary, Norman, and Cartwright (1965). The usual data representation, the sociogram, is described in Moreno (1934).

Numerous theoretical statements about structural systems of social or perceptual behavior have been formalized by their representation in graph or network terms. For example, the cognitive balance theory of Heider (1944) was formalized by Cartwright and Harary (1956), Davis (1967), and Flament (1963). Homans's propositions (1950) about behavior in groups were formalized by Davis and Leinhardt (1972) and Holland and Leinhardt (1971). Radcliff-Brown's (1940) and Nadel's (1957) theoretical statements on kinship and role systems were formalized by White (1963) and Lorrain and White (1972). Such models are global in that they imply that the entire organization of the system can be represented by relatively simple patterns.

Sociometric data have commonly been collected by investigators interested in closed small-scale social systems. In the years since the introduction of sociometric procedures, data have been compiled on a large variety of groups by numerous investigators (see, for example, Davis and Leinhardt, 1972). The complexity of these data has led to the development of various techniques and algorithms for organizing and simplifying them (see, for example, Moreno and others, 1960, and Boyle, 1969). Most of these techniques have focused on the identification of global properties such as cliques, status hierarchies, and communication paths. Thus, there has been a consistency in the development of small-scale social-structural theory and data-analytic procedures in that the emphasis has been on global rather than local regularities in the organization of interpersonal relations.

While the development of theory, models, data collection, and data-analytic procedures has been vigorous in this area, it seems fair to say that our understanding of small-scale social structure has not advanced much beyond the fundamental insights of Moreno, Heider, and Homans. We believe that a major reason for this lack of advancement is the discrepancy between the local level at which the data are collected and the global level at which the models are conceptualized. We propose to bridge this gap by examining local structural properties and determining whether they hold, on the average, across entire social systems. This approach permits the formalization and empirical study of propositions concerning the effect of social organization on individual perception and behavior. Such propositions are quite common in the sociological and social psychological literature, as evidenced in Davis (1963), where 56 major sociological and social psychological propositions are restated in graph-theoretic terms and shown to be statements about the consequences of local structure in interpersonal relations.

In the absence of a statistical methodology for testing empirical tendencies in local structure, even Davis's theoretical reformulation would remain little more than an interesting exercise in the art of formalization. Our purpose here is to propose such a methodology (for an earlier statement of our approach see Holland and Leinhardt, 1970). With this methodology, a variety of propositions concerning local structure in networks of interpersonal relations can be modeled in graph-theoretic terms and tested by determining the discrepancy between an empirically observed structure and the structure that is predicted from lower-order relational tendencies such as dyad frequencies and other marginal constraints. While we emphasize the sociometric interpretation of graphs, such interpretation is not essential to the development or use of the method.

TRIAD CENSUS OF A DIGRAPH

We briefly discuss a few graph-theoretic concepts we use repeatedly. For a detailed discussion of many concepts and results from graph theory, we refer the reader to Harary, Norman, and Cartwright (1965).

The basic mathematical entity that concerns us is the binary directed graph, or digraph. A digraph is a set of nodes and a set of directed lines or edges connecting pairs of nodes. In sociometric choice data, nodes correspond to individual group members, and a directed line connects node i to node j if and only if person i chooses person j according to the sociometric choice criterion employed. The adjective *binary* in the definition of a digraph refers to the added restriction that we consider only whether a line connects i to j. We ignore the possibility that choices may have strengths attached to them or may be of different types. Thus the digraph is the mathematical representation of the simplest form of sociometric choice data—unranked choices on a single criterion.

We denote the number of nodes in a digraph by g—the group size. If the nodes are numbered in some arbitrary way from 1 to g, then we may create a useful matrix representation of the digraph as follows. Let $\mathbf{X}$ be a g by g matrix whose (i, j) entry is

$$X_{ij} = \begin{cases} 1 \text{ if } i \rightarrow j \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

Note that we use $i \rightarrow j$ to mean that there is a directed line from node i to node j; there may also be a directed line from node j to node i, but this possibility is neither implied nor denied by the notation $i \rightarrow j$. The matrix $\mathbf{X}$ is called the adjacency matrix in graph theory and the sociomatrix in sociometry. We use the latter term. The main diagonal of $\mathbf{X}$ corresponds to self-choice in the sociometric context, and for many reasons it is convenient for us to assume that $X_{ii} = 0$. Because self-choice is often disallowed in sociometric data, this restriction is generally not significant. Some applications of digraphs allow self-choice (especially if the nodes correspond to groups of people rather than to individuals), but we ignore this possibility here. The sociomatrix $\mathbf{X}$ is not exactly the same thing as the digraph because it implies that the nodes have been arbitrarily labeled from 1 to g. Strictly speaking, each sociomatrix corresponds to a labeled digraph, whereas two different sociomatrices can represent the same unlabeled digraph. In the latter situation, the two sociomatrices will differ only by a simultaneous row-column permutation. A certain number of labeled digraphs correspond to any unlabeled digraph. These are called the labeled versions of the digraph.

Fundamental to the methodology we develop is the notion of a subgraph of a digraph. If in a digraph we delete some of the nodes and all the lines that either go to or come from the deleted nodes, the resulting entity is a subgraph of the digraph. If we delete all but k of the nodes, then we call the result a k-subgraph.

Since a k-subgraph is also a digraph with k nodes, it may be represented by a k by k sociomatrix, and it has a certain number of labeled versions. There are $\binom{g}{k}$ k-subgraphs in a digraph with g nodes.

Two digraphs with g nodes are said to be isomorphic if they are the same in the sense that they can both be represented by the same sociomatrix. This means that if digraph 1 is represented by sociomatrix $\mathbf{X}$ and digraph 2 is represented by sociomatrix $\mathbf{Y}$, then there is a row-column permutation of $\mathbf{Y}$, let us call it $\mathbf{Z}$, such that as matrices $\mathbf{X} = \mathbf{Z}$. The notion of isomorphism partitions digraphs into isomorphism classes. For example, when $g = 2$, there are three isomorphism classes, or dyads, as illustrated in Figure 1. We use the terms *null, asymmetric,* and *mutual* to describe the three dyad types.

When $g = 3$, there are 16 isomorphism classes, or triads. These are illustrated and named in Figure 2. We have adopted the triad naming convention of Holland and Leinhardt (1970), described in the legend for Figure 2.

When $g \geq 4$, the number of isomorphism classes of digraphs grows rapidly—for $g = 4$ there are 218 and for $g = 5$ there are 9608 (see Harary, 1955). We refer to the isomorphism classes of digraphs with g nodes as digraph types.

In a digraph there are $\binom{g}{3}$ distinct 3-subgraphs formed by selecting each of the possible subsets of three nodes and their corresponding lines. These subgraphs can be classified by their isomorphism type. Let T_u denote the number of these 3-subgraphs of isomorphism type u (where u ranges over the 16 triads given in Figure 2). The triad census $\mathbf{T}$ is the column vector of 16 components given by

$$\mathbf{T}' = (T_{003}, T_{012}, \ldots, T_{300}) \tag{2}$$

We have adopted the following ordering of the components of $\mathbf{T}$ to

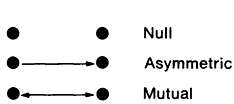

Figure 1. The three isomorphism classes for digraphs with g = 2 (that is, the dyad types).

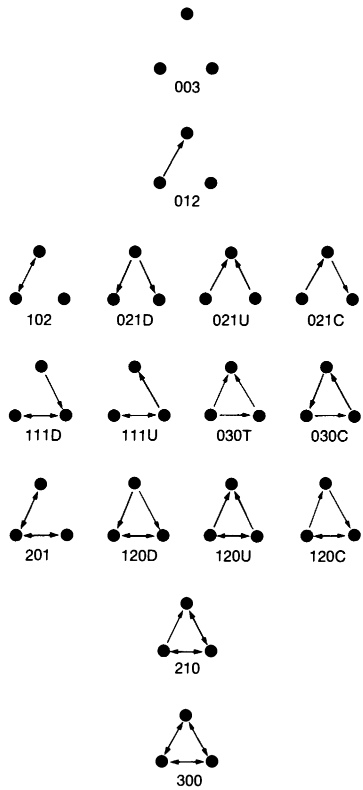

Figure 2. The 16 isomorphism classes for digraphs with g = 3 (that is, the triad types). Triad naming convention: the first digit is the number of mutual dyads; the second digit is the number of asymmetric dyads; the third digit is the number of null dyads; trailing letters further differentiate among triad types.

simplify communication: 003, 012, 102, 021D, 021U, 021C, 111D, 111U, 030T, 030C, 201, 120D, 120U, 120C, 210, 300.

A triad census may be regarded as a way of reducing the entire sociomatrix **X** to a smaller set of 16 summary statistics. When $g \geq 5$ this reduction in information is considerable since **X** contains more than 16 elements of data. In general, knowing the triad census of a digraph does not uniquely determine the digraph. However, as we show below, **T**, contains a surprising amount of useful information about **X**.

A triad census is a special case of the more general concept of a k-subgraph census. Thus we might also consider a dyad census, a tetrad census, or even a pentad census. We show below that a dyad census can be computed from a triad census. Furthermore, there are so many tetrad types (218) and pentad types (9608) that a k-subgraph census for $k \geq 4$ is often more cumbersome than the original sociomatrix. Thus the triad census occupies an important position among k-subgraph censuses in that it is manageable and contains a substantial amount of useful information.

For various random digraphs, the first and second moments of a triad census are readily computed. These computations are illustrated below for a particular random digraph (Corollary 2). The moments permit us to test propositions about average local structure in specific digraphs using the triad census.

Because a triad census **T** is a summation over all 3-subgraphs, the information in **T** is relatively stable and is not significantly affected by a few changes in the lines of the digraph. For this reason, reduction to the triad census is not as affected by sociometric measurement error or masking (see Holland and Leinhardt, 1973) as are methods that focus on specific linkages and individual nodes. This property of the triad census has advantages and disadvantages. On the one hand, conclusions drawn from **T** are likely to be relatively robust, but on the other, if interest focuses only on a few specific nodes or lines, the triad census may be irrelevant.

