Markov Graphs

Ove Frank; David Strauss

Journal of the American Statistical Association, Vol. 81, No. 395 (Sep., 1986), 832-842.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198609%2981%3A395%3C832%3AMG%3E2.0.CO%3B2-C

Markov Graphs

OVE FRANK and DAVID STRAUSS*

Log-linear statistical models are used to characterize random graphs with general dependence structure and with Markov dependence. Sufficient statistics for Markov graphs are shown to be given by counts of various triangles and stars. In particular, we show under which assumptions the triad counts are sufficient statistics. We discuss inference methodology for some simple Markov graphs.

KEY WORDS: Log-linear network model; Markov field; Graph inference.

1. INTRODUCTION

Random graphs have been used to describe social networks and other empirical data structures involving pairwise relationships. Some examples are friendships among children, transactions between bank accounts, and diffusion between neighboring sites in the plane or in space. See Freeman, Romney, and White (in press) for some recent studies of various applications.

Much of the literature on random graphs is devoted to simple probabilistic models having only a few parameters. A Bernoulli graph, for instance, has n vertices and independent edges with a common probability p for all unordered pairs of vertices. Even for random graph models with a simple probabilistic structure it is often difficult to determine sampling distributions of common graphical invariants and other statistics. Karoński (1982) reviewed many probabilistic results for Bernoulli graphs and related models.

Various attempts to model systematic deviations from some pure randomness model for graphs or digraphs were discussed and analyzed by Rapoport (1949, 1950, 1957). See also Rapoport (1979) and Strauss and Freeman (in press) for recent reviews. Rapoport's models have parameters that can be interpreted as measures of popularity differentiation and reciprocation in choice structures among people.

Log-linear graph models with parameters representing attraction, reciprocity, and other sociometric properties of networks were investigated by Holland and Leinhardt (1981), Fienberg, Meyer, and Wasserman (1985), and others. The basic digraph model of Holland and Leinhardt assumes that the unordered pairs of vertices have, independently of each other, no arc, a single arc, or two mutual arcs. This stochastic independence between dyads is essential for the analysis and is assumed in all other log-linear graph models that have been investigated. Wasserman (1980) and others who considered dynamic changes in graphs used Markov processes for the transitions, but the Markov dependence is over time and applies to independent dyads.

In many applications it is natural to assume that the graph reflects some probabilistic interdependencies or interactions that cause the dyads to be dependent. For instance, in communication or flow networks there are interactions due to common sources of information or flow and due to available paths and capacities. Holland and Leinhardt (1981) suggested that the triad counts (the numbers of different induced subgraphs of order 3) might be appropriate statistics for log-linear graph models with dependence structures. They did not attempt to verify that the triad counts can be obtained as sufficient statistics for graph models with an explicit dependence structure.

Our purpose in this article is to show how assumptions about the dependence structure lead to various families of log-linear models for graphs. We will consider network data comprising information about binary relations between individuals and, more generally, information about one or more characteristics of the pairs of individuals. To describe dependence structures we will use the Hammersley–Clifford theorem (Besag 1974) in much the same way as it has been used to derive spatial interaction models (see also Ripley 1981; Speed 1978; Strauss 1977). The concepts and main results we need are reviewed in the next section. Section 3 shows how random graphs with a general dependence structure can be described by log-linear probability models. We generalize the dyad independence models by introducing the notion of Markov dependency between dyads. A graph is said to be a Markov graph if only incident dyads can be conditionally dependent. We find sufficient statistics of general Markov graphs and of homogeneous Markov graphs. In Section 4 we discuss statistical inference for Markov graphs and investigate various inference methods. In Sections 5 and 6 we develop corresponding ideas for directed graphs, multiple graphs, and valued graphs.

2. PRELIMINARIES

If N is a finite set we write $|N|$ for its number of elements. If $|N| = n$ and k is an integer in the interval $[0, n]$, N^k is the set of all ordered sequences of k elements from N, $N^{(k)}$ is the set of all ordered sequences of k distinct elements from N, and $\binom{N}{k}$ is the set of all k-subsets of N. The numbers of elements in N^k, $N^{(k)}$, and $\binom{N}{k}$ are n^k, $n^{(k)} = n!/(n-k)!$, and $\binom{n}{k} = n^{(k)}/k!$, respectively.

A graph G on N is a subset of the set of pairs of elements in N; $G \subseteq \binom{N}{2}$ for unordered pairs, $G \subseteq N^{(2)}$ for ordered pairs of distinct elements, and $G \subseteq N^2$ for ordered pairs of not necessarily distinct elements. The elements of N can be displayed as points connected by undirected or directed lines corresponding to the elements of G. The elements in N, $\binom{N}{2}$, $N^{(2)}$, and $N^2 - N^{(2)}$ are called vertices, edges, arcs, and loops, respectively.

* Ove Frank is Professor, Statistics Department, University of Stockholm, Box 6701, S-113 85 Stockholm, Sweden. David Strauss is Associate Professor, Statistics Department, University of California, Riverside, CA 92521. The work reported here was initiated while Frank visited the Statistics Department at the University of California, Riverside, during the academic year 1982–1983. Partial support was given by the Swedish Council for Research in the Humanities and Social Sciences.

