NBER WORKING PAPER SERIES

RANDOM DIRECTED GRAPH DISTRIBUTIONS AND THE TRIAD CENSUS IN SOCIAL NETWORKS

Stanley S. Wasserman*

Working Paper No. 113

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

November 1975

Preliminary: not for quotation

*NBER Computer Research Center. Research supported in part by National Science Foundation Grants SOC 73-05489 to Carnegie-Mellon University and DCR70-03456-A04 to the National Bureau of Economic Research, Inc.

Abstract

This paper uses the concept of the triad census first introduced by Holland and Leinhardt, and describes several distributions on directed graphs. Methods are presented for calculating the mean and the covariance matrix of the triad census for the uniform distribution that conditions on the number of choices made by each individual in the social network. Several complex distributions on digraphs are approximated, and an application of these methods to a sociogram is given.

Acknowledgements

This paper was prepared as a Ph.D. qualifying paper for the Department of Statistics at Harvard University. I am extremely grateful to Paul Holland who guided the research and made valuable comments on the data analyses conducted in Section 6. In addition, he suggested the form of the figure that appears in Section 7.

I would also like to thank Professors Fred Mosteller and David Oakes who made helpful suggestions on an earlier draft of this paper, and Sandy Moriarty and Sheila Howard who typed this manuscript, and an earlier version.

Contents

1. Introduction

This paper discusses some recently developed methods for the analysis of social networks. The directed graph, or more briefly "digraph", a set of "nodes" or "points" and a set of directed "lines" or "edges" connecting pairs of nodes, is the basic mathematical concept in this paper.

This paper utilizes concepts of graph theory which have been found useful in discussing social networks. Recently, many structural models of social science have adopted the graph theory notation. These concepts will be introduced as needed in this paper, without going into a full exposition. Those readers unfamiliar with these ideas will find Harary, Norman, and Cartwright (1965) a valuable reference.

In a digraph, node $\underline{i}$ and node $\underline{j}$ are connected with a directed line running from i to j, if and only if person $\underline{i}$ chooses person $\underline{j}$ according to the sociometric choice criterion employed. Note that the digraph defined here is a "binary" directed graph. The "strengths" attached to individual choices are irrelevant. A binary digraph does not allow signed choices (each line is designated +1 or -1) or multiple choices (each line receives a value). It is also important to note that a directed line from node i to node j does not rule out the possibility of a directed line from node j to node i.

The central analytical tool in this presentation is the triad census, introduced by Holland & Leinhardt (1970). The main interest in this paper is the random directed graph distribution obtained as a result of conditioning on the number of choices made by each person. Under this distribution, all social networks which have exactly the specified number of choices for each person are equally likely.

Sections 3 and 4 of this paper are devoted to brief discussions of various random directed graph distributions and the triad census. Section 5 shows how to calculate the first two moments of the triad census under the digraph distribution mentioned above. An example of a social network is given in section 6, and then analyzed with the distribution.

2. Notation and Definitions

Let g denote the number of members in the group. Order the individuals in the group from 1 to g in an arbitrary manner. Define the (gxg) sociomatrix $\underset{\sim}{X}$ as a representation of a labeled binary directed graph. (A different ordering of the individuals produces a sociomatrix which differs from $\underset{\sim}{X}$ by a simultaneous row-column permutation.) The notation $i \to j$ imples that in the digraph there is a directed line from node i to node j. Let the (i,j) entry of $\underset{\sim}{X}$ be defined as follows:

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

Self-choices are not allowed; consequently, the diagonal elements of $\underset{\sim}{X}$, X_{ii}, $i = 1,2,\ldots,g$, remain empty, or conveniently are set to zero.

Two sets of quantities associated with $\underset{\sim}{X}$ are of particular interest to the investigator of the group. The outdegree of node i, written X_{i+}, is the number of lines in the digraph originating at node i. The indegree of node j, written X_{+j}, is the number of lines in the digraph terminating at node j. The row sums and column sums, respectively, of the sociomatrix give the outdegree and indegree of each node. These two sets of quantities $\{X_{i+}\}$ and $\{X_{+j}\}$ may be calculated as follows:

$$X_{i+} = \sum_j X_{ij} \text{ ; } i=1,\ 2,\ \ldots,\ g$$

and

$$X_{+j} = \sum_i X_{ij} \; ; \; j=1, 2, \ldots, g.$$

In a group of size g, the number of choices made by each person and the number of choices received by each person, the outdegree and indegree, respectively, take a value between 0 and (g-1).

A <u>mutual</u> relationship between person i and person j exists when $i \rightarrow j$ and $j \rightarrow i$ in the digraph. The mutual bond is denoted by $i \leftrightarrow j$. In the sociomatrix $\underset{\sim}{X}$, this situation occurs when $x_{ij}=1$ and $x_{ji}=1$. An asymmetric relationship occurs if and only if $i \rightarrow j$ or $j \rightarrow i$ but there is no mutual relationship present. A <u>null</u> relationship between i and j occurs if there is neither a mutual nor an asymmetric bond between these persons. Let M, A, and N, respectively, be the number of mutual, asymmetric, and null relationships in the group.

It is possible to represent each of these three relationships graphically. These representations of pairs of nodes are commonly referred to as <u>dyads</u>. Each pair of points, and the lines connecting them, are isomorphic to one of these three representations; consequently, the three dyad types are often referred to as isomorphism classes. The classes are named null, asymmetric, and mutual to correspond to the sociological concepts. Figure 2-1 illustrates the dyad types.

$\cdot \qquad\quad \cdot$ Null

$\cdot \longrightarrow \cdot$ Asymmetric

$\cdot \longleftrightarrow \cdot$ Mutual

Figure 2-1: The three Isomorphism classes for Digraphs with g=2 (i.e., the Dyad Types). Figure taken from Holland and Leinhardt (1976).

Now consider all possible isomorphism classes of triples of points, or triads. By enumeration, it is easy to show that there are 16 classes. These are illustrated in Figure 2.2. The naming convention employed in the figure was introduced by Holland and Leinhardt (1970), and uses the number of mutual, asymmetric, and null dyads within each triad as its basis.

In a digraph with g nodes, there are $\binom{g}{3}$ triads formed by selecting each triple of nodes and all lines connecting them. Suppose each of these triads is examined in turn, and the isomorphism class of each recorded. Let T_u denote the number of triads of type u (u ranges over the 16 triad types shown in Figure 2-2). The triad census $\underset{\sim}{T}$ is the 16 component vector defined by

$$\underset{\sim}{T} = (T_{003}, T_{012}, \ldots, T_{300}).$$

Adhering to an established convention, the ordering of the components of T is as follows:

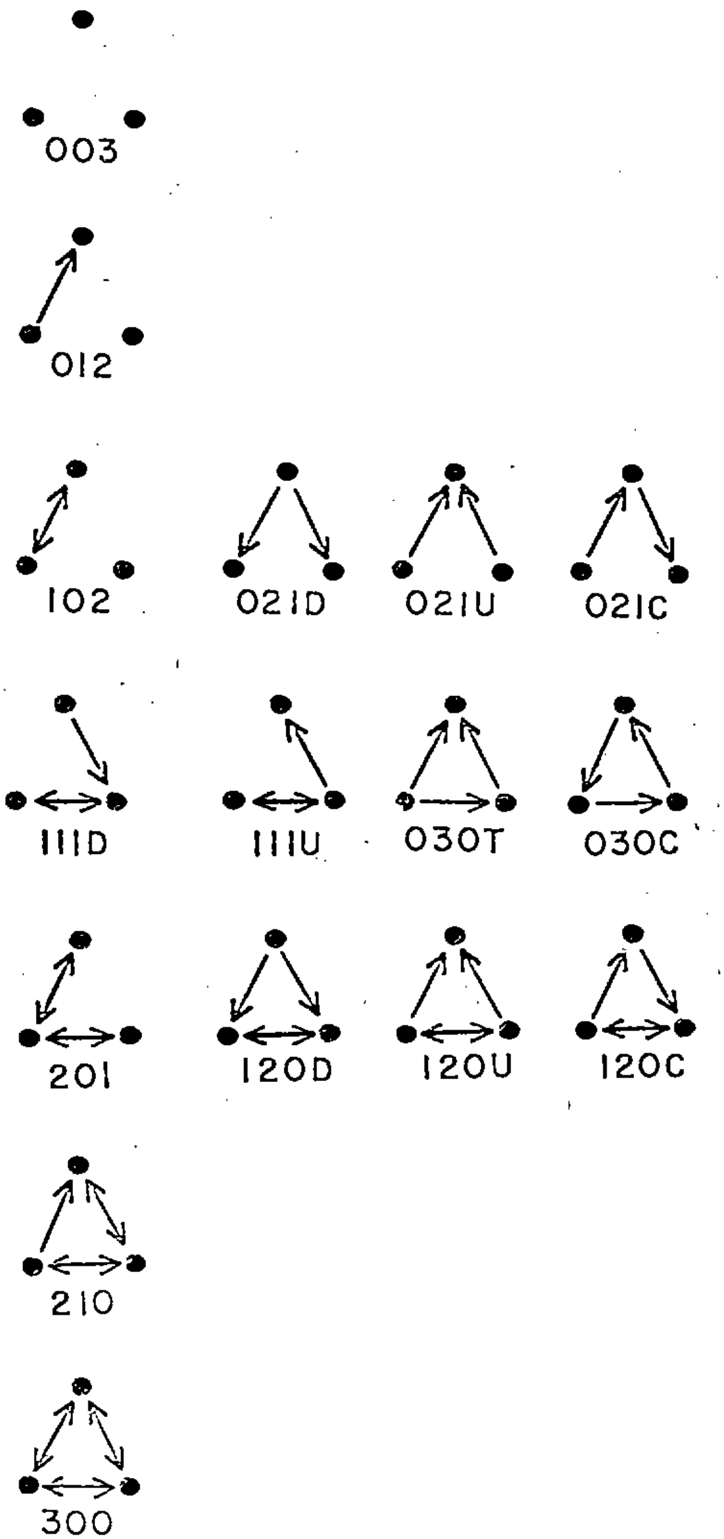
$$\begin{array}{l} 003,\ 012,\ 102,\ 021D,\ 021U,\ 021C,\ 111D,\ 111U \\ 030T,\ 030C,\ 201,\ 120D,\ 120U,\ 120C,\ 210,\ 300. \end{array} \qquad (2.1)$$

The trailing letters U, C, D, and T stand for, respectively, up, cyclic, down, and transitive. The triad census is discussed in more detail in section 4.