CONFIGURATIONS IN DIGRAPHS

Davis and Leinhardt (1972) present a procedure in which a version of the triad census is used to test a model of small-scale social structure. Distilling a set of structural propositions from Homans (1950), they examined the triad types to determine which ones were logically inconsistent with the propositions. They felt that a fair interpretation of Homans, formalized in this fashion, was that the inconsistent triads would be empirically rare. They then showed that a social structure combining clusters (Davis, 1967) with a transitive hierarchy (French, 1956) was implied by the propositions. Their empirical analysis of sociometric data lent some support to the model, as did a later analysis by Davis (1970). However, a reexamination of this model revealed that it and most other structural models of interpersonal affect assumed that the affective choice relation is transitive (Holland and Leinhardt, 1971). The models could all be expressed as transitive graphs (t-graphs) upon which some additional constraints had been applied. Transitivity, however, is not a characteristic of a triad. Rather, it is a property of a lower-order configuration that is contained in varying degree by some triad types. For example, assume that the triad type 300 represents the sociogram of a three-person group (see Figure 2 for triad names). This type is transitive from the point of view of each member. That is, the transitive condition, $i \rightarrow j$, $j \rightarrow k$, $i \rightarrow k$, holds for all six permutations of i, j, k. But, in the 120C triad type, the transitivity condition holds only once, while it is contradicted, that is, $i \rightarrow j, j \rightarrow k$, *and not* $i \rightarrow k$, twice. Thus, if Homans's propositions are in fact statements about the propensity for affective choice to tend toward transitivity and to avoid intransitivity, then a failure to recognize the complex nature of triads may yield erroneous conclusions. Indeed, our reanalysis of Davis's (1970) results yielded a high level of confirmation for the t-graph model (see Holland and Leinhardt, 1971).

Reasons for Considering Configurations

It is useful to pursue the notion of transitivity in interpersonal affect because it demonstrates the utility of thinking of structural propositions in terms of configurations of social relations, an idea central to the method of this chapter. Consider, for example, the statement by Heider (1958, p. 206) in regard to positive interpersonal sentiment in triadic situations: "In the *p-o-x* triad, the case of three positive relations may be considered psychologically ... transitive." For Heider, transitivity defines cognitive balance in affective situations. His conceptualization of balanced psychological structures is configurational and not triadic in the sense of, say, Davis and Leinhardt (1972), in that Heider emphasizes transitivity from the point of view of p, the perceiver, and not necessarily from the simultaneous points of view of the other entities.

In an alternative to the t-graph model, Mazur (1971, p. 308) puts forth another structural proposition. He argues that interpersonal affect data can be explained by the proposition: "Friends are likely to agree, and unlikely to disagree; close friends are very likely to agree, and very unlikely to disagree." If one assumes that an asymmetric pair represents "friends" while a mutual pair represents "close friends," then Mazur's proposition amounts to a set of hypotheses concerning the relative empirical frequency of simultaneous choice of individuals by "friends" and by "close friends." Since the third individual's choices are not at issue, Mazur's proposition, like Heider's, is not about triads but refers instead to configurations.

An examination of Davis (1963) reveals numerous other structural propositions that can be formalized in terms of configurations, but these may involve more than three nodes. The concepts of cross-pressures (Berelson and others, 1954), homophily-heterophily (Lazarsfeld and Merton, 1954), structural balance (Cartwright and Harary, 1956), distributive justice (Homans, 1961), cliques (Homans, 1950; Lazarsfeld and Merton, 1954; Lipset and others, 1956), innovation in ideas (Katz and Lazarsfeld, 1955), attitude change (Homans, 1950), and conflict (Coleman, 1975), as reformulated by Davis, are all propositions about configurational tendencies in local structure.

Structure of Configurations

Configurations and subgraphs are similar except that in a configuration only some of the lines between a subset of nodes are of interest. We shall try to make this vague idea precise. We introduced above the concept of labeled and unlabeled subgraphs of a digraph and the concept of a sociomatrix, **X**. For any 3-subgraph we may construct a 3 by 3 sociomatrix such as this one:

$$\begin{pmatrix} - & X_{ij} & X_{ik} \\ X_{ji} & - & X_{jk} \\ X_{ki} & X_{kj} & - \end{pmatrix} \tag{3}$$

where i, j, and k are three distinct nodes. There are $2^6 = 64$ possible zero-one sociomatrices of the form (3), and they correspond to the set of all possible labeled three-node digraphs. These in turn are the labeled versions of the 16 nonisomorphic unlabeled three-node digraphs. It is easier to illustrate how to construct a configuration than to give a precise definition of this concept. We begin with the pairs of subscripts that appear on the entries of (3). The first step is to select a subset of these ordered pairs of subscripts. For example, to define the configuration that corresponds to intransitivity, we select these three ordered pairs: ij, jk, ik. The order of the pairs indicates the direction of the relation. Next, associate a 0 or a 1 with each ordered pair of subscripts that has been selected. For example, for intransitivity we associate a 1 with ij, another 1 with jk, and a 0 with ik. These correspond to $i \rightarrow j$, $j \rightarrow k$, *and not* $i \rightarrow k$. For convenience these items can be arranged in a matrix as follows:

$$\begin{pmatrix} ij & jk & ik \\ 1 & 1 & 0 \end{pmatrix} \tag{4}$$

The first row of the matrix that describes a configuration is the reading rule for the configuration. The second row defines the type of configurations. Two different configurations can have the same reading rule but be of different types. For example, the configuration that corresponds to transitivity has the same reading rule as (4) but a different type. It is given by

$$\begin{pmatrix} ij & jk & ik \\ 1 & 1 & 1 \end{pmatrix} \tag{5}$$

It is convenient to have a picture for a configuration. We adopt the following conventions for drawing them. For an ordered pair of nodes ab: (i) if ab is not in the reading rule, then no arrow is drawn from a to b; (ii) if ab is in the reading rule and has a 1 associated with it, then a solid arrow is drawn from a to b; (iii) if ab is in the reading rule and has a 0 associated with it, then a dashed arrow is drawn from a to b. For example, matrix (4) can be represented by Figure 3.

Another example of a configuration comes from Mazur's proposition about agreement among friends. The reading rule for configurations that portray situations in which friends agree or

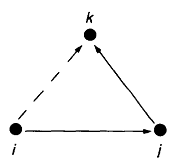

Figure 3. Pictorial representation of the configuration defined by (4).

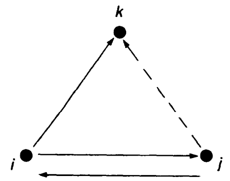

Figure 4. Pictorial representation of configurations given by (6) and (7).

disagree in their choices of a third individual is ij, ji, ik, jk. If close friends are pairs of individuals who choose one another, then disagreement among close friends can be represented by the following configuration matrix:

$$\begin{pmatrix} ij & ji & ik & jk \\ 1 & 1 & 1 & 0 \end{pmatrix} \tag{6}$$

It is obvious that this matrix is equivalent to the matrix

$$\begin{pmatrix} ij & ji & ik & jk \\ 1 & 1 & 0 & 1 \end{pmatrix} \tag{7}$$

The pictorial representation of (6) and (7) is given in Figure 4.

While we have only discussed configurations that involve three nodes, that is, 3-configurations, it is clear that three is not an essential part of the concept. In general, one may have reading rules that involve k distinct nodes, and these result in various types of k-configurations. If a 3-configuration has a reading rule that involves all six of the ordered pairs ij, ji, ik, ki, jk, kj, then it is equivalent to a labeled triad.

The main reason for considering configurations is that they are a more refined set of concepts than triads. A single triad may contain many different configurations. Furthermore, many sociological and social psychological propositions about networks make predictions about configurations rather than about triads. However, we show below that the triad census can be used to enumerate all 3-configurations. This use, along with other known properties of the triad census, is what makes it a basic tool.

LINEAR COMBINATIONS OF TRIAD FREQUENCIES

Once we consider a vector such as the 16 triad frequencies that make up a triad census **T**, it is natural to consider any linear combination of the elements of the vector. We use the vector notation

$$\mathbf{l}'\mathbf{T} = \sum_u l_u T_u \tag{8}$$

where u is always assumed to run over the triad types given in Figure 2.

Dyads from Triads

A simple property of a triad census $\mathbf{T}$ is that it determines how many nodes there are in a digraph. $\mathbf{T}$ has this property because the total number of triads in a digraph with g nodes is $\binom{g}{3}$, and furthermore, if $\mathbf{e}' = (1, 1, \ldots, 1)$, then

$$\mathbf{e}'\mathbf{T} = \sum_u T_u = \binom{g}{3} \tag{9}$$

Thus, to find g we merely compute $\binom{g}{3}$ from (9) and solve for g. Equation (9) is a simple and yet important example of a linear combination of triad frequencies. Let M, A, and N denote the number of mutual, asymmetric, and null dyads in the digraph. Furthermore, let m_u, a_u, and n_u denote the number of mutual, asymmetric, and null dyads that are contained in a triad of type u. Since every dyad in a digraph is contained in exactly $g - 2$ triads, it is easy to see that

$$\mathbf{m}'\mathbf{T} = \sum_u m_u T_u = (g - 2)\, M \tag{10}$$

$$\mathbf{a}'\mathbf{T} = \sum_u a_u T_u = (g - 2)\, A \tag{11}$$

and

$$\mathrm{n}'\mathbf{T} = \sum_u n_u T_u = (g - 2)\, N \tag{12}$$

If we regard g as given, then M, A, and N can be written as linear combinations of the triad frequencies where the weights are $m_u/(g - 2)$, $a_u/(g - 2)$, and $n_u/(g - 2)$. The weights used for enumerating M, A, and N are given in Table 1.

TABLE 1
Selected Weighting Vectors I.

Triad Types	e	m_u	a_u	n_u	c_u	$b_{in,u}$	$b_{out,u}$	Transitivity	Intransitivity	Close Friends Disagreeing
003	1	0	0	3	0	0	0	0	0	0
012	1	0	1	2	1	0	0	0	0	0
102	1	1	0	2	2	0	0	0	0	0
021D	1	0	2	1	2	0	1	0	0	0
021U	1	0	2	1	2	1	0	0	0	0
021C	1	0	2	1	2	0	0	0	1	0
111D	1	1	1	1	3	1	0	0	1	0
111U	1	1	1	1	3	0	1	0	1	1
030T	1	0	3	0	3	1	1	1	0	0
030C	1	0	3	0	3	0	0	0	3	0
201	1	2	0	1	4	1	1	0	2	2
120D	1	1	2	0	4	2	1	2	0	0
120U	1	1	2	0	4	1	2	2	0	0
120C	1	1	2	0	4	1	1	1	2	1
210	1	2	1	0	5	2	2	3	1	1
300	1	3	0	0	6	3	3	6	0	0
$\sum_u l_u T_u$	$\binom{g}{3}$	$(g-2)\,M$	$(g-2)\,A$	$(g-2)\,N$	$(g-2)\,C$	B_{in}	B_{out}			

Since the total number of directed lines (or choices) C is given by

$$C = 2M + A \tag{13}$$

it is clear that C, too, can be expressed as a linear combination of the triad frequencies. The weights for C are given in Table 1.