Consider a sequence $Z = (Z_1, \ldots, Z_m)$ of m discrete random variables. Let $M = \{1, \ldots, m\}$ be the index set, let R_i be the range space of Z_i, and let R be the range space of Z. Assume that $|R| < \infty$. Denote the probability function of Z by $\Pr(Z = z) = P(z)$ for $z \in R$. The function P can be defined by

$$P(z) = c^{-1}\exp Q(z) \tag{2.1}$$

with

$$c = \sum_{z \in R} \exp Q(z) \tag{2.2}$$

for any real function Q on R. Conversely, any P with $P(z) > 0$ for $z \in R$ can be represented in exponential form with

$$Q(z) = \log P(z) - \log P(z^*) \tag{2.3}$$

for a fixed $z^* \in R$. By application of the inclusion–exclusion principle it can be shown that $Q(z)$ is representable as a sum,

$$Q(z) = \sum_{A \subseteq M} \lambda_A(z_A), \tag{2.4}$$

where z_A is the subsequence of z consisting of z_i with $i \in A$. For identifiability of each λ_A we must impose certain side conditions. We will consider two different choices of constraints.

The first choice is obtained by selecting a fixed element z^* in R and specifying the values of $\lambda_A(z_A)$ for all z_A with at least one component z_i equal to z_i^*. This choice is convenient and will be adopted if we have binary random variables. Following Besag (1974) we can without loss of generality assume that a zero sequence is included in R, take $z^* = 0$, and set $\lambda_A(z_A) = 0$ if $z_i = 0$ for some $i \in A$. In the binary case it follows that

$$\lambda_A(z_A) = \alpha_A \prod_{i \in A} z_i, \tag{2.5}$$

where α_A is the value of $\lambda_A(z_A)$ when $z_i = 1$ for all $i \in A$.

The other choice of constraints on λ_A is given by

$$\sum_{z_i \in R_i} \lambda_A(z_A) = 0 \tag{2.6}$$

for $i \in A$ and $z \in R$. This can be shown to imply that

$$\lambda_A(z_A) = \sum_{B \subseteq A} (-1)^{|B|} \overline{Q}_B(z_B), \tag{2.7}$$

where $\overline{Q}_B(z_B)$ is the average of $Q(z)$ for $z_i \in R_i$ $(i \notin B)$.

The function λ_A is called the A interaction in Z. These interactions are useful to describe dependence structures among the random variables in the sequence Z. In fact, the Hammersley–Clifford theorem (Besag 1974) provides the important link between dependence structures and interactions. To state this theorem in a form suitable for us we introduce a *dependence graph* D on M that defines the conditional dependencies present among the pairs of random variables in Z. D consists of all pairs $\{i, j\} \in \binom{M}{2}$ such that Z_i and Z_j are dependent conditional on the rest of Z. For example, a sequence $(Z_1, \ldots, Z_m)$ of independent random variables has a dependence graph $D = \emptyset$ that is empty and a Markov sequence $(Z_1, \ldots, Z_m)$ has a dependence graph $D = \{\{1, 2\}, \{2, 3\}, \ldots, \{m - 1, m\}\}$ consisting of all pairs of nearest neighbors in index space M. It should be noted that the dependence graph D is a subset of $\binom{M}{2}$ and its vertex set M is the index set of the random variables under investigation. A *clique* of D is a nonempty subset A of M such that either $|A| = 1$ or $\binom{A}{2} \subseteq D$. According to the Hammersley–Clifford theorem $\lambda_A(z_A) = 0$ for all $z \in R$ unless A is a clique of D. Thus in the expansion of $Q(z)$, we need only those A interactions that correspond to cliques of D.

It is readily seen that for a sequence of independent random variables $(Z_1, \ldots, Z_m)$ there are only singleton cliques of D and consequently,

$$Q(z) = \sum_{i=1}^{m} \lambda_i(z_i)$$

and

$$P(z) = c^{-1} \prod_{i=1}^{m} \exp \lambda_i(z_i),$$

so $\exp \lambda_i(z_i)$ is proportional to $\Pr(Z_i = z_i)$.

For a Markov sequence $(Z_1, \ldots, Z_m)$ there are singleton cliques and nearest neighbor pair cliques of D, which implies that

$$Q(z) = \sum_{i=1}^{m} \lambda_i(z_i) + \sum_{i=1}^{m-1} \lambda_{i,i+1}(z_i, z_{i+1})$$

and

$$P(z) = c^{-1}\left[\prod_{i=1}^{m} \exp \lambda_i(z_i)\right]\left[\prod_{i=1}^{m-1} \exp \lambda_{i,i+1}(z_i, z_{i+1})\right],$$

so $\exp \lambda_i(z_i)$ is proportional to $\Pr(Z_i = z_i)$ and $\exp \lambda_{i,i+1}(z_i, z_{i+1})$ is proportional to