3. Random Digraph Distributions

In this section several distributions on digraphs are described. Simple distributions are presented first, followed by conditional uniform distributions of increasing complexity. The probability mass functions

Figure 2.2: The 16 Isomorphism Classes for digraphs with g=3 (i.e., the Triad Types). Triad naming convention: first digit=number of mutual dyads; second digit=number of asymmetric dyads; third digit=number of null dyads; trailing letters further differentiate among triad types. Figure taken from Holland and Leinhardt (1976).



defined here are not used in latter sections of this paper; however, a review of the distributions will enable the reader to better understand the calculation in section 5 of the first two moments of the triad census.

Define D_g as the set of all labeled binary directed graphs on g nodes. A sociomatrix $\underset{\sim}{X}$ will denote the random digraph generated by the distribution of interest. A possible value of $\underset{\sim}{X}$ will be denoted by $\underset{\sim}{x}$, with elements x_{ij}. Most of the distributions will be described in terms of $\underset{\sim}{X}$, whose elements are X_{ij}.

Some of the material presented in this section is borrowed from Holland (1972).

A. Simple Distributions on D_g

D_g may be considered a finite set with $2^{g(g-1)}$ elements. The uniform distribution, U, on D_g, with probability mass function

$$P\{\underset{\sim}{X}=\underset{\sim}{x}\} = (2^{g(g-1)})^{-1} \text{ for all } \underset{\sim}{x} , \tag{3.1}$$

considers all elements of D_g as equally likely. It is perhaps simpler to describe the X_{ij} as independent, Bernoulli random variables with

$$P\{X_{ij}=1\} = \begin{cases} p_{ij} = 1/2 \ ; \ i \neq j \\ 0 \ ; \ i=j \ . \end{cases} \tag{3.2}$$

A digraph is easily simulated under this distribution using (3.2).

The uniform distribution may be generalized to a family of Bernoulli distributions on D_g. Specifically, let

$$P\{X_{ij}=1\} = \begin{cases} p_{ij} \ ; \ i \neq j \\ 0 \ ; \ i=j \end{cases} \tag{3.3}$$

where $0 \leq p_{ij} \leq 1$. This distribution permits some directed lines in the digraph to have greater probabilities of being present than other lines. If $p_{ij}=1/2$ for all $i \neq j$, then U is obtained. After specifying the full set of p_{ij}, these distributions are also easily simulated using uniform pseudo-random numbers.

Several classes of uniform distributions on digraphs can be formed by conditioning U on certain functions of $\underset{\sim}{X}$ fixed at specific values. Let $x_{++} = \sum_{ij} x_{ij}$, the total number of directed lines in the digraph. Define the random variable C as the number of lines in the random sociomatrix $\underset{\sim}{X}$. The simplest conditional uniform distribution conditions on the random variable C.

The $U|C = x_{++}$ is the conditional uniform distribution which gives equal probability to all digraphs with C lines and zero probability to the elements of D_g with $C \neq X_{++}$. It has probability mass function

$$P\{\underset{\sim}{X}=\underset{\sim}{x}\} = \begin{cases} \binom{g(g-1)}{C}^{-1} & \text{if } x_{++}=C \\ 0 & \text{, otherwise.} \end{cases} \tag{3.4}$$

For computer simulation, it is more informative to consider $U|C=x_{++}$ as allocating C directed lines at random to the $g(g-1)$ possible edges without replacement.

B. <u>The U|MAN Distribution</u>

The U|M=m, A=a, N=n distribution is the conditional uniform distribution which puts equal probability on all digraphs in D_g with M=m mutual, A=a asymmetric, and N=n null dyads. Note that unless $m+a+n=\binom{g}{2}$, the subset of digraphs of D_g with the given values of M, A, and N will be empty. The U|MAN distribution has been popularized by Holland and Leinhardt. It provides a large amount of conditioning, and calculations are made more easily than under other conditional distributions.

To find the probability mass function of $\underset{\sim}{X}$ under U|MAN, number the $\binom{g}{2}$ pairs of nodes in the digraph from 1 to $\binom{g}{2}$. Choose <u>m</u> of these numbers at random and without replacement. Mutual dyads are assigned to the m pairs of nodes corresponding to the chosen numbers. From the remaining $\binom{g}{2}$-m numbers, select a numbers also at random and without replacement as asymmetric dyads. The direction of each of these asymmetric dyads is decided randomly (e.g., by tossing a fair coin). The remaining pairs whose numbers have not been chosen, are assigned null dyads. Thus U|MAN is given by

$$P\{X=x\} = \begin{cases} \dfrac{1}{2^a}\left(\dfrac{\binom{g}{2}!}{m!\ a!\ n!}\right)^{-1} & ;\ \text{if } M=m,\ A=a,\ N=n \\ 0 & ;\ \text{otherwise.} \end{cases} \tag{3.5}$$

This distribution will be compared to the more complex uniform conditional distributions on digraphs in section 3-E.

C. The $U|\{X_{i+}\}$ distribution

The $U|\{X_{i+}=r_i\}$ distribution is the uniform distribution conditional on a fixed set of outdegrees. Equal weight is given to all digraphs in D_g with $X_{1+}=r_1$, $X_{2+}=r_2, \ldots, X_{g+}=r_g$. Each r_i may take on all integer values between 0 and (g-1). Sociometric interpersonal preference data may be collected under either a "fixed choice" or a "free choice" procedure. In a fixed choice experiment, the investigator may instruct each member in the group to "Name your four best friends in the group". If each group member fully cooperates then the outdegree of each node is fixed at a specific value. A free choice experiment places no restrictions on the number of individuals chosen by each group member. In either situation, the $U|\{X_{i+}\}$ distribution is very useful in calculations because it allows the investigator to "control" for the outdegree of each node. This outdegree adjustment removes the effects of the procedure used to gather the data.

The $U|\{X_{i+}=r_i\}$ distribution has probability mass function

$$P\{\underset{\sim}{X}=\underset{\sim}{x}\} = \begin{cases} \prod_i \binom{g-1}{r_i}^{-1} ; & \text{if } x_{i+}=r_i \text{ for all } i, \\ 0 ; & \text{otherwise.} \end{cases} \tag{3.6}$$

$U|\{X_{i+}\}$ may be generated by regarding each row of $\underset{\sim}{X}$ as stochastically independent. If $\underset{\sim}{X}_i$ denotes the i^{th} row of $\underset{\sim}{X}$, then r_i ones are distributed at random and without replacement to the $(g-1)$ possible locations in $\underset{\sim}{X}_i$ (remember $X_{ii}=0$). When all the r_i are equal to a fixed value, r, (3.6) simplifies to

$$P\{\underset{\sim}{X}=\underset{\sim}{x}\} = \begin{cases} \binom{g-1}{r}^{-g} & ; \text{ if } x_{i+}=r,\ i=1,\ 2,\ \ldots,\ g \\ 0 & ; \text{ otherwise.} \end{cases} \tag{3.7}$$

D. <u>The $U|\{X_{+j}\}$ distribution</u>

The $U|\{X_{+j}=c_j\}$ distribution is identical to the $U|\{X_{i+}\}$ distribution except that the conditioning is on the set of indegrees of $\underset{\sim}{X}$. The probability mass function of $U|\{X_{+j}\}$ is

$$P\{\underset{\sim}{X}=\underset{\sim}{x}\} = \begin{cases} \prod_j \binom{g-1}{c_j}^{-1} & ; \text{ if } x_{+j}=c_j \text{ for all } j \\ 0 & ; \text{ otherwise.} \end{cases} \tag{3.8}$$

This distribution may be simulated in a manner similar to $U|\{X_{i+}\}$ by regarding the columns of $\underset{\sim}{X}$ as stochastically independent. Conditioning on the set of indegrees of the digraph is not as useful as conditioning on the outdegrees; however, using the calculations of the means and variances of the triad census developed in section 5 of this paper, the $U|\{X_{+j}\}$ distribution is helpful in approximating more complex conditional uniform distributions.

E. Complex Conditional Distributions

There are several highly important conditional uniform distributions that are so complex that no simple way exists for generating random digraphs with these distributions. This section briefly discusses several of these.

The $U|\{X_{i+}\},\{X_{+j}\}$ distribution simultaneously conditions on both the indegrees and the outdegrees of the digraph. All digraphs with the specified values of $\{X_{i+}\}$ and $\{X_{+j}\}$ are equally likely. This distribution is extremely important in sociometric data analysis, since it controls for both the choices made by each group member and choices received. Ford and Fulkerson (1962) give necessary and sufficient conditions for the existence of at least one element in D_g with the specified outdegrees and indegrees. Unfortunately, no one has been able to develop a sophisticated technqiue to simulate this distribution.

Also worth noting is the $U|M,\{X_{i+}\}$ distribution. Since $\sum_i X_{i+} = C$ and C-2M = A, this distribution also controls the total number of choices made and the number of asymmetric choices. Little is known about $U|M,\{X_{i+}\}$, even though it is important.

Perhaps the most important distribution in sociometric data analysis is the $U|M,\{X_{i+}\},\{X_{+j}\}$ distribution. Its value derives from the fact that it controls for choices-made, choices-received, and mutuality. As with $U|M,\{X_{i+}\}$, and $U|\{X_{i+}\}, \{X_{+j}\}$, there is no sophisticated way to generate random sociomatrices under this distribution.

To reiterate, there are many possible distributions on D_g. At the present time, only a few of these distributions are fully understood. The U|MAN distribution was chosen by Holland and Leinhardt (1970) because it best approximated the $U|M,\{X_{i+}\}, \{X_{+j}\}$ distribution. Unfortunately, it does not control for either the set of indegrees or outdegrees in the digraph. The attitude taken in this paper is that although U|MAN is the most complex uniform distribution in use at present, sociometric data analysis should not overlook the information to be gained from considering the $U|\{X_{i+}\}$ and $U|\{X_{+j}\}$ distributions.