In- and Out-Degrees from Triads

The row and column sums of a sociomatrix $\mathbf{X}$ will be denoted by $\{X_{i+}\}$ and $\{X_{+j}\}$, respectively. In the sociometric context, X_{i+} is the number of choices made by individual i, while X_{+j} is the number of choices received by individual j. X_{i+} and X_{+i} are also called the out-degree and the in-degree, respectively, of node i. In this notation C is determined by (13) and by

$$C = X_{++} = \sum_i X_{i+} = \sum_j X_{+j} \tag{14}$$

It is sometimes convenient to summarize a set of numbers by their mean and variance. We now do this for the two sets of degrees $\{X_{i+}\}$ and $\{X_{+j}\}$. From (14) we see that the mean in-degree and the mean out-degree are the same and are given by

$$\bar{X} = C/g \tag{15}$$

Since C can be expressed as a linear combination of triad frequencies, it follows that $\bar{X}$ can also be so expressed. More interesting is the fact that the variances of the in-degrees and of the out-degrees can be determined by both C and linear combinations of the triad frequencies. Because this fact is useful to us but not widely known, we include a proof of it now. We denote the variance of the $\{X_{i+}\}$ by S^2_{out} and of the $\{X_{+j}\}$ by S^2_{in} where

$$S^2_{\text{out}} = \frac{1}{g}\sum_i (X_{i+} - \bar{X})^2 \tag{16}$$

and

$$S^2_{\text{in}} = \frac{1}{g}\sum_j (X_{+j} - \bar{X})^2 \tag{17}$$

The proof is divided into two parts:

LEMMA 1. *Let*

$$B_{\text{in}} = \sum_i \sum_{j<k} X_{ji} X_{ki}$$

and

$$B_{\text{out}} = \sum_i \sum_{j<k} X_{ij} X_{ik}$$

then

$$S^2_{\text{in}} = (2/g)\, B_{\text{in}} - \bar{X}(\bar{X} - 1) \tag{18}$$

$$S^2_{\text{out}} = (2/g)\, B_{\text{out}} - \bar{X}(\bar{X} - 1) \tag{19}$$

where $\bar{X}$ *is defined in* (15) *and* (14).

Proof. Since the proofs of (18) and (19) are nearly identical, we only prove (18).

$$S^2_{\text{in}} = \frac{1}{g} \sum_i (X_{+i} - \bar{X})^2 = \frac{1}{g} \sum_i (X_{+i})^2 - (\bar{X})^2 \tag{20}$$

But

$$\begin{aligned}
\sum_i (X_{+i})^2 &= \sum_{i,j,k} X_{ji} X_{ki} \\
&= \sum_i \sum_{j<k} X_{ji} X_{ki} + \sum_i \sum_{k<j} X_{ji} X_{ki} + \sum_i \sum_j X^2_{ji} \qquad (21) \\
&= 2 \sum_i \sum_{j<k} X_{ji} X_{ki} + \sum_i \sum_j X_{ji}
\end{aligned}$$

In (21) we made use of the fact that

$$X^2_{ij} = X_{ij} \tag{22}$$

Hence we have

$$\frac{1}{g} \sum_i (X_{+i})^2 = \frac{2}{g} B_{\text{in}} + \bar{X} \tag{23}$$

Putting (23) into (20) completes the proof. ‖

The next lemma shows that both B_{in} and B_{out} may be ex-

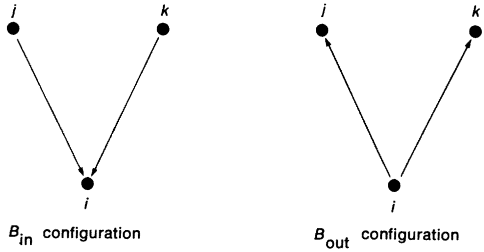

Figure 5. Pictorial representation of B_{in} and B_{out} configurations.

pressed as linear combinations of the triad frequencies so that together with Lemma 1 it proves that S^2_{in} and S^2_{out} are determined by the triad frequencies. Before we state Lemma 2, we need to define two special configurations, which we call B_{in} and B_{out} configurations.

A B_{in} configuration is defined by the matrix

$$\begin{pmatrix} ji & ki \\ 1 & 1 \end{pmatrix} \tag{24}$$

while a B_{out} configuration is defined by the matrix

$$\begin{pmatrix} ij & ik \\ 1 & 1 \end{pmatrix} \tag{25}$$

In Figure 5 we illustrate B_{in} and B_{out} configurations using the conventions of Figures 3 and 4.

It is easy to see that B_{in} gives the number of B_{in} configurations and that B_{out} gives the number of B_{out} configurations. Lemma 2 shows that B_{in} and B_{out} can also be computed from the triad census.

LEMMA 2. *B_{in} and B_{out} are given by*

$$B_{\text{in}} = \sum_u b_{\text{in},u} T_u \tag{26}$$

$$B_{\text{out}} = \sum_u b_{\text{out},u} T_u \tag{27}$$

where $b_{\text{in},u}$ and $b_{\text{out},u}$ are given in Table 1.

Proof. Again, we prove only (26). Every B_{in} configuration in the graph is contained in exactly one 3-subgraph. Hence to count B_{in} configurations, we only need to weight each triad frequency by the number of B_{in} configurations contained in that triad type. To finish the proof, we observe that these weights are exactly the values of $b_{\text{in},u}$ given in Table 1. ||

Counting Configurations from Triads

The proof of Lemma 2 uses the fact that the number of 3-configurations of certain types can be counted using only the triad census **T**. This leads us to consider counting all the possible types of 3-configurations using only the triad census. Evidently, this enumeration requires no new ideas since any configuration that involves three nodes is contained in exactly one 3-subgraph. Thus, to count the number of 3-configurations of a given type that arise in a digraph, we need to find out only how many of these are in each of the 16 triad types and then sum the correspondingly weighted triad frequencies. If a configuration involves only two nodes, the same rule applies except that each 2-configuration is contained in $g - 2$ triads, and hence we must divide the weighted sum of triad frequencies by $g - 2$. This is why the factor $g - 2$ arises when we count dyad types from the triad frequencies.

As an example, suppose we wish to find the number of intransitive configurations in a given digraph. An intransitive configuration is a 3-configuration and can therefore be enumerated by the above rule. The weights used for counting intransitive as well as transitive configurations from **T** are given in Table 1 (p. 13).

As a final example of counting a 3-configuration using **T**, we consider the one given in (5), (6), and Figure 4. Each weight given in Table 1 for this configuration (the column "Close Friends Disagreeing") is found by counting the number of mutual dyads in the given triad type for which only one member of the mutual pair chooses the other triad member.

The discussion above shows why the linear combinations of triad frequencies are important. They give additional information about the graph that may not be obvious from the triad census itself but is implied by it. Thus the set of all linear combinations of the triad frequencies is a natural extension of the triad census. While not all linear combinations of triad frequencies have interpretations in terms of counting configurations, many do, and the concept of configurations is a key to understanding the set of all linear combinations of the triad frequencies.

DISTRIBUTION OF A RANDOM SUBGRAPH CENSUS

Various notions of random digraphs have statistical utility in the analysis of sociometric data (see Holland and Leinhardt, 1970, and Appendixes A and B for further discussion). When we obtain a triad census from a sociomatrix, it is also useful to know what triad census we would expect from a random digraph. While the exact probability distribution of a triad census from a random digraph is complicated, when there is a sufficient number of nodes (say, at least 10) and when certain other nondegeneracies obtain (such as a sufficient number of dyads of each type), a random triad census has an approximate multivariate normal distribution. (See Holland and Leinhardt, 1970, for some simulation results.) In this section, we give formulas for the means, variances, and covariances of the components of $\mathbf{T}$ for a general class of random digraphs. We also give special results for a particular class of random digraphs we have found useful.

Moments of a Subgraph Census

We begin by considering the number of k-subgraphs of a given digraph that are of a particular isomorphism class. This added generality is justifiable on the grounds that it entails no additional difficulties and that the general results for k-subgraphs may be more useful for related problems. We then specialize to triads.

Let u and v denote two isomorphism classes for k-subgraphs. In the triad case, the range of values for u and v is given in Figure 2. For k-subgraphs where $k \geq 4$, the possibilities are numerous; some are given in Moon (1968).

In Theorem 1, below, we give formulas for the means, variances, and covariances for the number of k-subgraphs of types u and v for a random digraph generated by a completely general stochastic mechanism. In Corollary 1 we specialize this result to the case of a triad census, and in Corollary 2 we specialize even further to the moments of a triad census where the random digraph is of the special variety used in Holland and Leinhardt (1970).

We need some notation. Let K and L be subscripts that refer to the $\binom{g}{k}$ distinct k-subgraphs of a given digraph. Thus, we speak of K being of a particular isomorphism class and so on.

We next define the indicator variables $H_K(u)$ by

$$H_K(u) = \begin{cases} 1 \text{ if } K \text{ is of isomorphism class } u \\ 0 \text{ otherwise} \end{cases} \tag{28}$$

The number of k-subgraphs of the given digraph that are of type u, H_u, is then given by the sum over K of $H_K(u)$, that is,

$$H_u = \sum_K H_K(u) \tag{29}$$

Since we provide formulas that hold for a wide class of random digraphs, we need a notation for various probabilities that arise in the calculations and that must be computed explicitly for any particular application of the general results. Thus we define

$$p_K(u) = P\{K \text{ is of type } u\} = P\{H_K(u) = 1\} \tag{30}$$

$$\begin{aligned} p_{K,L}(u, v) &= P\{K \text{ is of type } u \text{ and } L \text{ is of type } v\} \\ &= P\{H_K(u) = 1 \text{ and } H_L(v) = 1\} \end{aligned} \tag{31}$$

We also need a notation for certain average probabilities that can be computed from $p_K(u)$ and $p_{K,L}(u, v)$. The first of these is easy since it is just the average value of $p_K(u)$ over all values of K, that is,

$$\bar{p}(u) = \frac{1}{\binom{g}{k}} \sum_K p_K(u) \tag{32}$$

In order to define the other average probabilities that arise, we need to examine the relationship between two k-subgraphs carefully.

Let

$$|K \cap L| = \text{the number of nodes that } K \text{ and } L \text{ have in common} \tag{33}$$

The possible values for $|K \cap L|$ are $0, 1, \ldots, k$. When $|K \cap L| = 0$, K and L are disjoint, while when $|K \cap L| = k$, K and L are identical. In general, there are

$$\binom{g}{k}\binom{g-k}{k-j}\binom{k}{j} \tag{34}$$

pairs of k-subgraphs of a digraph on g nodes for which $|K \cap L| = j$. Now we define the average value of $p_{K,L}(u, v)$ over all choices of K and L such that $|K \cap L| = j$ as

$$\bar{p}_j(u, v) = \frac{1}{\binom{g}{k}\binom{g-k}{k-j}\binom{k}{j}} \sum_{|K \cap L| = j} p_{K,L}(u, v) \tag{35}$$

It should be emphasized that the various probabilities defined in (30), (31), (32), and (35) must be calculated explicitly in order to apply the general results to particular cases. Thus, we use Theorem 1 to specify what probabilities must be computed and combined to get the moments of a subgraph census.