$$\Pr(Z_i = z_i, Z_{i+1} = z_{i+1})/\Pr(Z_i = z_i)\Pr(Z_{i+1} = z_{i+1}).$$

3. DEPENDENCE STRUCTURES FOR RANDOM GRAPHS

Dependence graphs D on a vertex set M can be used to describe dependence structures of arbitrary sets of random variables indexed by the elements of M. In particular, if we choose the random variables as the edge indicators of a random graph G on a vertex set N, then the index set of these random variables is the set $M = \binom{N}{2}$ of possible edges of G. In fact, if G is a random graph on $N = \{1, \ldots, n\}$, the edge indicators are given by the elements of the adjacency matrix of G. The adjacency matrix of G is a symmetric matrix $Y = (Y_{uv})$ of binary random variables defined by

$$Y_{uv} = 1 \text{ if } \{u, v\} \in G$$

$$= 0 \text{ otherwise}$$

for $(u, v) \in N^2$. Since $Y_{vv} = 0$ for $v \in N$ and $Y_{uv} = Y_{vu}$ for $(u, v) \in N^{(2)}$, there are $\binom{n}{2}$ random variables Y_{uv} with

$1 \le u \le v \le n$ that describe the random graph G. The dependence graph D of G is a nonrandom graph that specifies the dependence structure between the $\binom{n}{2}$ random variables Y_{uv} $(1 \le u < v \le n)$. The vertices of D are the possible edges of G, that is, the pairs $\{u, v\} \in M$, and the edges of D are the pairs of edges of G that are conditionally dependent. This means that D has an edge between $\{s, t\}$ and $\{u, v\}$ in M iff Y_{st} and Y_{uv} are dependent conditional on the rest of Y.

According to the Hammersley–Clifford theorem we obtain the following characterization of the probability function of a general random graph G.

Theorem 1. Any random undirected graph G on N with dependence structure D has probability

$$\Pr(G) = c^{-1}\exp \sum_{A \subseteq G} \alpha_A, \tag{3.1}$$

where c is a normalizing constant given by

$$c = \sum_{\substack{G \\ \Pr(G)>0}} \exp \sum_{A \subseteq G} \alpha_A \tag{3.2}$$

and α_A is an arbitrary constant if A is a clique of D and $\alpha_A = 0$ otherwise.

We note that the formula for $\Pr(G)$ is a log-linear model with parameters α_A and sufficient statistics

$$\prod_{\{u,v\} \in A} Y_{uv} \tag{3.3}$$

for all A that are cliques of D. Thus the sufficient statistics are indicators of various subgraphs of G. These subgraphs will be called *sufficient subgraphs.*

We also note that if A is a clique of D, then all nonempty subsets of A are also cliques of D. Thus a maximal clique A of size r contains $2^r - 1$ nonempty subsets $B \subseteq A$ that are also cliques of D. The $2^r - 1$ parameters α_B corresponding to these cliques $B \subseteq A$ can be considered as a particular parameterization of a general probability distribution for the r binary random variables Y_{uv} for $\{u, v\} \in A$. If the maximal cliques of D are disjoint, then it follows that a general probability distribution of G is specified by separate probability distributions of the maximal sufficient subgraphs of G.

To illustrate these concepts we now consider some examples.

Example 1. The simplest dependence structure possible for G is obtained if all edge indicators Y_{uv} are mutually independent Bernoulli (p_{uv}) for $1 \le u < v \le n$. We call this G a *Bernoulli graph* with edge probabilities p_{uv} for $1 \le u < v \le n$. Then the dependence graph D is empty and the only cliques are the singleton subsets of M, that is, the possible edges of G. Theorem 1 yields

$$\Pr(G) = c^{-1}\exp \sum_{\{u,v\} \in G} \alpha_{uv} = c^{-1} \prod_{\{u,v\} \in G} \exp \alpha_{uv},$$

where

$$c = \sum_{G \subseteq M} \prod_{\{u,v\} \in G} \exp \alpha_{uv} = \prod_{u<v} (1 + \exp \alpha_{uv}).$$

Thus

$$\Pr(G) = \left[\prod_{\{u,v\} \in G} ((\exp \alpha_{uv})/(1 + \exp \alpha_{uv})) \right] \times \left[\prod_{\{u,v\} \notin G} (1/(1 + \exp \alpha_{uv})) \right]$$

and it follows that

$$p_{uv} = (\exp \alpha_{uv})/(1 + \exp \alpha_{uv}),$$

so α_{uv} is the log odds of edge $\{u, v\}$:

$$\alpha_{uv} = \log(p_{uv}/(1 - p_{uv})).$$

In this case there are $\binom{n}{2}$ sufficient subgraphs, namely, all of the possible edges in G.

Example 2. Let G be a random graph on a vertex set N of six points in the plane, as shown in Figure 1. There are 15 possible edges in G, and these are the vertices of the dependence graph D. Assume that the dependence structure of G specifies that two edges of G are conditionally independent if they are not parallel, that is, if they correspond to two pairs of points in the plane that cannot be joined by parallel lines. The dependence graph D of G is shown in Figure 2. In this case there are six disjoint maximal cliques of D. Three of these have size 4 and three are singletons. Any nonempty subset of a maximal clique is a clique, so in total there are 48 cliques, namely, 15 of size 1, 18 of size 2, 12 of size 3, and 3 of size 4. Theorem 1 yields a representation of $\Pr(G)$ with 48 parameters α_A. In this case the maximal cliques of D correspond to the six directions parallel and perpendicular to the sides of the outer triangle in Figure 1. With independent directions the general probability distribution of G is obtained from six independent marginal distributions. Three of these have four edge indicators for the edges parallel to one side of the triangle, and three have only one edge indicator for the edge perpendicular to one of the sides of the triangle. Three distributions with 2^4 outcomes and three distributions with 2 outcomes can be specified with $3(2^4 - 1) + 3(2 - 1) = 48$ parameters in general, which is just what we obtained by the theorem.