4. The Triad Census

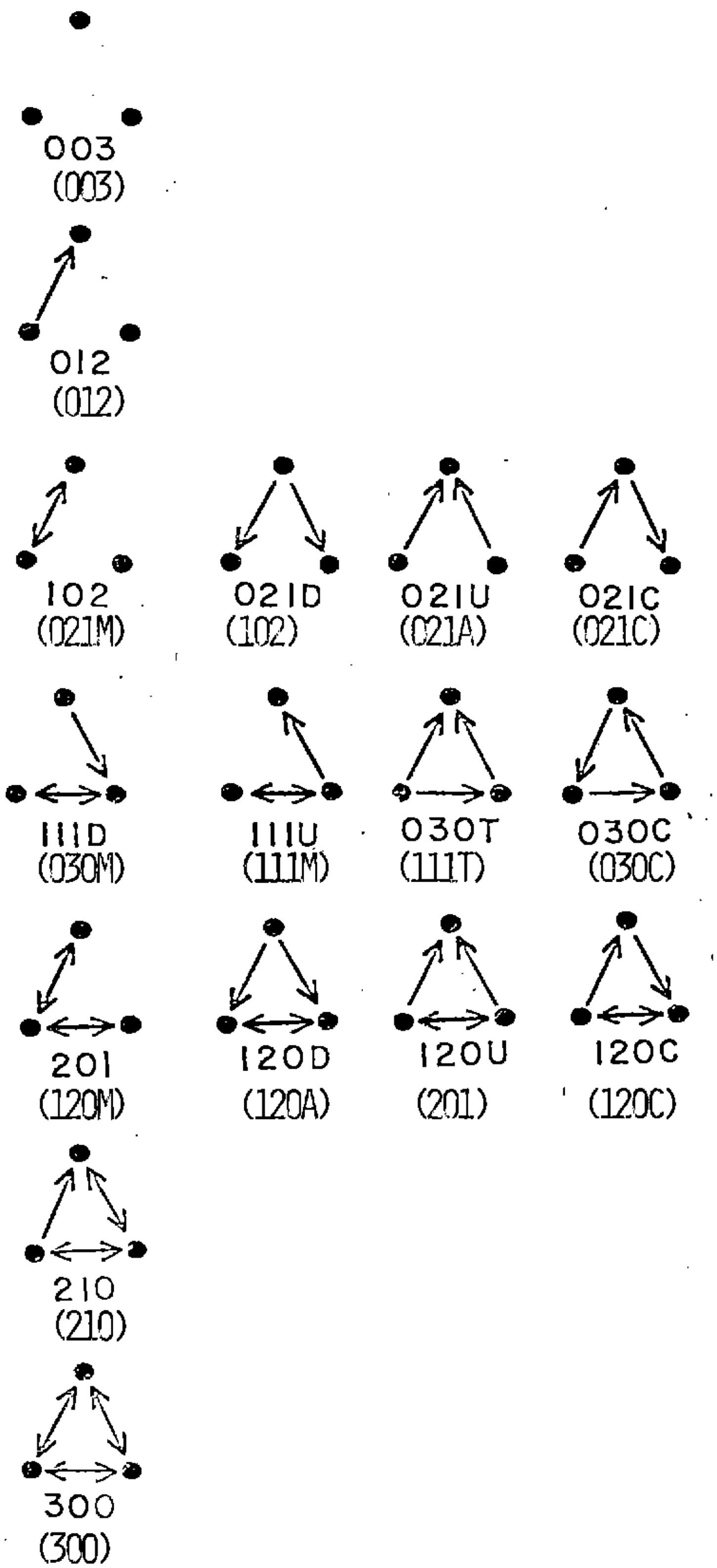
The triad census has been effectively used in the analysis of sociometric data, reducing the entire sociomatrix $\underset{\sim}{X}$ to a set of 16 summary statistics. In sections B and C, summarizing some of the earlier work of Holland and Leinhardt, various aspects of the triad census are discussed.

A. Naming Conventions

The U|MAN distribution has been the only distribution employed for computing the first and second moments of the triad census. Holland and Leinhardt (1971), Davis (1970), and Davis and Leinhardt (1972) have used the first 2 moments of the triad census for testing structural hypothesis concerning social networks. Consequently, the convention for labeling the 16 components of the triad census utilized in Figure 2.2 is well established in the literature, but impractical in the discussion of the $U|\{X_{i+}\}$ distribution. The labels should communicate the outdegree of each node in the triad, rather the number of mutual, asymmetric and null dyads. Figure 4.1 illustrates the 16 isomorphism classes for 3-subgraphs with both the U|MAN and $U|\{X_{i+}\}$ designations. The trailing letters in the figure, M, A, C, and T, respectively, stand for mutual, asymmetric, cyclic and transitive.

So as not to confuse the reader familiar with the U|MAN naming convention, the $U|\{X_{i+}\}$ labeling scheme will not be employed in this paper. However, a knowledge of this new scheme will aid in the interpretation of the calculations in section 5.

Figure 4.1: The 16 isomorphism classes for 3-subgraphs with $U|\{X_{i+}\}$ labeling. The label directly under each triad is defined by the number of mutuals, asymmetric, and null dyads in the triad (see Figure 2.2). The label in parentheses is based on the outdegree of each node in the triad: first digit = number of nodes with outdegree equal to 2; second digit = number of nodes with outdegree equal to 1; third digit = number of nodes with outdegree equal to 0; trailing letters further differentiate triad types.



B. Linear Combinations of the Triad Census

Many quantities can be determined by taking linear combinations of the 16 triad frequencies of $\underset{\sim}{T}$. Suppose $\underset{\sim}{\ell}$ is a vector with 16 elements. A linear combination of $\underset{\sim}{T}$ will be denoted

$$\underset{\sim}{\ell}' \underset{\sim}{T} = \sum_u \ell_u T_u \tag{4.1}$$

where the subscript u runs over the 16 triad types enumerated in (2.1).

Denote the variance of the $\{X_{i+}\}$ by S^2_{out} and the variance of the $\{X_{+j}\}$ by S^2_{in} where

$$S^2_{out} = (1/g) \sum_i (X_{i+} - \overline{X})^2$$

and

$$S^2_{in} = (1/g) \sum_j (X_{+j} - \overline{X})^2$$

Among the quantities calculable from $\underset{\sim}{T}$ are g, M, A, N, C and S^2_{in}, and S^2_{out}, the variances of the indegrees and outdegrees of the digraph. To compute g, recall that there are $\binom{g}{3}$ triads in a digraph with g nodes. Therefore, if $\underset{\sim}{\ell}'=\underset{\sim}{e}=(1,1,\ldots,1)$, then

$$\underset{\sim}{e} \underset{\sim}{T} = \sum_u T_u = \binom{g}{3} \tag{4.2}$$

and g may be found by finding the single real root of the cubic equation

$$\left[\binom{g}{3} - \sum_u T_u = 0\right] .$$

Holland and Leinhardt (1976) discuss in detail the calculations involved in finding the values of M, A, N, C, S^2_{in} and S^2_{out} from the triad census. These details will not be reproduced here. Table (4.1) presents the various $\underset{\sim}{\ell}$ vectors needed to calculate these quantities, and the last row of the table gives $\sum_u \ell_u T_u$ for each vector.

Table 4.1: Selected Weighting Vectors.

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

The vectors $\underset{\sim}{B}_{in}$ and $\underset{\sim}{B}_{out}$ shown in the last two columns of Table (4.1), are used to calculate S^2_{in} and S^2_{out}. Holland and Leinhardt (1976) prove

$$S^2_{in} = (2/g) \sum_u B_{in,u} T_u - \overline{X}(\overline{X}-1) \tag{4.3}$$

and

$$S^2_{out} = (2/g) \sum_u B_{out,u} T_u - \overline{X}(\overline{X}-1) \tag{4.4}$$

where $\overline{X}=C/g=X_{++}/g$, the average number of choices per group member. Since C can be expressed as a linear combination of triad frequencies, so can $\overline{X}$. Thus S^2_{in} and S^2_{out} are easily calculated using linear combinations of $\underset{\sim}{T}$.

C. <u>Testing Structural Hypothesis</u>

Perhaps the most important use of the triad census is testing propositions about local structure in a sociomatrix. Holland and Leinhardt originally proposed the triad census to test the proposition that interpersonal choices tended to transitive; i.e., if person i chooses person j and person j chooses person k, then person i <u>should</u> choose person k. In their 1970 paper, the following triads were classified as "intransitive": 021C, 030C, 111D, 111U, 120C, 201, and 210. The occurrence of any of these 7 triad types indicated that the group violated the transitivity hypothesis, since each of these triads had at least one intransitivity. Holland and Leinhardt developed a measure, $\tau(\underset{\sim}{\ell})$, which was used as a "transitivity index".

They define $\tau(\underset{\sim}{\ell})$ as follows:

$$\tau(\underset{\sim}{\ell}) = \frac{\underset{\sim}{\ell}'\underset{\sim}{T} - \underset{\sim}{\ell}'\underset{\sim}{\mu}_T}{\sqrt{\underset{\sim}{\ell}' \underset{\sim}{\Sigma}_T \underset{\sim}{\ell}}} \tag{4.5}$$

where $\underset{\sim}{\ell}$ is the weighting vector that counts the number of intransitive triads, and $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$ are the mean and covariance matrix of $\underset{\sim}{T}$ as computed using the U|MAN distribution. A computer simulation showed that for large g, $\tau(\underset{\sim}{\ell})$ was approximately distributed as a standard Gaussian random variable.

Holland and Leinhardt (1976) generalize this procedure so that any structural hypothesis may be tested. The triads that will inviolate the hypothesis in question merely need to be discovered, and the corresponding weighting vector found. Mazur's (1971) proposition concerning "close friends disagreeing" is discussed and $\tau(\underset{\sim}{\ell})$ for this hypothesis is computed for 408 sociomatrices.

5. Moments of a Triad Census

In this section the means, variances, and covariances of the triad census of a random digraph are given, assuming that the digraph is distributed according to the $U|\{X_{i+}\}$ distribution. Also discussed are the formulas for the above quantities assuming that the digraph follows the $U|\{X_{+j}\}$ distribution. The section concludes with a consideration of the moments of a linear combination of the triad census under a general distribution, and the development of a method useful in approximating the moments of the triad census under the uniform conditional distributions mentioned in Section 3 as too difficult to work with.

A. Notation and Few Derivations

There are $\binom{g}{3}$ triads in a digraph with g nodes. Let $\underline{K}$ and $\underline{L}$ be subscripts that refer to the $\binom{g}{3}$ distinct triads of a given digraph. The letters $\underline{u}$ and $\underline{v}$ will refer to two of the 16 isomorphism classes of the triad census, discussed in Section 4-A. This section concludes with the formula to be used in section C for calculating the variances and covariances of the 16 triad types.