THEOREM 1. *Using the notation given above and assuming that a random digraph is generated by some stochastic mechanism, the first and second moments of H_u defined in* (29) *are given by*

$$E(H_u) = \binom{g}{k}\bar{p}(u) \tag{36}$$

$$\begin{aligned} \operatorname{var}(H_u) &= \binom{g}{k}\bar{p}(u)\,(1 - \bar{p}(u)) \\ &\quad + \binom{g}{k}\sum_{j=0}^{k-1}\binom{g-k}{k-j}\binom{k}{j}[\bar{p}_j(u, u) - (\bar{p}(u))^2] \end{aligned} \tag{37}$$

$$\begin{aligned} \operatorname{cov}(H_u, H_v) &= -\binom{g}{k}\bar{p}(u)\bar{p}(v) \\ &\quad + \binom{g}{k}\sum_{j=0}^{k-1}\binom{g-k}{k-j}\binom{k}{j}[\bar{p}_j(u, v) - \bar{p}(u)\bar{p}(v)] \end{aligned} \tag{38}$$

Proof. First we prove (36).

$$E(H_u) = E\sum_K H_K(u) = \sum_K E(H_K(u))$$

but $H_K(u)$ is an indicator variable, so that its expected value is merely its probability of being one; thus

$$E(H_u) = \sum_K P\{H_K(u) = 1\} = \binom{g}{k}\bar{p}(u)$$

proving (36).

Since formula (37) is the special case of formula (38) for $u = v$, we prove only (38).

$$\begin{aligned}\text{cov}(H_u, H_v) &= \text{cov}\left(\sum_K H_K(u), \sum_L H_L(v)\right)\\ &= \sum_{K,L} \text{cov}(H_K(u), H_L(v))\end{aligned}$$

But

$$\begin{aligned}\text{cov}(H_K(u), H_L(v)) &= P\{H_K(u) = 1 \quad \text{and} \quad H_L(v) = 1\}\\ &- P\{H_K(u) = 1\}P\{H_L(v) = 1\}\\ &= p_{K,L}(u,v) - p_K(u)p_L(v)\end{aligned}$$

thus

$$\begin{aligned}\text{cov}(H_u, H_v) &= \sum_{K,L} p_{K,L}(u,v) - \sum_{K,L} p_K(u)p_L(v)\\ &= \sum_{j=0}^{k} \sum_{|K\cap L|=j} p_{K,L}(u,v) - \left[\binom{g}{k}\bar{p}(u)\right]\left[\binom{g}{k}\bar{p}(v)\right] \qquad (39)\\ &= \sum_{j=0}^{k} \binom{g}{k}\binom{g-k}{k-j}\binom{k}{j}\bar{p}_j(u,v) - \binom{g}{k}^2 \bar{p}(u)\bar{p}(v)\end{aligned}$$

We now use the following fact about binomial coefficients:

$$\sum_{j=0}^{k} \binom{g-k}{k-j}\binom{k}{j} = \binom{g}{k} \qquad (40)$$

Using (40) we may rewrite (39) as follows:

$$\begin{aligned}\operatorname{cov}(H_u, H_v) &= \binom{g}{k}\sum_{j=0}^{k}\binom{g-k}{k-j}\binom{k}{j}\bar{p}_j(u,v)\\ &-\binom{g}{k}\sum_{j=0}^{k}\binom{g-k}{k-j}\binom{k}{j}\bar{p}(u)\bar{p}(v)\\ &=\binom{g}{k}\sum_{j=0}^{k}\binom{g-k}{k-j}\binom{k}{j}[\bar{p}_j(u,v)-\bar{p}(u)\bar{p}(v)]\end{aligned} \tag{41}$$

We get (38) from (41) and the observation that if $u \neq v$ then $\bar{p}_k(u,v) = 0$ since a k-subgraph can only be of one isomorphism type. We get (37) from (41) and the fact that if $u = v$ then $\bar{p}_k(u,u) = \bar{p}(u)$.

Moments of a Triad Census

The following corollary follows immediately from Theorem 1 in that it specializes it to $k = 3$—the triad census. It is almost exactly the same as Theorem 1 of Holland and Leinhardt (1970) except that it allows a more general stochastic mechanism to generate the random digraph.

COROLLARY 1. *Under the assumptions of Theorem* 1, *the first and second moments of a triad census* T *are given by*

$$E(T_u) = \binom{g}{3}\bar{p}(u) \tag{42}$$

$$\begin{aligned}\operatorname{var}(T_u) &= \binom{g}{3}\bar{p}(u)(1-\bar{p}(u))\\ &+\binom{g}{3}\sum_{j=0}^{2}\binom{g-3}{3-j}\binom{3}{j}[\bar{p}_j(u,u)-(\bar{p}(u))^2]\end{aligned} \tag{43}$$

$$\begin{aligned}\operatorname{cov}(T_u, T_v) &= -\binom{g}{3}\bar{p}(u)\bar{p}(v)\\ &+\binom{g}{3}\sum_{j=0}^{2}\binom{g-3}{3-j}\binom{3}{j}\\ &\quad[\bar{p}_j(u,v)-\bar{p}(u)\bar{p}(v)]\end{aligned} \tag{44}$$

In Holland and Leinhardt (1970) we derived some of the relevant probabilities for a particular random digraph that appears in Davis and Leinhardt (1972). Since we have discussed these derivations extensively elsewhere, we shall give only a description of the random digraph and the complete tables of the relevant probabilities.

The U | *MAN* random digraph distribution is the probability distribution on the set of all digraphs with g nodes that makes all digraphs with given values of M, A, and N (defined in (10), (11) and (12)) equally likely. In other words, it is the uniform distribution on the set of all labeled digraphs having given values of M, A, and N—hence the notation U | *MAN*. One way to generate random digraphs from this distribution is as follows. First randomly allocate M mutual pairs to the $\binom{g}{2}$ possible pairs. Next randomly allocate A asymmetric pairs to the remaining $\left(\binom{g}{2} - M\right)$ pairs, and then randomly and independently orient the asymmetric pairs. The result is a random labeled digraph with the given values of M, A, and N. All such digraphs are equally likely to be generated.

A random digraph with the U | *MAN* distribution possesses certain properties that simplify the calculation of $p_K(u)$, $p_{K,L}(u, v)$, $\bar{p}_j(u, v)$ and $\bar{p}(u)$. The U | *MAN* distribution on the set of digraphs is homogeneous in the sense that it is invariant under permutations of the labels given to the nodes. Because of this property, $p_K(u)$ does not depend on K so that we have

$$\begin{aligned} p_K(u) &= \bar{p}(u) = p(u) \\ &= P\{\text{triad involving nodes } 1, 2, 3 \text{ is of type } u\} \end{aligned} \tag{45}$$

Furthermore, the probabilities $p_{K,L}(u, v)$ depend only on u, v, and $|K \cap L|$ so that if $|K \cap L| = j$, then

$$p_{K,L}(u, v) = \bar{p}_j(u, v) = p_j(u, v) \tag{46}$$

Finally, it is also easy to see that for the U | *MAN* distribution

$$p_0(u, v) = p_1(u, v) \tag{47}$$

Corollary 2 summarizes the above discussion.

COROLLARY 2. *For a triad census* **T** *from a random digraph* *with the U/MAN distribution, the means, variances, and covariances are given by*

$$E(T_u) = \binom{g}{3} p(u) \tag{48}$$

$$\begin{aligned} \operatorname{var}(T_u) = &\binom{g}{3} p(u)(1 - p(u)) + \binom{g}{3}^2 \left(\frac{g+7}{g-2}\right) [p_0(u,u) \\ &- (p(u))^2] \\ &+ 3\binom{g}{3}(g-3)[p_2(u,u) - (p(u))^2] \end{aligned} \tag{49}$$

$$\begin{aligned} \operatorname{cov}(T_u, T_v) = &-\binom{g}{3} p(u)p(v) + \binom{g}{3}^2 \left(\frac{g+7}{g-2}\right) [p_0(u,v) \\ &- p(u)p(v)] \\ &+ 3\binom{g}{3}(g-3)[p_2(u,v) - p(u)p(v)] \end{aligned} \tag{50}$$

where $p(u)$, $p_0(u,v)$ and $p_2(u,v)$ are given in Tables 2, 3, *and* 4.

TABLE 2
Numerators for $p(u)$ Under U/MAN Distribution

u	$p(u)$
003	$n^{(3)}$
012	$3an^{(2)}$
102	$3mn^{(2)}$
021D	$\frac{3}{4}na^{(2)}$
021U	$\frac{3}{4}na^{(2)}$
021C	$\frac{3}{2}na^{(2)}$
111D	$3man$
111U	$3man$
030T	$\frac{3}{4}a^{(3)}$
030C	$\frac{1}{4}a^{(3)}$
201	$3nm^{(2)}$
120D	$\frac{3}{4}ma^{(2)}$
120U	$\frac{3}{4}ma^{(2)}$
120C	$\frac{3}{2}ma^{(2)}$
210	$3am^{(2)}$
300	$m^{(3)}$

In Holland and Leinhardt (1970) we illustrated how some of the values of $p(u)$, $p_0(u, v)$, and $p_2(u, v)$ are calculated so we do not repeat these derivations here. Tables 2, 3, and 4 give the formulas for $p(u)$, $p_0(u, v)$, and $p_2(u, v)$ for the U | *MAN* distribution, extending the corresponding tables in Holland and Leinhardt (1970) which gave values only for the intransitive triad types, that is, those that contain at least one intransitive configuration.

In Tables 2, 3, and 4, we have used the descending factorial notation in which $X^{(k)} = X(X - 1) \ldots (X - k + 1)$. Furthermore, Tables 2, 3, and 4 contain only the numerator for the probabilities. The denominators are respectively D_1, D_2, and D_3, where

$$D_1 = \binom{g}{2}^{(3)} \tag{51}$$

$$D_2 = \binom{g}{2}^{(6)} \tag{52}$$

$$D_3 = \binom{g}{2}^{(5)} \tag{53}$$

Moments of Linear Combinations of a Triad Census

Once all the work has been done to generate the means, variances, and covariances for a random triad census $\mathbf{T}$, it is simple to calculate the corresponding moments of any linear combination $\mathbf{l}'\mathbf{T}$. An elementary result from probability theory is that if $\mathbf{l}'\mathbf{T}$ and $\mathbf{s}'\mathbf{T}$ denote two linear combinations of the components of a triad census, then

$$E(\mathbf{l}'\mathbf{T}) = \mathbf{l}'\boldsymbol{\mu}_T \tag{54}$$

$$\mathrm{var}(\mathbf{l}'\mathbf{T}) = \mathbf{l}' \sum_{T^1} \tag{55}$$

and

$$\mathrm{cov}(\mathbf{l}'\mathbf{T}, \mathbf{s}'\mathbf{T}) = \mathbf{l}' \sum_T \mathbf{s} \tag{56}$$

where $\boldsymbol{\mu}_T = E(\mathbf{T})$ is the vector of expected values of the T_u, and $\sum_T$ is the covariance matrix of $\mathbf{T}$.