We now introduce a Markov property that will specify the cliques of D and thus specify the terms in the exponent

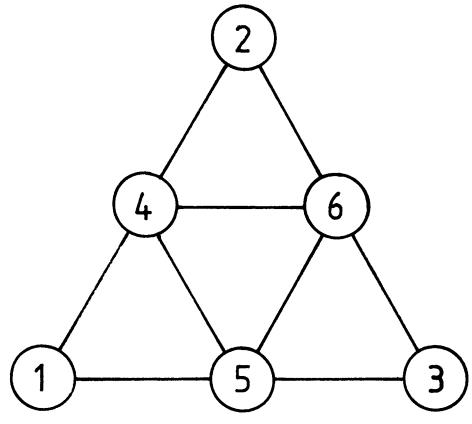

Figure 1. The Vertex Set of G in Example 2, Consisting of Six Points in a Regular Plane Triangular Lattice.

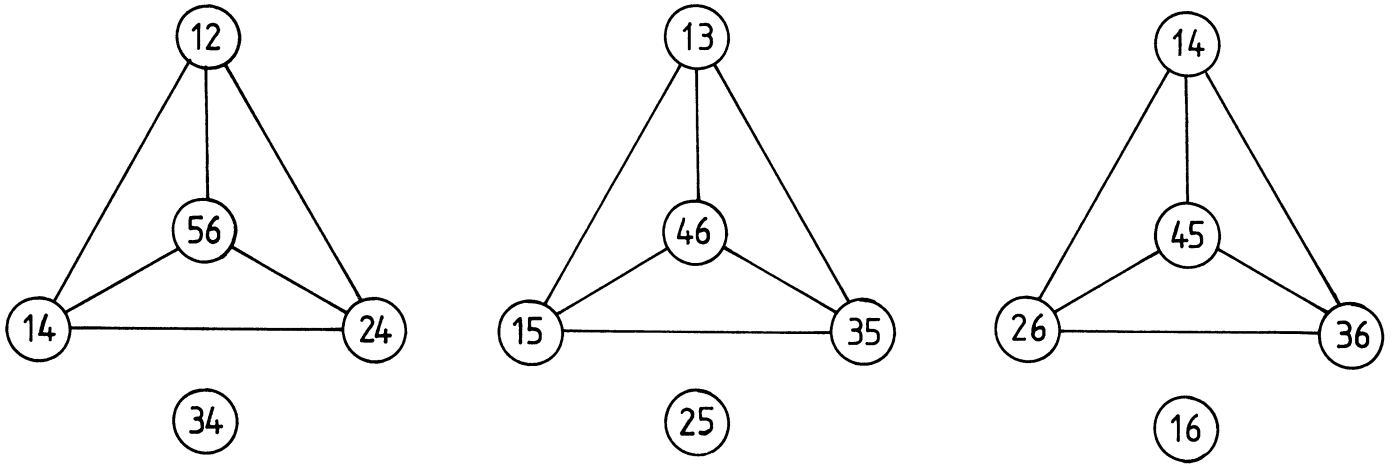

Figure 2. Dependence Graph D of G in Example 2.

of the probability function. We will say that a graph G is a *Markov graph,* or has *Markov dependence,* if D contains no edge between disjoint sets $\{s, t\}$ and $\{u, v\}$ in M. This means that nonincident edges in G are conditionally independent. This definition of Markov dependence is consistent with the nearest neighbor definition (Besag 1974) if the nearest neighbors of $\{u, v\} \in M$ are $\{u, w\}$ and $\{v, w\}$ for all $w \in N$ different from u and v. Figure 3 illustrates the dependence graph D for a Markov graph of order $n = 4$.

For general Markov graphs the cliques of D correspond to sets of edges such that any pair of edges within the set must be incident. Such sets are just the triangles and stars; that is, triangles

$$T_{uvw} = \{\{u, v\}, \{v, w\}, \{w, u\}\}, \quad (u, v, w) \in N^{(3)}, \tag{3.4}$$

and k-stars

$$S_{v_0 \cdots v_k} = \{\{v_0, v_i\}: i = 1, \ldots, k\}, \quad (v_0, \ldots, v_k) \in N^{(k+1)} \tag{3.5}$$

for $k = 1, \ldots, n - 1$. These sufficient subgraphs are displayed in Figure 4 for $n = 4$. We note that T_{uvw} is invariant under permutations of (u, v, w). Furthermore, $S_{v_0 v_1} = S_{v_1 v_0}$ and for $k > 1$ $S_{v_0 \cdots v_k}$ is invariant under permutations of $(v_1, \ldots, v_k)$. Thus there are $\binom{n}{3}$ distinct triangles, $\binom{n}{2}$ distinct 1-stars, and $n\binom{n-1}{k}$ distinct k-stars for $k = 2, \ldots, n - 1$. In total there are