Define the indicator variables $T_K(u)$ as follows:

$$T_K(u) = \begin{cases} 1 & \text{if triad } K \text{ is of isomorphism class } u \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

T_u, the number of triads of isomorphism class u is found by summing $T_K(u)$ over all possible triads:

$$T_u = \sum_K T_K(u) \quad . \tag{5.2}$$

Notation is also needed for the various probabilities that arise in the calculations given here. Define

$$p_K(u) = P\{\text{triad } K \text{ is of type } u\} = P\{T_K(u) = 1\} \tag{5.3}$$

and

$$\begin{aligned} p_{KL}(u,v) &= P\{\text{triad } K \text{ is of type } u \text{ and triad } L \text{ is of type } v\} \\ &= P\{T_K(u) = 1 \text{ and } T_L(v) = 1\} \quad . \end{aligned} \tag{5.4}$$

Formula (5.4) is a joint probability involving triads K and L. Consider the number of nodes that triads K and L have in common. Let

$$|K \cap L| = \text{number of nodes that } K \text{ and } L \text{ have in common.} \tag{5.5}$$

Obviously, $|K \cap L|$ takes on the values 0, 1, 2, and 3. If $|K \cap L|=0$, the two triads are disjoint, and if $|K \cap L|=3$, triads K and L are identical. Let

$$p_{KL}^{(j)}(u,v) = P\{T_K(u) = 1,\ T_L(v) = 1, \text{ and } |K \cap L| = j\} \ , \quad j = 0, 1, 2, 3. \tag{5.6}$$

For a fixed u and v, the four probabilities defined in (5.6) for varying j provide a decomposition of the joint probability (5.4) as follows:

$$P_{KL}(u,v) = p^{(0)}_{KL}(u,v) + p^{(1)}_{KL}(u,v) + p^{(2)}_{KL}(u,v) + p^{(3)}_{KL}(u,v). \qquad (5.7)$$

Formula (5.7) is important for calculations involving variances and covariances.

B. Expressions for Means, Variances, and Covariances

Holland and Leinhardt (1976) give formulae for the first two moments of the triad census under a general distribution using the average values of $p_K(u)$, $p^{(0)}_{KL}(u,v)$, $p^{(1)}_{KL}(u,v)$, $p^{(2)}_{KL}(u,v)$ and $p^{(3)}_{KL}(u,v)$. The moments may also be given in terms of summations, which defines the $U|\{X_{i+}\}$ distribution more easily. Theorems 1 and 2 given in this paper are equivalent to Corollary 1 presented in Section 5.A of Holland and Leinhardt (1976).

Theorem 1: *Assuming that a random digraph is generated by some random digraph distribution, the first mement of T_u is given by:*

$$E(T_u) = \sum_K p_K(u).$$

Proof: Note, by (5.2), that

$$E(T_u) = E(\sum_K T_K(u)) = \sum_K E(T_K(u)).$$

Since $T_K(u)$ is an indicator variable, it follows that

$$E(T_u) = \sum_K E(T_K(u)) = \sum_K p_K(u). \quad \text{Q.E.D.}$$

Define $\underset{\sim}{\Sigma}_T$ as the covariance matrix of the 16 components of the triad census. The (u,v) element of $\underset{\sim}{\Sigma}_T$ is denoted $\sigma_T(u,v)$. For Theorem 2, let $\sum_{|K\cap L|=j}$ denote a summation over all pairs of triads with j nodes in common, where j = 0, 1, 2, or 3.

Theorem 2: *With the same assumptions as in Theorem 1, the (16x16) covariance matrix* $\underset{\sim}{\Sigma}_T$, *of the triad census has representative terms:*

$$\sigma_T(u,u) = \sum_{j=0}^{3} \sum_{|K\cap L|=j} p_{KL}^{(j)}(u,u) - (\sum_K p_K(u))^2$$

$$\sigma_T(u,v) = \sum_{j=0}^{3} \sum_{|K\cap L|=j} p_{KL}^{(j)}(u,v) - (\sum_K p_K(u))(\sum_K p_K(v)).$$

Proof: $\sigma_T(u,v)$ is the covariance between T_u, and T_v. Since T_u and T_v are sums of indicator variables, the proof of this theorem is straightforward, and will not be given here. The reader is referred to Theorem 1 in Holland and Leinhardt (1976) for an analogous proof. Q.E.D.

From these theorems, the quantities that must be computed to find

$\underset{\sim}{\mu}_T' = [E(T_{003}, E(T_{021}), \ldots , E(T_{300})]'$ and

$\underset{\sim}{\Sigma}_T$ are:

$$\sum_K p_K(u) \; ; \tag{5.8}$$

$$\sum_{|K\cap L|=0} p_{KL}^{(0)}(u,v), \quad \text{for all } u,v \; ; \tag{5.9}$$

$$\sum_{|K\cap L|=1} p_{KL}^{(1)}(u,v), \quad \text{for all } u,v \; ; \tag{5.10}$$

$$\sum_{|K\cap L|=2} p_{KL}^{(2)}(u,v), \quad \text{for all } u,v \; ; \tag{5.11}$$

and

$$\sum_{|K\cap L|=3} p_{KL}^{(3)}(u,v), \quad \text{for all } u,v \; . \tag{5.12}$$

Note that if u=v, $p_{KL}^{(3)}(u,v)$ reduces to $p_K(u)$, and if $u\neq v$, $p_{KL}^{(3)}(u,v)=0$.

C. Derivations of Probabilities Under $U|\{X_{i+}\}$

A random digraph with the U|MAN distribution is characterized by certain properties that greatly simplify the calculation of the quantities (5.8)-(5.12). Consider the triad 030T illustrated in Figure 5.1a.

The 3 nodes of this triad have been labeled, in a clockwise manner beginning at the lower left vertex, node i, node j, and node k. Note the same triad 030T in Figure 5.1b where the nodes j and k have been interchanged. Under the U|MAN distribution, the only relevant features of this triad, and in fact all triads, are the numbers of null, asymmetric, and mutual dyads. Thus, the two triads in Figures 5.1a and 5.1b while distinct by a permutation of the labels attached to the nodes, are considered identical by the U|MAN distribution. The U|MAN distribution on a digraph is "homogeneous" in the sense that it is invariant under permutations of the labels given to the nodes.

Now, consider the outdegree of each node in the two triads shown in Figure 5.1. The two triads have one node with the same outdegree (node i). In the top triad, node j has outdegree 0 and node k, outdegree 1. In the bottom triad, node j has outdegree 1 and node k has outdegree 0. These triads, under $U|\{X_{i+}\}$, are obviously not invariant under permutations. In general, the outdegree of each node changes with a rearrangement of the node labels.

Due to the lack of homogeneity of a digraph under $U|\{X_{i+}\}$, it is necessary to examine every triad, and every pair of triads with 0, 1, and 2 nodes in common, to compute each term of the sums (5.8) - (5.11). This is in stark contrast to the same calculations under U|MAN given by Holland and Leinhardt (1974). With the U|MAN distribution, the probabilities defined in (5.3) and (5.6) do not depend on K or L. so that

Figure 5.1: Triad 030T as illustration of a homogeneous digraph.

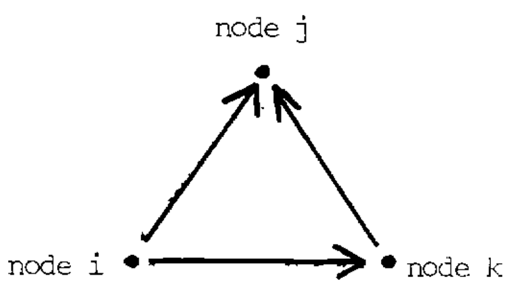


Figure 5.1a

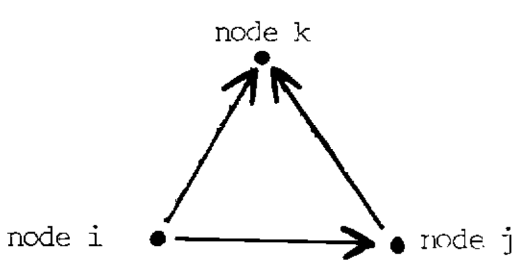


Figure 5.1b

$$p_K(u) \equiv p(u) = P\{\text{triad involving nodes 1, 2, 3 is of type } u\} \quad (5.13)$$

and

$$p_{KL}^{(j)}(u,v) \equiv p^{(j)}(u,v), \text{ for } j=0,\ 1, \text{ and } 2. \quad (5.14)$$

Finally, under $U|MAN$,

$$p^{(0)}(u,v) = p^{(1)}(u,v) \quad (5.15)$$

A relationship which does not hold under $U|\{X_{i+}\}$.

To illustrate the calculations of $p_K(u)$, $p_{KL}^{(0)}(u,v)$, $p_{KL}^{(1)}(u,v)$, and $p_{KL}^{(2)}(u,v)$ under the $U|\{X_{i+}\}$ distribution on digraphs, some additional notation will be needed. Let the ordered triple (i,j,k) refer to the nodes in triad K, and the triple (ℓ,m,n) refer to the nodes in triad L.

First consider the quantities $p_K(u)$ needed to compute $E(T_u)$. Define the variable $C_i(K)$ as the outdegree of node i in triad K. Obviously, $C_i(K)$ equals either 0, 1, or 2. Let $c_i(K)$ equal the actual value of $C_i(K)$. The variables $C_j(K)$ and $C_k(K)$ are defined in a similar way. When no confusion may arise, the triad in parantheses of $C_i(K)$ and $C_j(K)$ will be dropped. Let

$$[c_i,c_j,c_k] = P\{C_i=c_i,\ C_j=c_j,\ C_k=c_k\} \quad (5.16)$$

For a fixed (i,j,k), the possible values of $[c_i,c_j,c_k]$ form a 3x3x3 array of probabilities, with one dimension each for C_i, C_j, and C_k. There are $\binom{g}{3}$ such three dimensional tables. It will be convenient to abbreviate $[c_i(K),c_j(K),c_k(K)]$ as $[c_i,c_j,c_k]_K$.