TABLE 3

Numerator for $p_0(u, v)$ Under U | *MAN* Distribution

Part 1

	003	*012*	*102*	*021D*	*021U*	*021C*	*111D*	*111U*
003	$n^{(6)}$	—	—	—	—	—	—	—
012	$3an^{(5)}$	$9a^{(2)}n^{(4)}$	—	—	—	—	—	—
102	$3mn^{(5)}$	$9man^{(4)}$	$9m^{(2)}n^{(4)}$	—	—	—	—	—
021D	$\frac{3}{4}a^{(2)}n^{(4)}$	$\frac{9}{4}a^{(3)}n^{(3)}$	$\frac{9}{4}ma^{(2)}n^{(3)}$	$\frac{9}{16}a^{(4)}n^{(2)}$	—	—	—	—
021U	$\frac{3}{4}a^{(2)}n^{(4)}$	$\frac{9}{4}a^{(3)}n^{(3)}$	$\frac{9}{4}ma^{(2)}n^{(3)}$	$\frac{9}{16}a^{(4)}n^{(2)}$	$\frac{9}{16}a^{(4)}n^{(2)}$	—	—	—
021C	$\frac{3}{2}a^{(2)}n^{(4)}$	$\frac{9}{2}a^{(3)}n^{(3)}$	$\frac{9}{2}ma^{(2)}n^{(3)}$	$\frac{9}{8}a^{(4)}n^{(2)}$	$\frac{9}{8}a^{(4)}n^{(2)}$	$\frac{9}{4}a^{(4)}n^{(2)}$	—	—
111D	$3man^{(4)}$	$9ma^{(2)}n^{(3)}$	$9m^{(2)}an^{(3)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{2}ma^{(3)}n^{(2)}$	$9m^{(2)}a^{(2)}n^{(2)}$	—
111U	$3man^{(4)}$	$9ma^{(2)}n^{(3)}$	$9m^{(2)}an^{(3)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{2}ma^{(3)}n^{(2)}$	$9m^{(2)}a^{(2)}n^{(2)}$	$9m^{(2)}a^{(2)}n^{(2)}$
030T	$\frac{3}{4}a^{(3)}n^{(3)}$	$\frac{9}{4}a^{(4)}n^{(2)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{16}a^{(5)}n$	$\frac{9}{16}a^{(5)}n$	$\frac{9}{8}a^{(5)}n$	$\frac{9}{4}ma^{(4)}n$	$\frac{9}{4}ma^{(4)}n$
030C	$\frac{1}{4}a^{(3)}n^{(3)}$	$\frac{3}{4}a^{(4)}n^{(2)}$	$\frac{3}{4}ma^{(3)}n^{(2)}$	$\frac{3}{16}a^{(5)}n$	$\frac{3}{16}a^{(5)}n$	$\frac{3}{8}a^{(5)}n$	$\frac{3}{4}ma^{(4)}n$	$\frac{3}{4}ma^{(4)}n$
201	$3m^{(2)}n^{(4)}$	$9m^{(2)}an^{(3)}$	$9m^{(3)}n^{(3)}$	$\frac{9}{4}m^{(2)}a^{(2)}n^{(2)}$	$\frac{9}{4}m^{(2)}a^{(2)}n^{(2)}$	$\frac{9}{2}m^{(2)}a^{(2)}n^{(2)}$	$9m^{(3)}an^{(2)}$	$9m^{(3)}an^{(2)}$
120D	$\frac{3}{4}ma^{(2)}n^{(3)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{4}m^{(2)}a^{(2)}n^{(2)}$	$\frac{9}{16}ma^{(4)}n$	$\frac{9}{16}ma^{(4)}n$	$\frac{9}{8}ma^{(4)}n$	$\frac{9}{4}m^{(2)}a^{(3)}n$	$\frac{9}{4}m^{(2)}a^{(3)}n$
120U	$\frac{3}{4}ma^{(2)}n^{(3)}$	$\frac{9}{4}ma^{(3)}n^{(2)}$	$\frac{9}{4}m^{(2)}a^{(2)}n^{(2)}$	$\frac{9}{16}ma^{(4)}n$	$\frac{9}{16}ma^{(4)}n$	$\frac{9}{8}ma^{(4)}n$	$\frac{9}{4}m^{(2)}a^{(3)}n$	$\frac{9}{4}m^{(2)}a^{(3)}n$
120C	$\frac{3}{2}ma^{(2)}n^{(3)}$	$\frac{9}{2}ma^{(3)}n^{(2)}$	$\frac{9}{2}m^{(2)}a^{(2)}n^{(2)}$	$\frac{9}{8}ma^{(4)}n$	$\frac{9}{8}ma^{(4)}n$	$\frac{9}{4}ma^{(4)}n$	$\frac{9}{2}m^{(2)}a^{(3)}n$	$\frac{9}{2}m^{(2)}a^{(3)}n$
210	$3m^{(2)}an^{(3)}$	$9m^{(2)}a^{(2)}n^{(2)}$	$9m^{(3)}an^{(2)}$	$\frac{9}{4}m^{(2)}a^{(3)}n$	$\frac{9}{4}m^{(2)}a^{(3)}n$	$\frac{9}{2}m^{(2)}a^{(3)}n$	$9m^{(3)}a^{(2)}n$	$9m^{(3)}a^{(2)}n$
300	$m^{(3)}n^{(3)}$	$3m^{(3)}an^{(2)}$	$3m^{(4)}n^{(2)}$	$\frac{3}{4}m^{(3)}a^{(2)}n$	$\frac{3}{4}m^{(3)}a^{(2)}n$	$\frac{3}{2}m^{(3)}a^{(2)}n$	$3m^{(4)}an$	$3m^{(4)}an$

Part 2

	030T	*030C*	*201*	*120D*	*120U*	*120C*	*210*	*300*
003	—	—	—	—	—	—	—	—
012	—	—	—	—	—	—	—	—
102	—	—	—	—	—	—	—	—
021D	—	—	—	—	—	—	—	—
021U	—	—	—	—	—	—	—	—
021C	—	—	—	—	—	—	—	—
111D	—	—	—	—	—	—	—	—
111U	—	—	—	—	—	—	—	—
030T	$\frac{9}{16}a^{(6)}$	—	—	—	—	—	—	—
030C	$\frac{3}{16}a^{(6)}$	$\frac{1}{16}a^{(6)}$	—	—	—	—	—	—
201	$\frac{9}{4}m^{(2)}a^{(3)}n$	$\frac{3}{4}m^{(2)}a^{(3)}n$	$9m^{(4)}n^{(2)}$	—	—	—	—	—
120D	$\frac{9}{16}ma^{(5)}$	$\frac{3}{16}ma^{(5)}$	$\frac{9}{4}m^{(3)}a^{(2)}n$	$\frac{9}{16}m^{(2)}a^{(4)}$	—	—	—	—
120U	$\frac{9}{16}ma^{(5)}$	$\frac{3}{16}ma^{(5)}$	$\frac{9}{4}m^{(3)}a^{(2)}n$	$\frac{9}{16}m^{(2)}a^{(4)}$	$\frac{9}{16}m^{(2)}a^{(4)}$	—	—	—
120C	$\frac{9}{8}ma^{(5)}$	$\frac{3}{8}ma^{(5)}$	$\frac{9}{2}m^{(3)}a^{(2)}n$	$\frac{9}{8}m^{(2)}a^{(4)}$	$\frac{9}{8}m^{(2)}a^{(4)}$	$\frac{9}{4}m^{(2)}a^{(4)}$	—	
210	$\frac{9}{4}m^{(2)}a^{(4)}$	$\frac{3}{4}m^{(2)}a^{(4)}$	$9m^{(4)}an$	$\frac{9}{4}m^{(3)}a^{(3)}$	$\frac{9}{4}m^{(3)}a^{(3)}$	$\frac{9}{2}m^{(3)}a^{(3)}$	$9m^{(4)}a^{(2)}$	—
300	$\frac{3}{4}m^{(3)}a^{(3)}$	$\frac{1}{4}m^{(3)}a^{(3)}$	$3m^{(5)}n$	$\frac{3}{4}m^{(4)}a^{(2)}$	$\frac{3}{4}m^{(4)}a^{(2)}$	$\frac{3}{2}m^{(4)}a^{(2)}$	$3m^{(5)}a$	$m^{(6)}$

TABLE 4
Numerator for $p_2(u, v)$ Under U | *MAN* Distribution

Part 1

	003	*012*	*102*	*021D*
003	$n^{(5)}$	—	—	—
012	$2an^{(4)}$	$an^{(3)}(4a+n-7)$	—	—
102	$2mn^{(4)}$	$4man^{(3)}$	$mn^{(3)}(4m+n-7)$	—
021D	$\frac{1}{4}a^{(2)}n^{(3)}$	$\frac{1}{2}a^{(2)}n^{(2)}(a+n-4)$	$\frac{1}{2}ma^{(2)}n^{(2)}$	$\frac{1}{16}a^{(3)}n(a+4n-7)$
021U	$\frac{1}{4}a^{(2)}n^{(3)}$	$\frac{1}{2}a^{(2)}n^{(2)}(a+n-4)$	$\frac{1}{2}ma^{(2)}n^{(2)}$	$\frac{1}{16}a^{(3)}n(a+4n-7)$
021C	$\frac{1}{2}a^{(2)}n^{(3)}$	$a^{(2)}n^{(2)}(a+n-4)$	$ma^{(2)}n^{(2)}$	$\frac{1}{8}a^{(3)}n(a+4n-7)$
111D	$man^{(3)}$	$man^{(2)}(2a+n-4)$	$man^{(2)}(2m+n-4)$	$\frac{1}{4}ma^{(2)}n(a+2n-4)$
111U	$man^{(3)}$	$man^{(2)}(2a+n-4)$	$man^{(2)}(2m+n-4)$	$\frac{1}{4}ma^{(2)}n(a+2n-4)$
030T	0	$\frac{3}{4}a^{(3)}n^{(2)}$	0	$\frac{3}{8}a^{(4)}n$
030C	0	$\frac{1}{4}a^{(3)}n^{(2)}$	0	$\frac{1}{8}a^{(4)}n$
201	$m^{(2)}n^{(3)}$	$2m^{(2)}an^{(2)}$	$2m^{(2)}n^{(2)}(m+n-4)$	$\frac{1}{4}m^{(2)}a^{(2)}n$
120D	0	$\frac{1}{2}ma^{(2)}n^{(2)}$	$\frac{1}{4}ma^{(2)}n^{(2)}$	$\frac{1}{4}ma^{(3)}n$
120U	0	$\frac{1}{2}ma^{(2)}n^{(2)}$	$\frac{1}{4}ma^{(2)}n^{(2)}$	$\frac{1}{4}ma^{(3)}n$
120C	0	$ma^{(2)}n^{(2)}$	$\frac{1}{2}ma^{(2)}n^{(2)}$	$\frac{1}{2}ma^{(3)}n$
210	0	$m^{(2)}an^{(2)}$	$2m^{(2)}an^{(2)}$	$\frac{1}{2}m^{(2)}a^{(2)}n$
300	0	0	$m^{(3)}n^{(2)}$	0