$$\binom{n}{3} + \binom{n}{2} + \sum_{k=2}^{n-1} n\binom{n-1}{k} = n\,2^{n-1} + \binom{n}{3} - \binom{n+1}{2} \tag{3.6}$$

sufficient subgraphs. Writing $\alpha_A = \alpha(A)$, we now introduce the notation

$$\tau_{uvw} = \alpha(T_{uvw}), \qquad \sigma_{v_0 \cdots v_k} = \alpha(S_{v_0 \cdots v_k}) \tag{3.7}$$

and obtain the following result from Theorem 1.

Theorem 2. Any undirected Markov graph has probability

$$\Pr(G) = c^{-1}\exp\left[\sum \tau_{uvw} + \sum_{k=1}^{n-1} \sum \sigma_{v_0 \cdots v_k}/k!\right], \tag{3.8}$$

where the first sum is over $\{u, v, w\} \in \binom{N}{3}$ with $T_{uvw} \subseteq G$ and the last sum is over $(v_0, \ldots, v_k) \in N^{(k+1)}$ with $S_{v_0 \cdots v_k} \subseteq G$. This log-linear model has $n\,2^{n-1} + \binom{n}{3} - \binom{n+1}{2}$ parameters and sufficient statistics corresponding to triangles and stars in G.

It is interesting to note that no tetrads or other induced

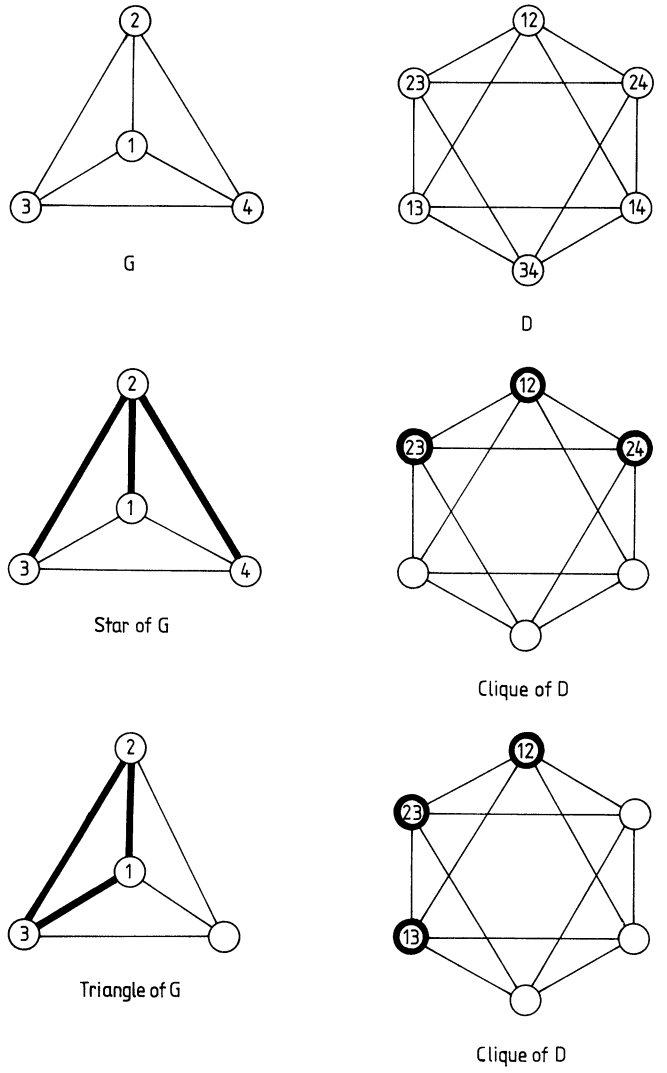

Figure 3. A Markov Graph G of Order n = 4 and Its Dependence Graph D (two illustrations of the correspondence between cliques of D and sufficient subgraphs of G).

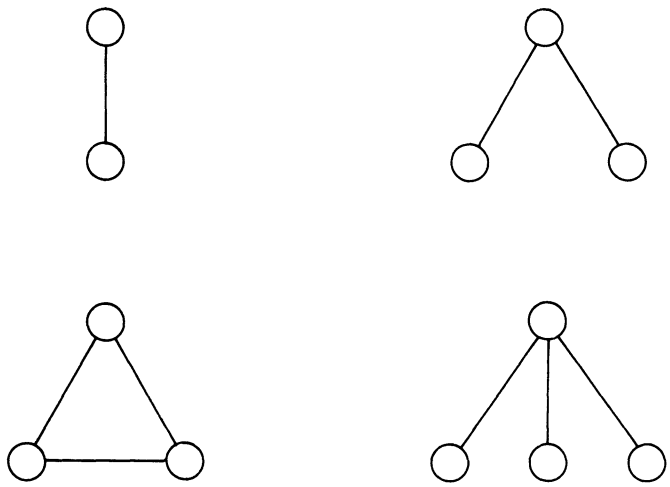

Figure 4. Sufficient Subgraphs of Markov Graphs of Order n = 4.

subgraphs of higher order are needed. All triads are needed, though, since we require 1-stars, 2-stars, and triangles.