It is not difficult to compute $[c_i,c_j,c_k]$. Specifically,

$$[c_i, c_j, c_k] = \frac{\binom{2}{c_i}\binom{g-3}{X_i - c_i}}{\binom{g-1}{X_i}} \cdot \frac{\binom{2}{c_j}\binom{g-3}{X_j - c_j}}{\binom{g-1}{X_j}} \cdot \frac{\binom{2}{c_k}\binom{g-3}{X_k - c_k}}{\binom{g-1}{X_k}} \tag{5.17}$$

Where X_i is a shorthand notation for X_{i+}, the outdegree of node i in the digraph. If c_i should exceed X_i (or c_j exceed X_j or c_k exceed X_k), (5.17) is identically zero. The notation $\sum_K [c_i, c_j, c_k]_K$ will refer the summation of the (c_i, c_j, c_k) cells over all $\binom{g}{3}$ three dimensional tables. These 27 quantities, formed by collapsing all the $\binom{g}{3}$ tables into one table, will be used to calculate the 16 components of $\underset{\sim}{\mu}_T$.

An example will illustrate this calculation. Consider the triad 021U. With a fixed triple of nodes (i,j,k), there are three ways that this triad may be "oriented", as shown in Figure 5.2.

Examination of the three orientations and the outdegree of each node within each orientation yields the following expression for $E(T_{021U})$:

$$E(T_{021U}) = \sum_K [1,0,1]_K + \sum_K [1,1,0]_K + \sum_K [0,1,1]_K. \tag{5.18}$$

Table 5.1 gives the expected values of the 16 isomorphism classes.

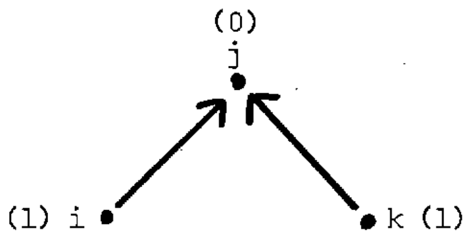
Consider $p_{KL}^{(0)}(u,v)$, the joint probabilities of triad K and triad L assuming $|K \cap L| = 0$. Triads K and L are disjoint; thus

$$\sum_{|K\cap L|=0} p_{KL}^{(0)}(u,v) = \sum_{K} \sum_{\substack{L \\ |K\cap L|=0}} p_K(u)\, p_L(v) = \sum_K p_K(u) \sum_{\substack{L \\ L \neq K}} p_L(v) \tag{5.19}$$

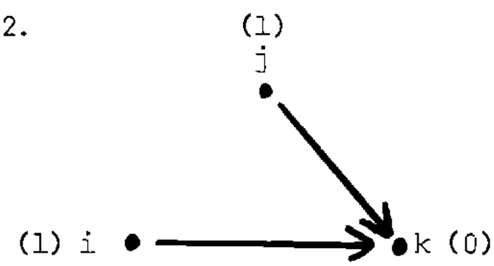
The quantities $[c_i, c_j, c_k]_K$ used to calculate the expected values of the triad census are also employed in the calculation of the probabilities $p_{KL}^{(0)}(u,v)$.

Figure 5.2: The 3 possible orientations of triad 021U. The outdegree of each node in the triad is given in parantheses next to the node.

Orientation 1.



Orientation 2.



Orientation 3.

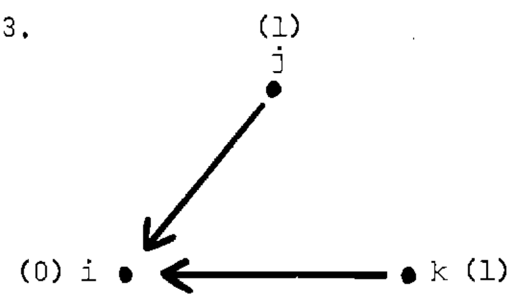


Table 5.1: Expected Values of Triad Census under $U|\{X_{i+}\}$. Sums are over all possible triads.

Triad	Expected Value
003	$\Sigma\{[0,0,0]_K\}$
012	$\Sigma\{2[1,0,0]_K + 2[0,1,0]_K + 2[0,0,1]_K\}$
102	$\Sigma\{[1,1,0]_K + [1,0,1]_K + [0,1,1]_K\}$
021D	$\Sigma\{[2,0,0]_K + [0,2,0]_K + [0,0,2]_K\}$
021U	$\Sigma\{[1,1,0]_K + [1,0,1]_K + [0,1,1]_K\}$
021C	$\Sigma\{2[1,1,0]_K + 2[1,0,1]_K + 2[0,1,1]_K\}$
111D	$\Sigma\{6[1,1,1]_K\}$
111U	$\Sigma\{[2,1,0]_K + [2,0,1]_K + [1,2,0]_K + [1,0,2]_K + [0,2,1]_K + [0,1,2]_K\}$
030T	$\Sigma\{[2,1,0]_K + [2,0,1]_K + [1,2,0]_K + [1,0,2]_K + [0,2,1]_K + [0,1,2]_K\}$
030C	$\Sigma\{2[1,1,1]_K\}$
201	$\Sigma\{[2,1,1]_K + [1,2,1]_K + [1,1,2]_K\}$
120D	$\Sigma\{[2,1,1]_K + [1,2,1]_K + [1,1,2]_K\}$
120U	$\Sigma\{[2,2,0]_K + [2,0,2]_K + [0,2,2]_K\}$
120C	$\Sigma\{2[2,1,1]_K + 2[1,2,1]_K + 2[1,1,2]_K\}$
210	$\Sigma\{2[2,2,1]_K + 2[2,1,2]_K + 2[1,2,2]_K\}$
300	$\Sigma\{[2,2,2,]_K\}$

The computations of $p_{KL}^{(1)}(u,v)$ and $p_{KL}^{(2)}(u,v)$ are quite involved and will be discussed only briefly here. The (16x16) tables of $\sum_{|K\cap L|=1} p_{KL}^{(1)}(u,v)$ and $\sum_{|K\cap L|=2} p_{KL}^{(2)}(u,v)$ will not be presented here in order to save space. First consider the probabilities $p_{KL}^{(1)}(u,v)$. Node k will be designated as the common node. Triad L contains nodes (k,ℓ,m). Let

$$[c_k,c_i,c_j,c_\ell,c_m] = \mathbf{P}\{C_k=c_k,\ C_i=c_i,\ C_j=c_j,\ C_\ell=c_\ell,\ C_m=c_m\} \tag{5.20}$$

C_i, C_j, C_ℓ, and C_m are defined identically as in (5.16), while C_k takes on the values 0, 1, 2, 3, and 4. (There are five variables within the brackets in formula (5.20) because node k is common to both triads K and L; consequently, there are five nodes, not six, to consider.) There are $g\binom{g-1}{2}\binom{g-3}{2}$ distinct 5x3x3x3x3 arrays of the $[c_k,c_i,c_j,c_\ell,c_m]$ probabilities. By "summing" all of these five dimensional arrays there remain 405 quantities necessary in the calculation of $\sum_{|K\cap L|=1} p_{KL}^{(1)}(u,v)$. The notation $\sum_{|K\cap L|=1} [c_k,c_i,c_j,c_\ell,c_m]_{KL}$ will refer to the summation of the (c_k,c_i,c_j,c_ℓ,c_m) cells over all tables. The value of $[c_k,c_i,c_j,c_\ell,c_m]$ is also a product of hypergeometric probabilities similar to (5.17).

An example will help to illustrate. Figure (5.3) shows all 9 possible orientations of triad 120D and triad 021U with one node (node k) in common. Table (5.2) lists all the orientations, and the outdegree of each node in each orientation.

From Table (5.2), after rearranging the orientations to have decreasing values for C_k, $\sum_{|K\cap L|=1} p_{KL}^{(1)}(120D,\ 021U)$ is as follows:

Figure 5.3: The 9 Orientations of Triads 120D and 021U with one node in common.

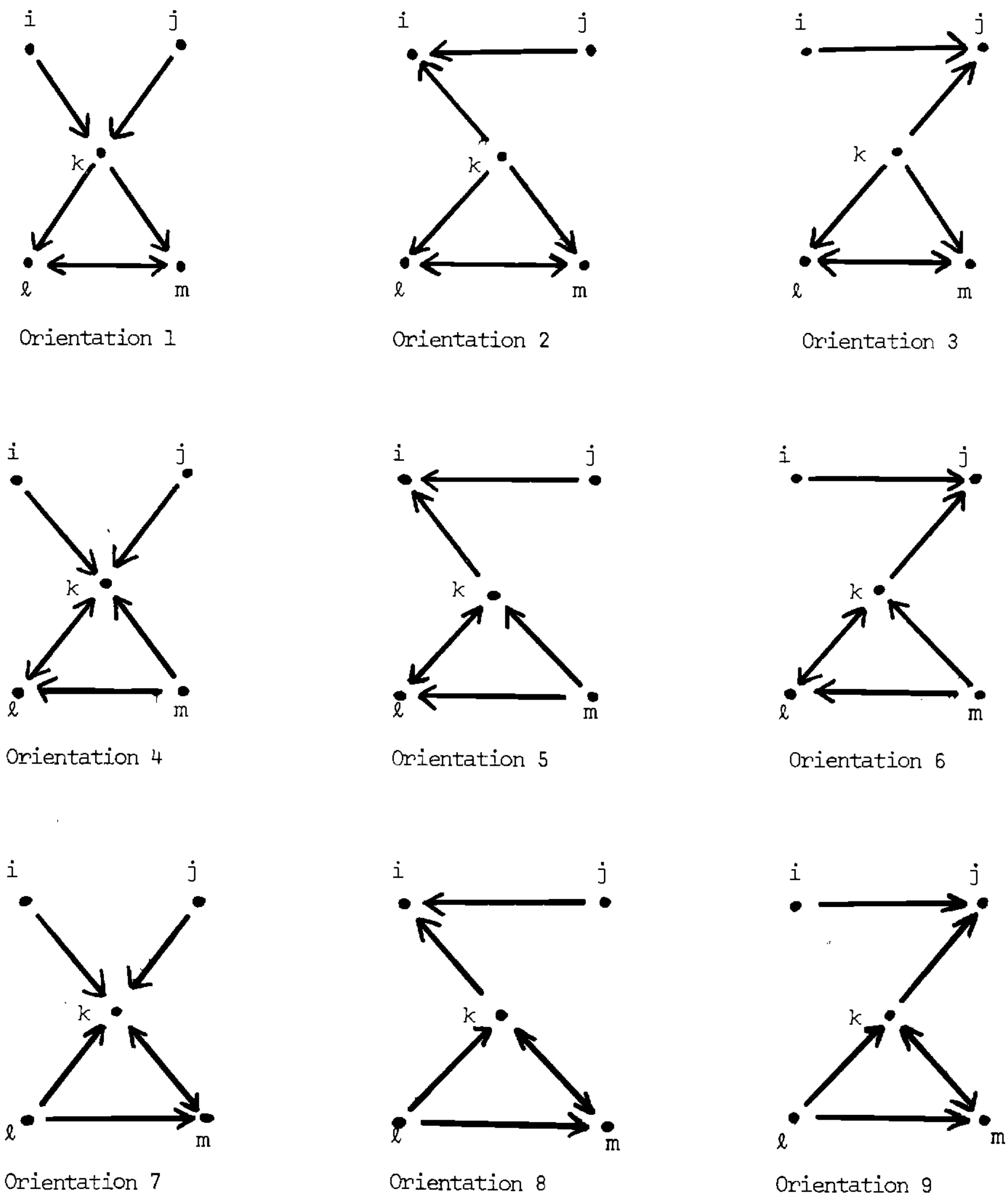


Table 5.2: The outdegree of each node of the 9 orientations of Triads 120D and 021U with one node in common.