Part 2

	021U	*021C*	*111D*	*111U*
003	—	—	—	—
012	—	—	—	—
102	—	—	—	—
021D	—	—	—	—
021U	$\frac{1}{16}a^{(3)}n(a+4n-7)$	—	—	—
021C	$\frac{1}{8}a^{(3)}n(a+4n-7)$	$\frac{1}{4}a^{(3)}n(a+4n-7)$	—	—
111D	$\frac{1}{4}ma^{(2)}n(a+2n-4)$	$\frac{1}{2}ma^{(2)}n(a+2n-4)$	$ma^{(2)}n^{(2)}+m^{(2)}an^{(2)}+m^{(2)}a^{(2)}n$	—
111U	$\frac{1}{4}ma^{(2)}n(a+2n-4)$	$\frac{1}{2}ma^{(2)}n(a+2n-4)$	$ma^{(2)}n^{(2)}+m^{(2)}an^{(2)}+m^{(2)}a^{(2)}n$	$ma^{(2)}n^{(2)}+m^{(2)}an^{(2)}+m^{2}a^{(2)}n$
030T	$\frac{3}{8}a^{(4)}n$	$\frac{3}{4}a^{(4)}n$	$\frac{3}{4}ma^{(3)}n$	$\frac{3}{4}ma^{(3)}n$
030C	$\frac{1}{8}a^{(4)}n$	$\frac{1}{4}a^{(4)}n$	$\frac{1}{4}ma^{(3)}n$	$\frac{1}{4}ma^{(3)}n$
201	$\frac{1}{4}m^{(2)}a^{(2)}n$	$\frac{1}{2}m^{(2)}a^{(2)}n$	$m^{(2)}an(m+2n-4)$	$m^{(2)}an(m+2n-4)$
120D	$\frac{1}{4}ma^{(3)}n$	$\frac{1}{2}ma^{(3)}n$	$\frac{1}{4}ma^{(2)}n(a+2m-4)$	$\frac{1}{4}ma^{(2)}n(a+2m-4)$
120U	$\frac{1}{4}ma^{(3)}n$	$\frac{1}{2}ma^{(3)}n$	$\frac{1}{4}ma^{(2)}n(a+2m-4)$	$\frac{1}{4}ma^{(2)}n(a+2m-4)$
120C	$\frac{1}{2}ma^{(3)}n$	$ma^{(3)}n$	$\frac{1}{2}ma^{(2)}n(a+2m-4)$	$\frac{1}{2}ma^{(2)}n(a+2m-4)$
210	$\frac{1}{2}m^{(2)}a^{(2)}n$	$m^{(2)}a^{(2)}n$	$m^{(2)}an(m+2a-4)$	$m^{(2)}an(m+2a-4)$
300	0	0	$m^{(3)}an$	$m^{(3)}an$

TABLE 4 (*Continued*)

Part 3

	030T	*030C*	*201*	*120D*
003	—	—	—	—
012	—	—	—	—
102	—	—	—	—
021D	—	—	—	—
021U	—	—	—	—
021C	—	—	—	—
111D	—	—	—	—
111U	—	—	—	—
030T	$\frac{9}{16}a^{(5)}$	—	—	—
030C	$\frac{3}{16}a^{(5)}$	$\frac{1}{16}a^{(5)}$	—	—
201	0	0	$m^{(3)}n[m + 4n - 7]$	—
120D	$\frac{3}{8}ma^{(4)}$	$\frac{1}{8}ma^{(4)}$	$\frac{1}{2}m^{(2)}a^{(2)}n$	$\frac{1}{16}ma^{(3)}[a + 4m - 7]$
120U	$\frac{3}{8}ma^{(4)}$	$\frac{1}{8}ma^{(4)}$	$\frac{1}{2}m^{(2)}a^{(2)}n$	$\frac{1}{16}ma^{(3)}[a + 4m - 7]$
120C	$\frac{3}{4}ma^{(4)}$	$\frac{1}{4}ma^{(4)}$	$m^{(2)}a^{(2)}n$	$\frac{1}{8}ma^{(3)}[a + 4m - 7]$
210	$\frac{3}{4}m^{(2)}a^{(3)}$	$\frac{1}{4}m^{(2)}a^{(3)}$	$4m^{(3)}an$	$\frac{1}{2}m^{(2)}a^{(2)}[m + a - 4]$
300	0	0	$2m^{(4)}n$	$\frac{1}{4}m^{(3)}a^{(2)}$

Part 4

	120U	*120C*	*210*	*300*
003	—	—	—	—
012	—	—	—	—
102	—	—	—	—
021D	—	—	—	—
021U	—	—	—	—
021C	—	—	—	—
111D	—	—	—	—
111U	—	—	—	—
030T	—	—	—	—
030C	—	—	—	—
201	—	—	—	—
120D	—	—	—	—
120U	$\frac{1}{16} m a^{(3)} [a + 4m - 7]$	—	—	—
120C	$\frac{1}{8} m a^{(3)} [a + 4m - 7]$	$\frac{1}{4} m a^{(3)} [a + 4m - 7]$	—	—
210	$\frac{1}{8} m^{(2)} a^{(2)} [m + a - 4]$	$m^{(2)} a^{(2)} [m + a - 4]$	$m^{(3)} a [m + 4a - 7]$	—
300	$\frac{1}{4} m^{(3)} a^{(2)}$	$\frac{1}{2} m^{(3)} a^{(2)}$	$2m^{(4)} a$	$m^{(5)}$

TESTING STRUCTURAL HYPOTHESES

In this section we make use of the tools we have developed so far and describe a procedure for testing propositions about local structure in a sociomatrix. The procedure has these steps: (1) operationalize the proposition into a hypothesis that a particular 3-configuration will tend to occur or fail to occur in the sociomatrix (this hypothesis will usually be directional); (2) find the weighting vector that when applied to the triad census will enumerate the critical 3-configuration; (3) use the weighting vector to enumerate the critical 3-configuration as well as to compute its mean and variance for a random triad census; (4) set up a test statistic that compares the observed and the expected number of critical configurations and use this discrepancy as a basis for testing the structural proposition. We discuss each of these steps in turn.

Operationalizing a Structural Proposition

In this initial step we take a proposition about the structure of a network and translate it, if possible, into a prediction about the number of 3-configurations of a particular type in observed sociomatrices. As an example, we consider Mazur's proposition (1971), mentioned above. Note that a simple sociomatrix indicating choice or nonchoice does not represent the strength of the relation, so Mazur's distinction between friends and close friends has to be made on some other basis if the proposition is to be operationalized and tested on binary choice data. Following Mazur's suggestion, we assume that mutual dyads indicate "close friends" while asymmetric dyads indicate "friends." Mazur's proposition leads us to examine not one but seven different 3-configurations. They all have the reading rule given by

$$ij \; ji \; ik \; jk \tag{57}$$

The first two pairs in the reading rule refer to the pair of individuals designated "close friends," "friends," or neither. The second two pairs in the reading rule refer to the choice or nonchoice of other group members by the pair in question. For example, the configuration that corresponds to "close friends agreeing on their choices" is given by

$$\begin{pmatrix} ij & ji & ik & jk \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{58}$$

But, "close friends agreeing on their nonchoices" is the configuration

$$\begin{pmatrix} ij & ji & ik & jk \\ 1 & 1 & 0 & 0 \end{pmatrix} \tag{59}$$

"Close friends disagreeing on their choices" is given by either

$$\begin{pmatrix} ij & ji & ik & jk \\ 1 & 1 & 0 & 1 \end{pmatrix} \tag{60}$$

or

$$\begin{pmatrix} ij & ji & ik & jk \\ 1 & 1 & 1 & 0 \end{pmatrix} \tag{61}$$

as mentioned earlier. Since the reading rule is understood here to be the one given in (51), we use the following shorthand notation for these configurations: (58) is denoted by 1111, (59) by 1100, and (60) or (61) by 1101/1110. We use only the lower part of the configuration matrix; if there are two equivalent forms, they are separated by a slash. In this notation, Mazur's proposition makes predictions about the following seven configurations: 1111, 1110, 1101/1110, 1011/0111, 1000/0100, 1010/0101, 1001/0110. The first three deal with agreement and disagreement among "close friends," while the last four deal with agreement and disagreement among "friends."

Now for the sign of the prediction. We propose to compare the observed number of the seven configurations with the corresponding number expected in a random sociomatrix having the same number of mutual, asymmetric, and null dyads. The expected values come from the $U \mid MAN$ distribution for a random digraph discussed above which appears especially appropriate for Mazur's proposition because it is based on mutual and asymmetric pairs. In view of the foregoing remarks, we would formulate Mazur's predictions as given by the second column of Table 5.

TABLE 5
Results of Testing Mazur's Proposition

	Configuration	Model	Predicted sign	Median $\tau(1)$ for 408 sociomatrices
Friends	1011/0111	Agreement	+	2.36
	1000/0100		+	1.82
	1010/0101	Disagreement	−	−1.73
	1001/0110		−	−1.49
Close Friends	1111	Agreement	+	3.48
	1100		+	1.73
	1101/1110	Disagreement	−	−3.81

Note that, under the U | MAN distribution, the seven predictions in Table 5 are not independent. For any observed sociomatrix, the configurations of the first four types in Table 5, grouped under "Friends," sum to $(g - 2)A$. The last three types, grouped under "Close Friends," sum to $(g - 2)M$. These summations also apply to the expected values under the U | MAN distribution. There are therefore only five independent predictions. Investigators must often remember such considerations when they examine more than one configuration.