To simplify the model further and reduce the number of parameters we introduce a *homogeneity* condition: that all isomorphic graphs G have the same probability $\Pr(G)$. In effect, we are assuming that vertices are a priori indistinguishable, and we do not introduce parameters specific to the different vertices. By writing

$$\tau = \tau_{uvw}, \qquad \sigma_k = \sigma_{v_0 \cdots v_k}, \tag{3.9}$$

the probability function reduces as follows.

Theorem 3. Any homogeneous undirected Markov graph has probability

$$\Pr(G) = c^{-1}\exp\left(\tau t + \sum_{k=1}^{n-1} \sigma_k s_k\right), \tag{3.10}$$

where t is the number of triangles in G and s_k is the number of k-stars in G.

Here we have sufficient statistics

$$t = \sum_{\{u,v,w\} \in \binom{N}{3}} Y_{uv} Y_{vw} Y_{wu} \tag{3.11}$$

and

$$s_k = \frac{1}{k!} \sum_{(v_0, \ldots, v_k) \in N^{(k+1)}} Y_{v_0 v_1} Y_{v_0 v_2} \cdots Y_{v_0 v_k}, \qquad k = 1, \ldots, n-1 \tag{3.12}$$

and n parameters $\tau, \sigma_1, \ldots, \sigma_{n-1}$.

For a simple model, capturing both transitivity and clustering, we may take the special case

$$\Pr_{\rho\sigma\tau}(G) = c^{-1}\exp\{\rho r + \sigma s + \tau t\}, \tag{3.13}$$

where $\rho = \sigma_1$, $r = s_1$, $\sigma = \sigma_2$, and $s = s_2$. We shall refer to this as the *triad model* or the $\rho\sigma\tau$ model. Its sufficient statistics r, s, and t are the numbers of edges, 2-stars, and triangles in G. An equivalent set of sufficient statistics is the set of triad counts of G, that is, the numbers of induced subgraphs of order 3 and size 0, 1, 2, and 3. These four triad counts, t_0, t_1, t_2, and t_3, say, sum to $\binom{n}{3}$, so there is need for only three of them as sufficient statistics. It can be easily shown that

$$t_0 = \binom{n}{3} - (n-2)r + s - t,$$

$$t_1 = (n-2)r - 2s + 3t,$$

$$t_2 = s - 3t, \qquad t_3 = t.$$

Triad count statistics for undirected graphs were used by Frank (1979) and Frank and Harary (1982). If σ and τ are 0, we have a Bernoulli graph with edge probability $p = (\exp \rho)/(1 + \exp \rho)$. In some cases ρ is a nuisance parameter to be eliminated by conditioning on r. Then we have the conditional probabilities

$$\Pr_{\sigma\tau}(G) = c^{-1}\exp(\sigma s + \tau t) \tag{3.14}$$

for graphs G with r edges.

In the general homogeneous case the various star parameters are hard to interpret jointly in view of the multiple counting. For example, every k-star contains $\binom{k}{j}$ j-stars for all $j < k$. It seems intuitively appealing to replace s_k by d_k, the number of vertices in G of degree k, for $k = 1, \ldots, n-1$. We have

$$s_k = \sum_{j \geq k} \binom{j}{k} d_j. \tag{3.15}$$

If we introduce new parameters

$$\delta_j = \sum_{k \leq j} \binom{j}{k} \sigma_k, \tag{3.16}$$

we have

$$\Pr(G) = c^{-1}\exp\left(\tau t + \sum_{j=1}^{n-1} \delta_j d_j\right). \tag{3.17}$$

Here $\tau > 0$ gives a bias toward transitivity and $\tau < 0$ against transitivity. The parameters δ_j control the bias for or against vertices of degree j.

4. INFERENCE FOR MARKOV GRAPHS

Standard likelihood techniques for the Markov models are not immediately applicable because of the complicated functional dependence of the normalizing constant c on the parameters. If the number of vertices is less than, say, six it is feasible to evaluate c by direct enumeration; otherwise fitting the model to data can be a difficult problem. In this section we indicate some possible approaches.

For illustration we consider first the clustering model

$$\Pr_\sigma(G) = \exp(\sigma s)/c(\sigma), \tag{4.1}$$

where s is the number of 2-stars in G. The normalizing constant $c(\sigma)$ is $\Sigma \exp(\sigma s)$, the sum being over all graphs with n vertices and r edges. Thus we have a triad model conditioned by r with clustering parameter σ and transitivity parameter $\tau = 0$. It is straightforward to show that the cumulant generating function of s is

$$K_s(\theta) = \log c(\theta + \sigma) - \log c(\sigma). \tag{4.2}$$

Setting $\sigma = 0$ gives us the Taylor expansion of $\log c$,

$$\log c(\theta) - \log c(0) = \sum_{i=1}^{\infty} \kappa_i \theta^i / i!, \tag{4.3}$$

where the κ_i are the null cumulants of s. By writing $m = \binom{n}{2}$, we have

$$\begin{aligned}\kappa_1 &= E_0(s) = n^{(3)} r^{(2)} / 2m^{(2)},\\ \kappa_2 &= \mathrm{var}_0(s) = 2m(n-2)(2n-5) r^{(3)}/m^{(3)}\\ &\quad + \left\{\frac{1}{2} n^{(3)}\left(\frac{1}{2} n^{(3)} - 1\right) - 2m(n-2)(2n-5)\right\}\\ &\quad \times \frac{r^{(4)}}{m^{(4)}} - \kappa_1^{(2)},\end{aligned} \tag{4.4}$$

and so on.