Orientation	C_k	C_i	C_j	C_ℓ	C_m
1	2	1	1	1	1
2	3	0	1	1	1
3	3	1	0	1	1
4	1	1	1	1	2
5	2	0	1	1	2
6	2	1	0	1	2
7	1	1	1	2	1
8	2	0	1	2	1
9	2	1	0	2	1

$$\sum_{|K\ L|=1} p_{KL}^{(1)}(120D,\ 120U) = \sum_{|K\ L|=1} \{[3,1,0,1,1]_{KL} + [3,0,1,1,1]_{KL} + [2,1,1,1,1]_{KL} + [2,1,0,2,1]_{KL} + [2,1,0,1,2]_{KL} + [2,0,1,2,1]_{KL} + [2,0,1,1,2]_{KL} + [1,1,1,2,1]_{KL} + [1,1,1,1,2]_{KL}\} \tag{5.21}$$

The remaining probabilities involving two triads with one node in common are found in a similar way.

Lastly, consider the probabilities $p_{KL}^{(2)}(u,v)$, the joint probabilities of triads K and L with 2 nodes in common. Nodes j and k are the common nodes, so that triad L contains nodes (j,k,ℓ). Let

$$[c_j,c_k,c_i,c_\ell] = P\{C_j=c_j,\ C_k=c_k,\ C_i=c_i,\ C_\ell=c_\ell\} \tag{5.22}$$

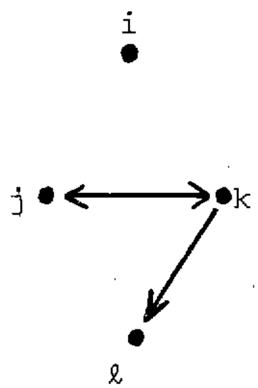
where C_i and C_ℓ equal 0, 1, or 2, and C_j and C_k take on the values 0, 1, 2 or 3. There are $\binom{g}{2}\binom{g-2}{2}$ distinct 4x4x3x3 arrays of these probabilities. As before, by summing all of these four dimensional arrays, the 144 quantities used to calculate $\sum_{|K\cap L|=2} p_{KL}^{(2)}(u,v)$ are found. The notation $\sum_{|K\cap L|=2} [c_j,c_k,c_i,c_\ell]_{KL}$ will be used here to indicate the summation.

Also note that

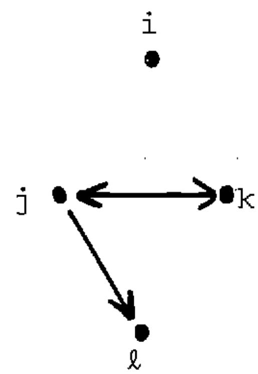
$$[c_j,c_k,c_i,c_\ell] = \frac{\binom{3}{c_j}\binom{g-4}{X_j-c_j}}{\binom{g-1}{X_j}} \cdot \frac{\binom{3}{c_k}\binom{g-4}{X_k-c_k}}{\binom{g-1}{X_k}} \cdot \frac{\binom{2}{c_i}\binom{g-3}{X_i-c_i}}{\binom{g-1}{X_i}} \cdot \frac{\binom{2}{c_\ell}\binom{g-3}{X_\ell-c_\ell}}{\binom{g-1}{X_\ell}} \tag{5.23}$$

Again, an example is helpful. Figure (5.4) illustrates the 6 possible positions for the triads 102 and 111U with nodes j and k in common.

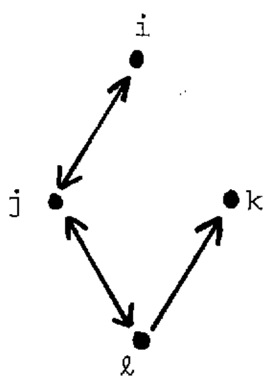
Figure 5.4: The 6 orientations of Triads 102 and 111U with two nodes in common.



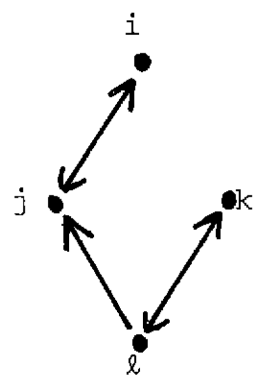
Orientation 1



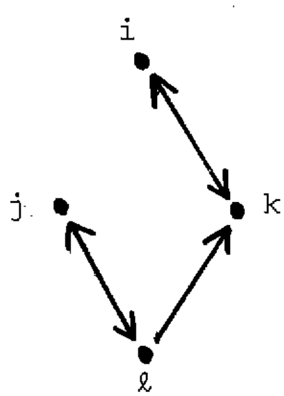
Orientation 2



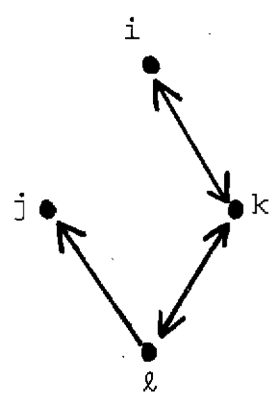
Orientation 3



Orientation 4



Orientation 5



Orientation 6

Recording the outdegree of each node as in Table (5.2) results in the following:

$$\sum_{|K \cap L|=2} P_{KL}^{(2)}(102,\ 111U) = \sum_{|K \cap L|=2} \{[2,1,0,0]_{KL} + [2,0,1,2]_{KL} + [1,2,0,0]_{KL} + 2[1,1,1,2]_{KL} + [0,2\ ,1,2]_{KL}\} \quad . \tag{5.24}$$

The calculation of $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$ under U X_{i+} have been written in FORTRAN code. The calculations performed in Section 6 were done on the TROLL interactive computing system, maintained by the Computer Research Center for Economics and Management Science of the National Bureau of Economic Research.

D. Moments of a Triad Census under $U|\{X_{+j}\}$

Once the machinery necessary for finding $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$ assuming the $U|\{X_{i+}\}$ distribution has been developed, it is quite simple to determine the same first two moments under the $U|\{X_{+j}\}$ distribution. Take the set of indegrees of the digraph which are to be regarded as fixed, and assume that this set of g numbers is actually a set of outdegrees of the digraph.

With the above assumption, three pairs of the 16 isomorphism classes will have to be interchanged. The other 10 triads are invariant under this forced reversal of all directed lines. The three pairs are

021D and 021U,
111D and 111U, (5.25)
120D and 120U,

i.e. all the "downs" become "ups" and vice versa. Thus, if $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$ are calculated under $U|\{X_{i+}\}$ by assuming the set of indegrees to be the conditioning set of outdegrees, only several pairs of rows of $\underset{\sim}{\mu}_T$ and several rows and columns of $\underset{\sim}{\Sigma}_T$ need to be switched. These rows (of $\underset{\sim}{\mu}_T$) or rows and columns (of $\underset{\sim}{\Sigma}_T$) are as follows:

Row 4 (and Column 4 of $\underset{\sim}{\Sigma}_T$) interchanged with
Row 5 (and Column 5 of $\underset{\sim}{\Sigma}_T$),
Row 7 (and Column 7 of $\underset{\sim}{\Sigma}_T$) interchanged with
Row 8 (and Column 8 of $\underset{\sim}{\Sigma}_T$),

and

Row 12 (and Column 12 of $\underset{\sim}{\Sigma}_T$) interchanged with
Row 13 (and Column 13 of $\underset{\sim}{\Sigma}_T$).

E. Moments of a Linear Combination of a Triad Census

After $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$ have been computed with either the U|MAN, $U|\{X_{i+}\}$, or $U|\{X_{+j}\}$ random digraph distributions, it is very simple to calculate the moments of any linear combination of $\underset{\sim}{T}$. If $\underset{\sim}{\ell}'T$ and $\underset{\sim}{s}'T$ are any two linear combinations of the triad census, then

$$E(\underset{\sim}{\ell}'\underset{\sim}{T}) = \underset{\sim}{\ell}'\underset{\sim}{\mu}_T \tag{5.26}$$

$$\mathrm{Var}(\underset{\sim}{\ell}'\underset{\sim}{T}) = \underset{\sim}{\ell}'\underset{\sim}{\Sigma}_T\,\underset{\sim}{\ell} \tag{5.27}$$

$$\mathrm{Cov}(\underset{\sim}{\ell}'\underset{\sim}{T},\ \underset{\sim}{s}'\underset{\sim}{T}) = \underset{\sim}{\ell}'\underset{\sim}{\Sigma}_T\,\underset{\sim}{s} \quad . \tag{5.28}$$