Weighting Vector

In this step, we compute the number of ways each critical configuration occurs within the 16 triad types. Each resulting set of 16 numbers forms the elements of the weighting vector that is applied to the triad census to enumerate the critical configuration. For example, consider configuration 1111. As we go through the list of triad types in Figure 2, we see that this configuration does not occur in any triad until we come to triad 120U. This triad contains exactly one configuration of type 1111. Triads 210 and 300 contain, respectively, one and three configurations of this type. Hence, the weighting vector for the 1111 configuration is (0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 3). Table 6 gives the weighting vectors for all the configurations critical to Mazur's proposition.

The mean and variance of the number of these configurations in a random triad census from the U | MAN distribution are given by formulas (54) and (55). A computer program that carries out these calculations has been written and is implemented on

TABLE 6
Weighting Vectors for Configurations Critical to Mazur's Proposition

Triad	1011 0111	1000 0100	1010 0101	1001 0110	1111	1100	1101 1110
003	0	0	0	0	0	0	0
012	0	1	0	0	0	0	0
102	0	0	0	0	0	1	0
021D	0	0	2	0	0	0	0
021U	0	2	0	0	0	0	0
021C	0	1	0	1	0	0	0
111D	0	0	0	1	0	1	0
111U	0	0	1	0	0	0	1
030T	1	1	1	0	0	0	0
030C	0	0	0	3	0	0	0
201	0	0	0	0	0	0	2
120D	2	0	0	0	0	1	0
120U	0	0	2	0	1	0	0
120C	1	0	0	1	0	0	1
210	1	0	0	0	1	0	1
300	0	0	0	0	3	0	0

the TROLL interactive computer system of the National Bureau of Economic Research, Inc. (see Appendix C).

Setting Up the Test Statistic

Let $\mathbf{l}$ denote the weighting vector for a critical configuration, and let $\mathbf{T}$ denote the triad census vector as usual. Then $\mathbf{l}'\mathbf{T}$ is the number of times the critical configuration occurs in the observed sociomatrix. Under the $\mathrm{U} \mid MAN$ distribution, the expected number of these configurations is $\mathbf{l}'\boldsymbol{\mu}_T$, where $\boldsymbol{\mu}_T$ is given by Corollary 2 (48). The variance is $\mathbf{l}'\sum_T\mathbf{l}$, where $\sum_T$ is a 16 by 16 matrix whose main diagonal elements are given by Corollary 2 (49) and off-diagonal elements are given by Corollary 2 (50). The difference

$$\mathbf{l}'\mathbf{T} - \mathbf{l}'\boldsymbol{\mu}_T \tag{62}$$

is the discrepancy between the observed and expected number of critical configurations. Under the assumption that the triad census is random ($\mathrm{U} \mid MAN$), the test statistic $\tau(\mathbf{l})$, defined by

$$\tau(\mathbf{l}) = (\mathbf{l}'\mathbf{T} - \mathbf{l}'\boldsymbol{\mu}_T) \Big/ \sqrt{\mathbf{l}' \sum_T \mathbf{l}} \tag{63}$$

has an approximate normal distribution with mean zero and variance one. Approximate significance levels for values of $\tau(\mathbf{1})$ may be obtained from tables of the normal distribution. See Holland and Leinhardt (1970) for simulation results bearing on the adequacy of this approximation. The test statistic $\tau(\mathbf{1})$ is affected by the sample size, which, for a triad census, is $\binom{g}{3}$. It is possible for large values of g to generate large values of $\tau(\mathbf{1})$, and it is wise to plot $\tau(\mathbf{1})$ against g in any comparison of sociograms for which g varies widely. A possible alternative measure is of the form

$$\delta(\mathbf{1}) = (\mathbf{1}'\mathbf{T} - \mathbf{1}'\boldsymbol{\mu})/\mathbf{1}'\boldsymbol{\mu} \qquad (64)$$

(see, for example, Davis and Leinhardt, 1972). This percentage difference measure does adjust for g, but it is usually hard to decide whether a given value of $\delta(\mathbf{1})$ is large or small unless the measure is referred to a standard deviation and this procedure leads back to $\tau(\mathbf{1})$. A more fundamental problem stems from the current lack of parametric models in sociometric analyses. The natural way to introduce measures of transitivity or other structural properties is to have stochastic models for generating sociomatrices that contain parameters for these properties. Estimates of such parameters lead to measures of structural properties and confidence intervals for them. While we believe this area is important for future research, the method of this chapter is not directed toward it.

Using a computer program, we have calculated the value of $\tau(\mathbf{1})$ in 408 sociomatrices (randomly selected from those collected by Davis and Leinhardt, 1972) for the seven critical configurations for Mazur's proposition. In the last column of Table 5 we give the median value of $\tau(\mathbf{1})$ over these 408 sociomatrices for each of the seven critical configurations. The results support by and large the predictions for $\tau(\mathbf{1})$ that derive from Mazur's proposition. All of the median $\tau(\mathbf{1})$ values have the predicted sign. Furthermore, Mazur's prediction that "close friends" will agree more than "friends" seems to be substantiated by direct comparison of the median $\tau(\mathbf{1})$ values.

Another Example: Transitivity

Consider the proposition stated by Davis, Holland, and Leinhardt (1971, p. 309): "Interpersonal choices tend to be transitive—if P chooses O and O chooses X, then P is likely to choose X.'' Elsewhere (Holland and Leinhardt, 1971), we have described the social structural consequences of the transitivity of interpersonal affect when social status is associated with asymmetry and clustering is associated with mutuality. Here we use our method to explore the empirical validity of the proposition.

There are two critical configurations, which are given in (4) (intransitivity) and (5) (transitivity). The corresponding weighting vectors are given in Table 1. Using the 408 sociomatrices mentioned above we obtain median $\tau(\mathbf{l})$ values of 5.18 for transitivity and -3.89 for intransitivity. These results support the proposition that interpersonal choices tend to be transitive. For another view see Killworth (1974).

However, these results and those found for Mazur's proposition are not independent of each other. Indeed, the weighting vector for transitivity is the vector sum of the weighting vector for the configuration 1011/0111 and two times the weighting vector 1111. The weighting vector for intransitivity is the sum of the weighting vectors for configurations 1010/0101 and 1101/1110. The transitivity prediction is therefore weaker than Mazur's in that it involves fewer configurations.

We may view the strong observed effect of transitivity as the sum of two more modest effects corresponding to the two Mazur configurations 1011/0111 and 1111. Conversely, the effects of these two configurations may be viewed as the result of being pulled along by the strong transitivity effect. In a future paper we will report the use of the techniques described in Appendix B to examine Mazur's proposition, controlling for the effects of transitivity.

SUMMARY AND DISCUSSION

Our purpose has been to describe a method for formalizing and testing theoretical propositions about regularities in social networks. This method involves the use of graph-theoretic concepts to restate verbal propositions about local structure in terms of configurations. Although configurations are relatively simple graph-theoretic concepts, we have shown that they can be used to represent important propositions about social structure. Arguing that most theories of structure in interpersonal relations concern average local properties, we have developed statistical procedures for testing the tendency of a particular property to hold across a social network. With these procedures, an investigator can determine whether the discrepancy between the empirical occurrence and the chance occurrence of local structure is statistically significant. We have presented examples to illustrate the use of this method and its interpretation.

Sociometric data are quite common. They represent an important and fundamental resource in the study of structure in interpersonal relations. However, these data are complex, and, like all empirical measurements, they contain an unknown amount of measurement error. Their inherent complexity and the likely presence of measurement error reduce their applicability to the study of global organization in small social systems. Moreover, such applications can also be questioned from the point of view of the level at which the data are collected. Interpersonal affect or choice data, the most frequently collected sociometric data, represent surveys of individual attitudes. In effect, individuals are asked to provide information on the nature of their local ties. The task each individual is presented with is to judge whether a relation exists between that individual and each other group member. Whatever overall consistency exists in these sets of responses derives from local regularities in interpersonal relations. Although some simple global models may fit these data, the data are more properly used in the study of local conditions.

The method we have described has several advantages when used to study local structure in small-scale social systems. First, it leads investigators to develop operational constructs that are subject to empirical testing. Many of the theoretical propositions advanced by sociologists and social psychologists concern the behavioral and experiential consequences of various arrangements of interpersonal relations. However, these propositions are rarely stated with a precision and absence of ambiguity that permit them to be tested. Second, the method focuses attention on the analysis of average local structure in sociometric data. It thereby exploits the essential feature of these data. Finally, the method facilitates the analysis of structural tendencies in large collections of groups. Only through analyses of large collections of network data will investigators be able to detect general tendencies in the structure of social relational systems and develop reliable propositions that hold for general classes of social structures.

APPENDIX A: CONDITIONAL UNIFORM DISTRIBUTIONS FOR RANDOM DIGRAPHS

Various considerations lead us to study random digraphs whose distributions are different from and often more complicated than the U | *MAN* distribution. In this appendix we describe some of these distributions.

Uniform Distribution

The uniform distribution is the basic distribution from which all others we discuss here may be obtained by controlling for particular statistics of the graph. For the uniform distribution, all labeled digraphs with g nodes are equally likely. It is easy to generate the sociomatrix (X_{ij}) for a uniformly distributed random digraph because the X_{ij} are independent zero-one random variables with

$$P\{X_{ij} = 1\} = 1/2 \qquad i \neq j \tag{65}$$

U | C Distribution

The U | C distribution is a simple conditional distribution of the uniform distribution just described. C defined in (12) denotes the number of directed edges in the graph. Thus U | C is the uniform distribution conditioned on C. It makes all labeled digraphs of size g with a specified value of C equally likely. It is also easy to generate the sociomatrix for such a random graph by selecting at random and without replacement C of the $g(g - 1)$ possible ordered pairs of nodes and allocating the C-directed edges to them. In the uniform distribution, C is a random variable with a binomial distribution, whereas, in the U | C distribution, C is not random and is fixed at a specified value.

U | $\{X_{i+}\}$ Distribution

The U | $\{X_{i+}\}$ distribution is the uniform distribution conditioned by the out-degrees. For the U | $\{X_{i+}\}$ distribution, all labeled digraphs of size g with the specified out-degrees are equally likely. To generate (X_{ij}) from U | $\{X_{i+}\}$ we observe that all the rows of (X_{ij}) are statistically independent, and in the i^{th} row we simply choose at random and without replacement X_{i+} columns (excluding the i^{th}) for the ones and set the rest equal to zero. This distribution is an important baseline for the allocation of choices in a sociogram. If we transpose (X_{ij}), the U | $\{X_{i+}\}$ becomes the U | $\{X_{+j}\}$ distribution. Note that in the U | $\{X_{i+}\{$ or the U | $\{X_{+j}\}$ distribution the value of C is fixed since

$$C = \sum_i X_{i+} = \sum_j X_{+j} \tag{66}$$

U | $\{X_{i+}\}$, $\{X_{+j}\}$ **Distribution**

In the U | $\{X_{i+}\}$, $\{X_{+j}\}$ distribution, all labeled digraphs of size g with the specified values of both $\{X_{i+}\}$ and $\{X_{+j}\}$ are equally likely. It is a highly nontrivial distribution, and no simple way seems to be known for generating random graphs with this exact distribution. Nevertheless, it is of potential importance in sociometric data analysis since it controls for both choices made and choices received.