The maximum likelihood estimator $\hat{\sigma}$ is a root of

$$s = (d/d\sigma) \log c(\sigma). \tag{4.5}$$

The right side is $E_\sigma(s)$ under the model (4.1). Since the distributions of s are stochastically ordered by σ, $\hat{\sigma}$ is unique.

In general it is impossible to evaluate c or its derivatives exactly. In fact, (4.2) shows that $c(\sigma)$ involves all of the null cumulants of s. In statistical mechanics, where constants such as c are known as *partition functions,* there is an extensive literature on series expansions, and in the literature on inference for exponential families of distributions a variety of approximations to the normalizing function have been discussed. For some references, see, for instance, Ruelle (1969), Strauss (1975; in press), Pickard (1977), and Ogata and Tanemura (1984).

In principle the best way to estimate σ is from a plot of $E_\sigma(s)$ against σ, as in Figure 5. This figure, based on graphs with $n = 12$ vertices and $r = 33$ edges, shows a characteristic sigmoid shape. We obtained such plots by simulating many random realizations and averaging the s values

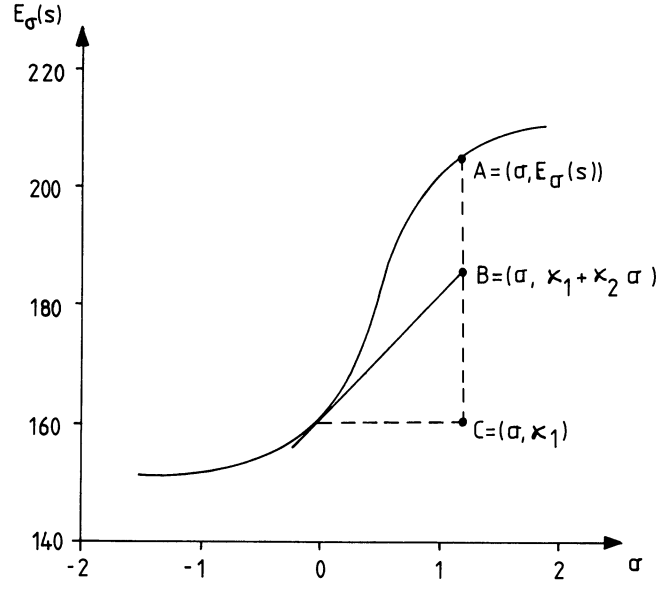

Figure 5. Expected Number of 2-Stars for the Clustering Model (4.1) With n = 12 Vertices and r = 33 Edges.

at each of a number of values of σ. The simulations were performed by a method of Strauss (in press), which is an adaptation of the Metropolis procedure (Hammersley and Handscomb 1964). In general we have to choose a fixed value of σ and find the average number $\bar{s}$ of 2-stars for a large number of simulated graphs with the same numbers of vertices and edges as the data. This should be repeated for different values σ to obtain a curve $\bar{s} = f(\sigma)$. From this curve the unknown parameter σ is estimated by the value $\hat{\sigma}$ that corresponds to the number s of 2-stars in the data, that is, $s = f(\hat{\sigma})$. To form the likelihood function evaluated at $\hat{\sigma}$ and to perform likelihood ratio tests we need an unbiased estimator of $1/c(\hat{\sigma})$. This can be achieved by simulating a large number of graphs for the value $\hat{\sigma}$ and taking the average over these of $\exp(-\hat{\sigma}s)$. This average can be shown to be an unbiased estimator of $c(0)/c(\hat{\sigma})$.

In practice the construction of the $(\sigma, E_\sigma(s))$ curve seems too large a task for routine data analysis and, therefore, we now turn to simpler methods of estimating σ.

For sufficiently small σ we can approximate $c(\sigma)$ by neglecting terms of order σ^3 or higher. Then from (4.1) the normal equation for this "truncated" estimator of σ is

$$\tilde{\sigma} = (s - \kappa_1)/\kappa_2, \tag{4.6}$$

where κ_1 and κ_2 are given by (4.4). The bias of the estimator may be examined with the help of a plot such as Figure 5. It may be seen, for instance, that the proportional expected bias $E\{(\tilde{\sigma} - \sigma)/\sigma\}$ is given by the ratio AB:BC in the figure. In this case the bias is serious for $|\sigma| > .2$.