Holland and Leinhardt (1974) suggest a "partial conditioning" scheme using a set of linear combinations as an approximation to one of the more complex conditional uniform distributions mentioned in Section 3. They reason that $\underset{\sim}{T}$ has an approximate multivariate Gaussian distribution, because $\underset{\sim}{T}$ is a sum of "loosely correlated indicator variables". It is a well known result that if $\underset{\sim}{T} \sim N(\underset{\sim}{\mu}\ \underset{\sim}{\Sigma})$ and $\underset{\sim}{L}\underset{\sim}{T}$ is a vector of linear combinations of the elements of $\underset{\sim}{T}$, then

$$E(\underset{\sim}{T}|\underset{\sim}{L}\underset{\sim}{T} = \underset{\sim}{L}\underset{\sim}{t}) = \underset{\sim}{\mu} + \underset{\sim}{\Sigma}\underset{\sim}{L}'(\underset{\sim}{L}\underset{\sim}{\Sigma}\underset{\sim}{L}')^{-1}\ \underset{\sim}{L}\ (\underset{\sim}{t} - \underset{\sim}{\mu}) \tag{5.29}$$

and

$$\mathrm{Cov}(\underset{\sim}{T}|\underset{\sim}{L}\underset{\sim}{T} = \underset{\sim}{L}\underset{\sim}{t}) = \underset{\sim}{\Sigma} - \underset{\sim}{\Sigma}\underset{\sim}{L}'(\underset{\sim}{L}\underset{\sim}{\Sigma}\underset{\sim}{L}')^{-1}\ \underset{\sim}{L}\underset{\sim}{\Sigma}\ . \tag{5.30}$$

Now that we know how to computer $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\mu}$ exactly under $U|\{X_{i+}\}$, (5.29) and (5.30) will give an approximation to $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\mu}$ computed under $U|\{X_{i+}\}$, $\{X_{+j}\}$ and $U|M$, $\{X_{i+}\}$, $\{X_{+j}\}$. For instance, if we condition on the vector $\underset{\sim}{m}$ shown in column 2 of Table (4.1), we obtain an approximation to $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\mu}$ computed under $U|M$, $\{X_{i+}\}$. Also, if we apply the above formulas to $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\mu}$ computed under $U|\{X_{i+}\}$ and let $\underset{\sim}{L}$ be the vector $\underset{\sim}{B}_{in}$ defined in column 6 of Table (4.1), we obtain an approximation to $U|\{X_{i+}\}$, $\{X_{+j}\}$. Unfortunately, the approximation may be poor as we are actually conditioning on the linear combination used to compute S^2_{in}, and not on the set of indegrees themselves. However, I believe that calculations with the $U|\{X_{i+}\}$ distribution will be quite important in sociometric data analysis, because of the "handle" that it gives in approximating the more complex distributions via (5.29) and (5.30).

6. Example

The data analyzed in this section were taken from McKinney (1948). Twenty-nine individuals in a ninth-grade classroom were asked to "Express (your) attitude toward serving in a discussion group with the other members of the class." Each student rated his/her cohorts with an acceptance, indifference, or rejection. The (10 x 10) sociomatrix in Table 6.1 represents the "acceptances" made by the subgroup consisting of the first ten students. I have chosen to ignore rejections, or negative choices, and treat them as indifferences, or null choices.

In the table, a mutual relationship between person i and person j is characterized by a M in the (i,j) and (j,i) cells. $M_{i.} = M_{.i}$ denotes the number of mutual relationships involving person i. $A_{i.}$ and $A_{.i}$ are, respectively, the number of asymmetric choices made by person i and the number of choices received by person i. Figure 6.1 displays the data in the form of a sociogram.

The sociogram displays the complex structure of this group. Persons 1, 2, 7, and 10 form a pure M-clique. Person 6 receives 5 unreciprocated choices, while person 3 makes 5 unreciprocated choices. Person 1 appears to be the individual with the most acquaintances in the group. Examining the individual choices sheds little light on the group structure; however, choices 4→1, 10→9 and 10→5 seen out of place because of the configuration of the sociogram.

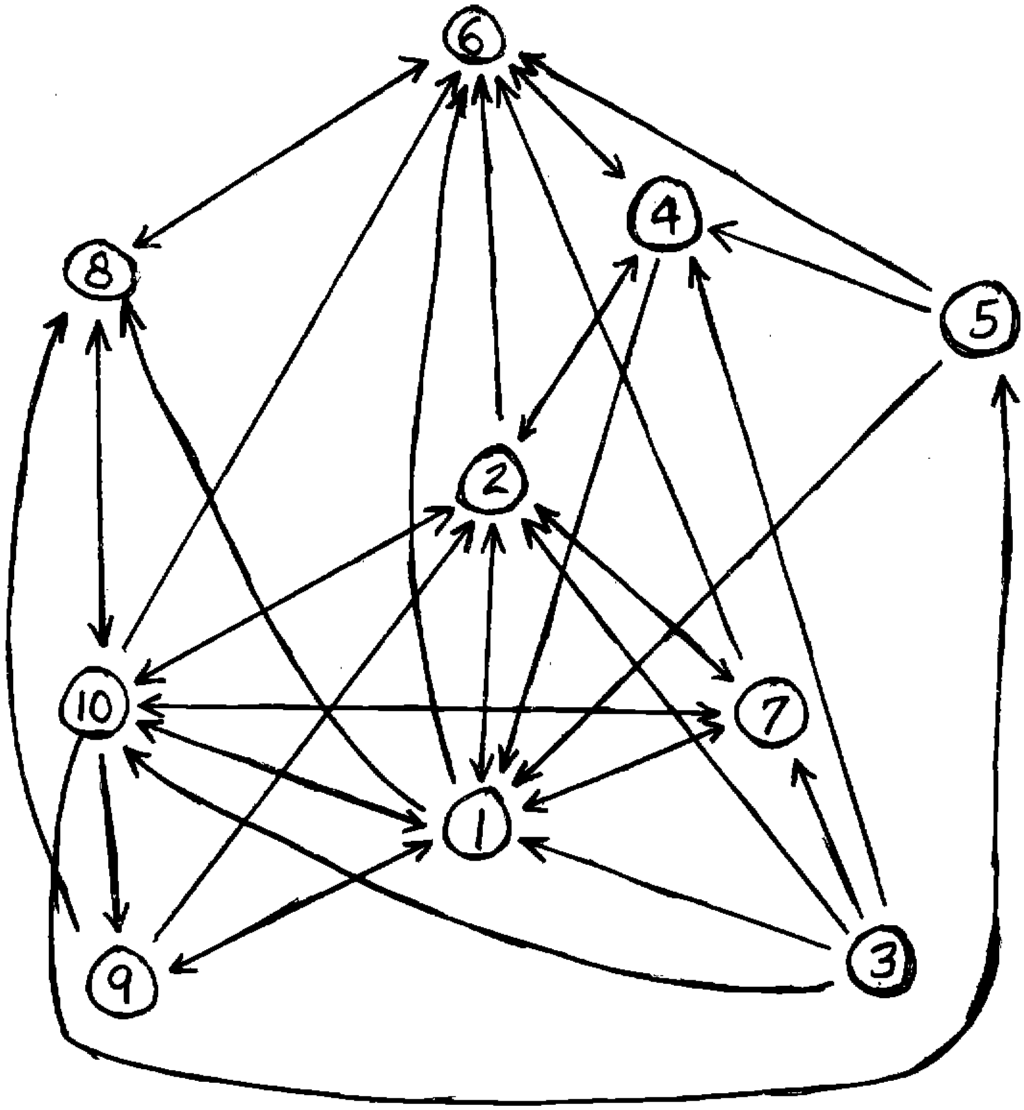
For this sociomatrix, I found the triad census and computed $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$ under U|MAN and U|$\{X_{i+}\}$. Using these two pairs of moments and the formulas (5.29) and (5.30), I obtained approximations to the conditional distributions U|M, $\{X_{i+}\}$ and U|M, $\{X_{i+}\}$, $\{X_{+j}\}$. By partially conditioning U|MAN on the vector $\underset{\sim}{B}_{out}$ (column 7 of Table 4.1) and partially conditioning U|$\{X_{i+}\}$ on the vector $\underset{\sim}{m}$ (column 2 of Table 4.1), I obtained the distributions which I will

Table 6.1: Sociomatrix Derived from Choices Made in a Ninth-Grade Classroom (see McKinney (1948)).

STUDENT

		1	2	3	4	5	6	7	8	9	10	X_{i+}	$M_{i\cdot}$	$A_{i\cdot}$
	1	-	M	0	0	0	1	M	1	M	M	6	4	2
	2	M	-	0	M	-	1	M	-	-	M	5	4	1
	3	1	1	-	1	0	0	1	0	0	1	5	0	5
S	4	1	M	0	-	0	M	0	0	0	0	3	2	1
T	5	1	0	0	1	-	1	0	0	0	0	3	0	3
U	6	0	0	0	M	0	-	0	M	0	0	2	2	0
D	7	M	M	0	0	0	1	-	0	0	M	4	3	1
E	8	0	0	0	0	0	M	0	-	0	M	2	2	0
N	9	M	1	0	0	0	0	0	1	-	0	3	1	2
T	10	M	M	0	0	1	1	M	M	1	-	7	4	3
	X_{+j}	7	6	0	4	1	7	4	4	2	5	40		
	$M_{\cdot j}$	4	4	0	2	0	2	3	2	1	4			
	$A_{\cdot j}$	3	2	0	2	1	5	1	2	1	1			

Figure 6.1: Sociogram derived from choices made in a ninth-grade classroom (see McKinney (1948)).



denote $U|MAN$, $\underset{\sim}{B}_{out}$ and $U|\{X_{i+}\}$, $\underset{\sim}{m}$ respectively. These two distributions are approximations to $U|M, \{X_{i+}\}$. I next obtained the approximate distributions $U|MAN$, $\underset{\sim}{B}_{out}$, $\underset{\sim}{B}_{in}$ by partially conditioning $U|MAN$ on the vectors $\underset{\sim}{B}_{out}$ and $\underset{\sim}{B}_{in}$ (columns 6 and 7 of Table 4.1), and $U|\{X_{i+}\}$, $\underset{\sim}{B}_{in}$, $\underset{\sim}{m}$ by partially conditioning $U|\{X_{i+}\}$ on the vectors $\underset{\sim}{B}_{in}$ and $\underset{\sim}{m}$. $U|MAN$, $\underset{\sim}{B}_{out}$, $\underset{\sim}{B}_{in}$ and $U|\{X_{i+}\}$, $\underset{\sim}{B}_{in}$, $\underset{\sim}{m}$ are approximations to the distributions $U|M, \{X_{i+}\}, \{X_{+j}\}$. Table 6.2 presents the triad census and the expected value of the triad census under these 6 distributions.