U | M, $\{X_{i+}\}$ **Distribution**

In the U | M, $\{X_{i+}\}$ distribution, we combine both U | MAN and U | $\{X_{i+}\}$. We need to specify only M and $\{X_{i+}\}$, since $\{X_{i+}\}$ fixes C and M and C fix A. N is determined by M, A, and g. This distribution is also highly nontrivial, and no simple way is known for generating graphs from it. Again, it is of potential importance in sociometry because it controls for both choices made and mutuality. One would have preferred to use this distribution rather than the U | MAN, but currently this is not possible because of mathematical intractability.

U | M, $\{X_{i+}\}$, $\{X_{+j}\}$ **Distribution**

In the U | M, $\{X_{i+}\}$, $\{X_{+j}\}$ distribution, all digraphs with the specified values of M, $\{X_{i+}\}$, and $\{X_{+j}\}$ are equally likely. Again, this distribution is very difficult to work with, and no simple way is known for generating graphs from it. Its value to sociometric data analysis stems from the interpretation that it controls for choices made (possibly constrained by the experimental technique), choices received (a measure of status and isolation), and mutuality (a measure of friendship).

There are a variety of possible types of random graphs besides the U | MAN distribution. We have catalogued a few of the important ones that stem from the notion of the uniform distribution. There are a host of possibilities for nonuniform distributions, but we will not discuss them here. The main virtue of U | MAN is that it is the most highly conditioned uniform distribution that fixes M and A and for which we can currently compute relevant probabilities.

APPENDIX B: APPROXIMATE DISTRIBUTIONS FOR THE TRIAD CENSUS

While the U | MAN distribution is useful, it does not control for the out-degrees (which may reflect experimental constraints such as the fixed-choice procedure) or the in-degrees (which reflect status and isolation). We are left with the possibility that triad census $\mathbf{T}$, observed from real data, departs substantially from its U | MAN expected value $\boldsymbol{\mu}_T$ not because there is structure in the sociogram, but because $\boldsymbol{\mu}_T$ does not control for all the simple effects we would like to condition out in our analysis.

Since a direct attack on the exact U | M, $\{X_{i+}\}$ or U | M, $\{X_{i+}\}$, $\{X_{+j}\}$ distribution appears to be too difficult, at least at present, we propose the following indirect and approximate approach, which makes use of the simple nature of conditional distributions for the multivariate normal distribution.

We observe that $\mathbf{T}$ is a sum of loosely correlated indicator variables (see (28)), and thus it is plausible that, under various random distributions, $\mathbf{T}$ has an approximate multivariate normal distribution for large values of g. Thus we have

$$\mathbf{T} \sim \mathfrak{N}(\boldsymbol{\mu}, \Sigma) \tag{67}$$

for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ computed, say, from U | MAN. Next we make use of the fact that if $\mathbf{T}$ has the multivariate normal distribution in (67) and $\mathbf{LT}$ is a vector of linear combinations of the elements of $\mathbf{T}$, then

$$E(\mathbf{T} \mid \mathbf{LT} = \mathbf{Lt}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{L}'(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}')^{-1}\mathbf{L}(\mathbf{t} - \boldsymbol{\mu}) \tag{68}$$

and

$$\operatorname{cov}(\mathbf{T} \mid \mathbf{LT} = \mathbf{Lt}) = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{L}'(\mathbf{L}\boldsymbol{\Sigma}\mathbf{L}')^{-1}\mathbf{L}\boldsymbol{\Sigma} \tag{69}$$

Thus it is fairly simple to approximately condition **T** on linear combinations of its elements. Now we make one further approximation. Instead of conditioning on $\{X_{i+}\}$, we condition on the mean $\overline{C}$ and variance S^2_{out} of the $\{X_{i+}\}$. While this is clearly a reduction in the level of conditioning used, it promises to be useful. For example, if $S^2_{\text{out}} = 0$, all the $\{X_{i+}\}$ are equal (fixed-choice procedure). Now we use the result that $\overline{C}$ and S^2_{out} are essentially linear combinations of the triad frequencies to transform the problem to one of conditioning **T** on some of its linear combinations. This approach promises to be powerful, and in a future paper we will report the result of using it to test structural propositions on a large data bank of sociograms.

APPENDIX C: COMPUTATIONAL PROCEDURES

The methods we have described in this chapter contain two essential computational components: a triad census that must be produced from an empirical sociomatrix, and a mean vector and covariance matrix that must be computed for a U | *MAN* random triad census. We have constructed various computer algorithms for performing these computations. An early routine that enumerates triads, computes the entire mean vector, and computes selected portions of the covariance matrix is the program SOCPAC I (Leinhardt, 1971). Separate routines for counting triads and computing the mean vector and covariance matrix written in FORTRAN have been run in a time-sharing mode under TSS on the Carnegie-Mellon University IBM 360/67. Another version of these routines has been implemented as two functions in TROLL—the interactive computing system of the National Bureau of Economic Research. This system is available through a national computer network.

Using Tables 2, 3, and 4 and Corollary 2, one can straightforwardly write a program to compute the mean vector and covariance matrix for a random triad census under U | *MAN*. Those who wish to do so, must exercise care to avoid round-off errors. In particular, we found it necessary to use double precision throughout the computations.

To count triads, we have found the following simple algorithm efficient: (a) Iterate over (i, j, k), such that $1 \leq i < j <$

$k \leq g$ in the standard way, using nested loops. (b) For each iteration (that is, a particular value of (i, j, k)), form the vector

$$(X_{ij}, X_{jk}, X_{ki}, X_{ik}, X_{kj}, X_{ji})$$

where these six components are the appropriate entries in the sociomatrix $\mathbf{X}$. There are $2^6 = 64$ possible realizations of this vector, each one corresponding to a labeled triad type. A simple mapping associates one of the 16 isomorphism classes with each of these 64 possible realizations. (c) Using this mapping, directly increment the count for the isomorphism class observed at this iteration. At the end of this iterative process all the $\binom{g}{3}$ triads will be enumerated according to their triad types. This procedure can be generalized to count k-subgraphs, but even for tetrads it is substantially more involved since it requires mapping the 4096 labeled 4-graphs into their 218 isomorphism classes.

REFERENCES

BERELSON, B. R., LAZARSFELD, P. F., AND MCPHEE, W. N.

1954 *Voting: A Study of Opinion Formation in a Presidential Campaign.* Chicago: University of Chicago Press.

BOYLE, R. P.

1969 "Algebraic systems for normal and hierarchical sociograms." *Sociometry* 32:99–119.

CARTWRIGHT, D., AND HARARY, F.

1956 "Structural balance: a generalization of Heider's theory." *Psychological Review* 63:277–293.

COLEMAN, J. S.

1957 *Community Conflict.* New York: Free Press.

DAVIS, J. A.

1963 "Structural balance, mechanical solidarity, and interpersonal relations." *American Journal of Sociology* 68:444–463.

1967 "Clustering and structural balance in graphs," *Human Relations* 20:181–187.

1970 "Clustering and hierarchy in interpersonal relations: testing two graph theoretical models on 742 sociograms." *American Sociological Review* 35:843–852.

DAVIS, J. A., HOLLAND, P. W., AND LEINHARDT, S.

1971 "Comments on Professor Mazur's hypothesis about interpersonal sentiments." *American Sociological Review* 36:309–311.

DAVIS, J. A. AND LEINHARDT, S.

1972 "The structure of positive interpersonal relations in small groups." In J. Berger (Ed.), *Sociological Theories in Progress.* Vol. 2. Boston-Houghton Mifflin.

FLAMENT, C.

1963 *Applications of Graph Theory to Group Structure.* Englewood Cliffs, N.J.: Prentice-Hall.

FRENCH, J. R. P.

1956 "A formal theory of social power." *Psychological Review* 63:181–195.

HARARY, F., NORMAN, R. Z., AND CARTWRIGHT, D.

1965 *Structural Models.* New York: Wiley.

HEIDER, F.

1944 "Social perception and phenomenal causality." *Psychological Review* 51:358–374.

1958 *The Psychology of Interpersonal Relations.* New York: Wiley.

HOLLAND, P. W., AND LEINHARDT, S.

1970 "A method for detecting structure in sociometric data." *American Journal of Sociology* 70:492–513.

1971 "Transitivity in structural models of small groups." *Comparative Group Studies* 2:107–124.

1973 "The structural implications of measurement error in sociometry." *Journal of Mathematical Sociology* 3:85–111.

HOMANS, G. C.

1950 *The Human Group.* New York: Harcourt Brace Jovanovich.

1961 *Social Behavior: Its Elementary Forms.* New York: Harcourt Brace Jovanovich.

KATZ, E., AND LAZARSFELD, P. F.

1955 *Personal Influence: The Part Played by People in the Flow of Mass Communications.* New York: Free Press.

KILLWORTH, P.

1974 "Intransitivity in the structure of small closed groups." *Social Science Research* 3:1–23.

LAZARSFELD, P. F., AND MERTON, R. K.

1954 "Friendship as a social process: a substantive and methodological analysis." In M. Berger, T. Abel, and C. H. Page (Eds.), *Freedom and Control in Modern Society* New York: Van Nostrand Reinhold, pp. 18–66.

LEINHARDT, S.

1971 "SOCPAC I: a FORTRAN IV program for structural analysis of sociometric data." *Behavioral Science* 16:515–516.

LIPSET, S. M., TROW, M. A., AND COLEMAN, J. S.
1956 *Union Democracy: The Internal Politics of the International Typographical Union.* New York: Free Press.

LORRAIN, F., AND WHITE, H. C.
1972 "Structural equivalence of individuals in social networks." *Journal of Mathematical Sociology* 1:49–80.

MAZUR, A.
1971 "Comments on Davis' graph model." *American Sociological Review* 36:308–311.

MOON, J. W.
1968 *Topics on Tournaments.* New York: Holt, Rinehart, and Winston.

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

MORENO, J. L., AND OTHERS.
1960 *The Sociometry Reader.* New York: Free Press.

NADEL, S. F.
1957 *The Theory of Social Structure.* London: Cohen and West.

RADCLIFF-BROWN, A. R.
1940 "On social structure." *Journal of the Royal Anthropological Institute* 70:1–12.

WHITE, H. C.
1963 *An Anatomy of Kinship: Mathematical Models for Structures of Cumulated Roles.* Englewood Cliffs, N.J.: Prentice-Hall.