This triad census has large numbers of 012, 021C, and 111U triads. This indicates that the group has a considerable number of asymmetric relationships. The lack of 030C triads is also of interest because the number of 021C and 111U triads, each with two-thirds of a complete "cycle", would suggest the opposite. The abundance of asymmetric choices and 012 and 021C triads can be easily seen by examining the sociogram in Figure 6.1. The small group size aids in drawing conclusions from this figure. (As g increases, so does the complexity of the group's sociogram, and the triad census becomes more important in understanding group structure.)

An examination of the expected values reveals that the partial conditioning slightly reduces the differences between the census and its expected values. The distributions based on $U|MAN$ have expected values which fit the data more closely, in that the absolute differences between the expected and observed quantities are smaller than those found using distributions based on $U|\{X_{i+}\}$. It also appears that the differences between $U|MAN$ and $U|\{X_{i+}\}$ decrease when examining μ_T computed under $U|MAN$, $\underset{\sim}{B}_{out}$, $\underset{\sim}{B}_{in}$ and $U|\{X_{i+}\}$, $\underset{\sim}{B}_{in}$, $\underset{\sim}{m}$ (note the changes in μ_{102} and μ_{030T} over distributions).

Table 6.2: Triad Census, and Expected Values for Data in Table 6.1

		Expected Values					
TRIAD	TRIAD CENSUS	$U\|MAN$	$U\|\{X_{i+}\}$	$U\|MAN, B_{out}$	$U\|\{X_{i+}\}, \underset{\sim}{m}$	$U\|MAN, B_{in}, B_{out}$	$U\|\{X_{i+}\}, B_{in}, \underset{\sim}{m}$
003	8	4.74	3.26	6.39	4.99	7.00	5.23
012	19	18.27	16.55	19.90	18.81	20.50	19.12
102	7	11.16	6.63	9.41	10.44	8.77	10.23
021D	5	5.17	7.70	8.22	5.93	7.95	5.71
021U	5	5.17	6.63	3.97	5.01	4.91	5.83
021C	11	10.35	13.26	7.95	10.01	7.07	9.20
111D	7	13.39	10.09	6.16	12.16	7.07	12.70
111U	18	13.39	11.72	17.14	14.29	14.94	13.07
030T	4	5.18	11.72	6.23	4.68	6.62	5.29
030C	0	1.73	3.37	0.66	1.30	0.28	0.75
201	6	7.44	4.25	6.77	7.14	6.52	7.08
120D	7	3.56	4.25	2.97	3.65	3.70	4.26
120U	9	3.56	4.73	5.89	4.12	5.79	4.00
120C	4	7.12	8.50	5.94	7.30	5.51	6.94
210	6	8.37	6.54	10.24	8.77	10.92	9.08
300	4	1.40	0.77	2.16	1.41	2.44	1.49

I tested the structural hypotheses of transitivity and intransitivity by computing $\tau(\ell)$ for both of these hypotheses using the six random digraph distributions and the ℓ vectors given in Holland and Leinhardt (1976). Table 6.3 presents these results.

The τ values decrease as I partially condition $U|MAN$, and increase as I partially condition $U|\{X_{i+}\}$. This may be due to the previously mentioned fact that $\underset{\sim}{\mu}_T$ computed under the approximate distributions based on $U|MAN$ provide a better "fit" to the observed triad census.

The example demonstrates the phenomenon that different conditional distributions may produce differing $\underset{\sim}{\mu}_T$ and $\underset{\sim}{\Sigma}_T$. An investigator using the triad census to test structural hypotheses should compute the relevant $\tau(\underset{\sim}{\ell})$ under a variety of distributions, and then seek an explanation for the apparent differences or similarities of the τ values.

Table 6.3 Values of τ for testing the transitivity and intransitivity hypotheses for the data in Table 6.1.

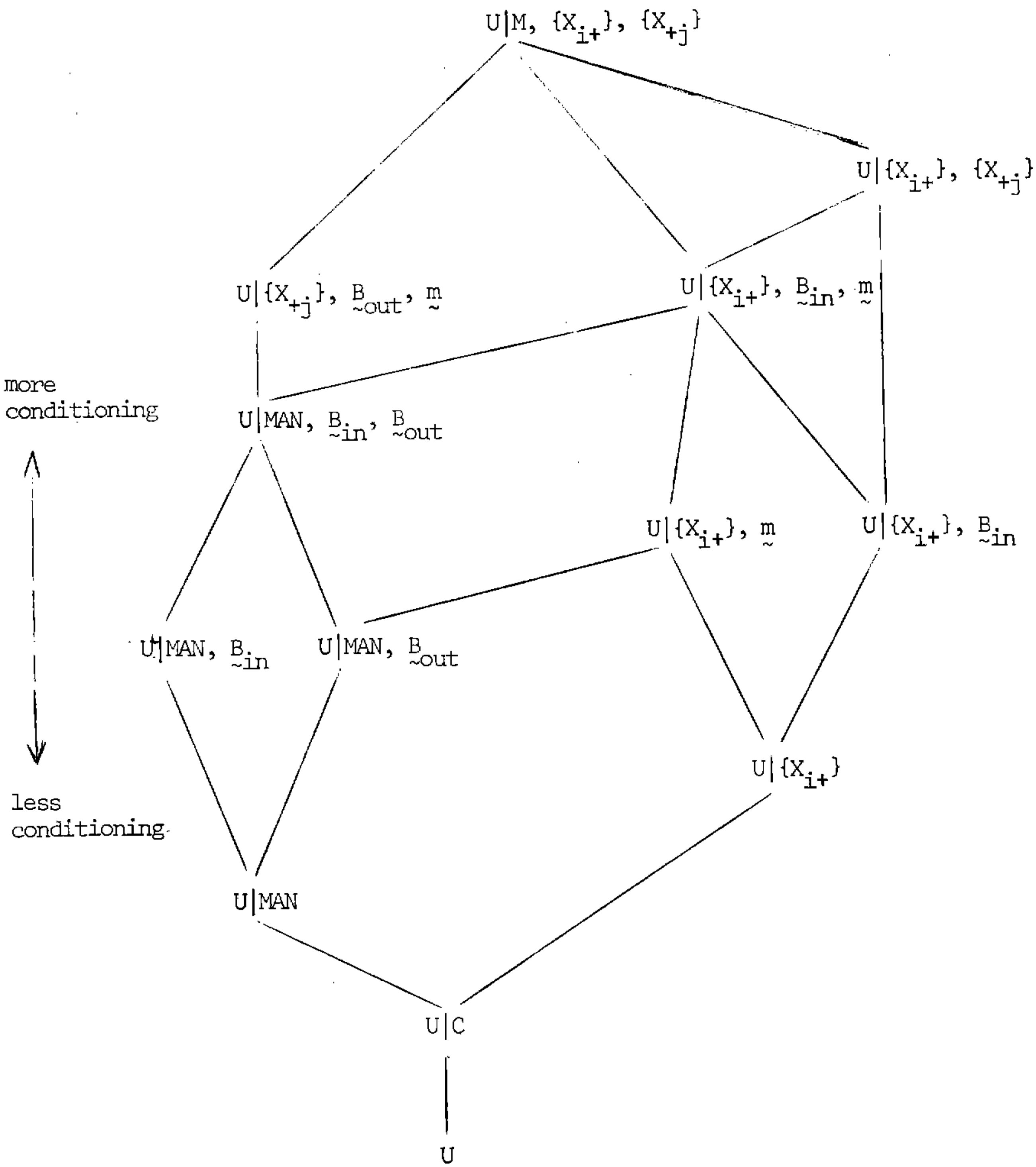
	Distribution					
	$U \mid MAN$	$U \mid \{X_{i+}\}$	$U \mid MAN, \underset{\sim}{B}_{out}$	$U \mid \{X_{i+}\}, \underset{\sim}{m}$	$U \mid MAN, \underset{\sim}{B}_{out}, \underset{\sim}{B}_{in}$	$U \mid \{X_{i+}\}, \underset{\sim}{m}, \underset{\sim}{B}_{in}$
Transitivity	3.00	3.68	1.34	3.70	0.744	4.07
Intransitivity	-2.61	-2.27	-1.13	-2.39	-0.572	-2.50

7. Summary

Figure 7.1 summarizes the relationships of the random digraph distributions and the partially conditioned digraph distributions studied in this paper. With the addition of the $U|\{X_{i+}\}$ distribution, the network becomes quite intricate. Note that the digraph distribution network is ordered so that as one moves from the bottom to the top of the figure, the amount of conditioning in each distribution increases. I have purposely left out the distributions based on $U|\{X_{+j}\}$ in order that the network remain comprehensible.

This paper has introduced two random directed graph distributions and has given the methods needed to compute the first two moments of these distributions. Individuals interested in the analysis of social networks now have powerful mathematical tools at their disposal to aid in their analyses. The example discussed in Section 6 demonstrates some of these. These methods show how statistical analyses can be applied to a specific field of study in the social sciences.

Figure 7.1: Network of Random Digraph Distributions



References

1. Davis, J.A. [1970] "Clustering and Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociograms", American Sociological Review, 35, 843-852.

2. Davis, J.A., and Leinhardt, S. [1972] "The Structure of Positive Interpersonal Relations in Small Groups", Sociological Theories in Progress, Volume 2, edited by J. Berger. Boston: Houghton Mifflin.

3. Ford, L.R., and Fulkerson, D.R. [1962] Flows in Networks. Princeton: Princeton University Press.

4. Harary, F., Norman, R.Z., and Cartwright, D. [1965] Structural Models: An Introduction to the Theory of Directed Graphs. New York: Wiley.

5. Holland, P.W. [1972] "Distributions on Digraphs", unpublished manuscript.

6. Holland, P.W., and Leinhardt, S. [1970] "A Method for Detecting Structure in sociometric data", American Journal of Sociology, 70, 492-513.

7. Holland, P.W., and Leinhardt, S. [1971] "Transitivity in Structural Models of Small Groups", Comparative Group Studies, 2, 107-124.

8. Holland, P.W., and Leinhardt, S. [1976] "Local Structure in Social Networks", Sociological Methodology 1976, edited by D. Heise. San Francisco: Jossey-Bass.

9. Mazur, A. [1971] "Comments on Davis Graph Model", American Sociological Review, 36, 308-311.

10. McKinney, J.C. [1948] "An Educational Application of a Two-Dimensional Sociometric Test", Sociometry, 11, 358-367.