Another method of estimation that can be applied for the Markov models is logit regression. We now illustrate this method by estimating the triad model (3.13) with parameters ρ, σ, and τ. Let Y be the adjacency matrix of G and set

$$P_{uv} = \Pr(Y_{uv} = 1 \mid \text{rest of } Y). \tag{4.7}$$

We have

$$\begin{aligned}P_{uv} &= \Pr(G = G_{uv}^+ \mid G = G_{uv}^+ \text{ or } G = G_{uv}^-)\\ &= \Pr(G = G_{uv}^+)/[\Pr(G = G_{uv}^+) + \Pr(G = G_{uv}^-)],\end{aligned} \tag{4.8}$$

where G_{uv}^- and G_{uv}^+ denote G with edge $\{u, v\}$ removed and added, respectively. It follows that

$$\begin{aligned}\text{logit } P_{uv} &= \log[P_{uv}/(1 - P_{uv})]\\ &= \log \Pr(G = G_{uv}^+) - \log \Pr(G = G_{uv}^-)\\ &= \rho r_{uv} + \sigma s_{uv} + \tau t_{uv},\end{aligned} \tag{4.9}$$

where r_{uv}, s_{uv}, and t_{uv} are the increments in r, s, and t when G is changed from G_{uv}^- to G_{uv}^+. Obviously $r_{uv} = 1$, and the parameters ρ, σ, and τ appear as regression coefficients in a logit regression of Y_{uv} on s_{uv} and t_{uv}, as follows:

$$\text{logit } E(Y_{uv} \mid \text{rest of } Y) = \rho + \sigma s_{uv} + \tau t_{uv}. \tag{4.10}$$

The estimates, which we denote by $\tilde{\rho}$, $\tilde{\sigma}$, and $\tilde{\tau}$, are thus easily obtained; indeed, routines are available in standard

packages. The BMDP package, for example, minimizes

$$\sum_{u<v} W_{uv}(Y_{uv} - P_{uv})^2, \qquad (4.11)$$

where the weight W_{uv} is $1/P_{uv}(1 - P_{uv})$, the reciprocal of the conditional variance of Y_{uv}.

Naturally (4.10) is not a standard logit model, since the values Y_{uv} are conditioned on other values in Y. The method is in fact analogous to the "pseudolikelihood" estimation procedure developed by Besag (1974) for binary lattice models. Nevertheless, our limited experiences suggest that the method has promise; in fact, it seems to give estimates rather similar to those obtained by maximum likelihood. It is not clear what choice of weights in (4.11) is optimal. We found that the BMDP weighting tends to underestimate $|\sigma|$ or $|\tau|$; this is hardly surprising, since when these are large the values P_{uv} are close to 0 or 1, and this gives large values of W_{uv}. We have preferred the unweighted version of (4.11), with $W_{uv} = 1$ for $u < v$, and have used it in various experiments. One is described forthwith. We also illustrate the method with an application to real data.

In one experiment, to compare the logit regression estimator $\tilde{\tau}$ with the maximum likelihood estimator $\hat{\tau}$ we generated random graphs of order $n = 30$, using the $\rho\sigma\tau$ model with $\rho = \sigma = 0$ and various values of τ. Using a procedure similar to the one that generated the curve in Figure 5, we approximated $E_\tau t$ as a function of τ. The result is shown in Figure 6. The maximum value of t corresponding to a complete graph is $\binom{30}{3} = 4{,}060$, and it can be seen that for $\tau > .2$ a nearly complete graph can be expected. For such τ values one cannot hope for any estimation method to be successful. Similar remarks apply to the case of low τ values, such as $\tau < -.15$. The maximum likelihood estimate $\hat{\tau}$ corresponding to an observed value of t can be approximated by simply reading off the abscissa of the point at height t on the curve in Figure 6.

At each of several τ values we generated four independent realizations of the model and found the logit regression estimates $\tilde{\tau}$ together with approximate maximum likelihood estimates $\hat{\tau}$. The results are shown in Table 1. As expected, neither method works well at $\tau = -.20$, and even at $\tau = -.10$ the curve of Figure 6 is too flat for reliable estimation. For the other τ values both of the estimators do quite well. At least in this experiment, the logit regression estimator seems to be a satisfactory substitute for the computationally burdensome maximum likelihood estimator.

Table 1. Numbers of Triangles t, Logit Regression Estimates $\tilde{\tau}$, and Maximum Likelihood Estimates $\hat{\tau}$ Obtained in Four Independent Realizations of a Random Graph of Order 30

τ	t	$\tilde{\tau}$	$\hat{\tau}$
−.20	173	−.17	−.18
	87	−.26	−.30
	138	−.22	−.20
	90	−.34	−.35
−.10	206	−.12	−.11
	121	−.16	−.22
	188	−.14	−.14
	119	−.15	−.22
0	531	.00	.00
	619	.01	.02
	645	.01	.02
	452	.00	−.01
.05	1291	.06	.05
	1324	.05	.05
	1525	.06	.06
	1137	.05	.05
.10	2621	.09	.09
	3266	.11	.11
	3371	.11	.11
	3169	.10	.11
.15	3789	.14	.16
	3707	.13	.13
	3914	.15	.17
	3680	.13	.13

NOTE: The graphs are generated according to a transitivity model for various values of the parameter τ.

We conclude this section with an application of logit

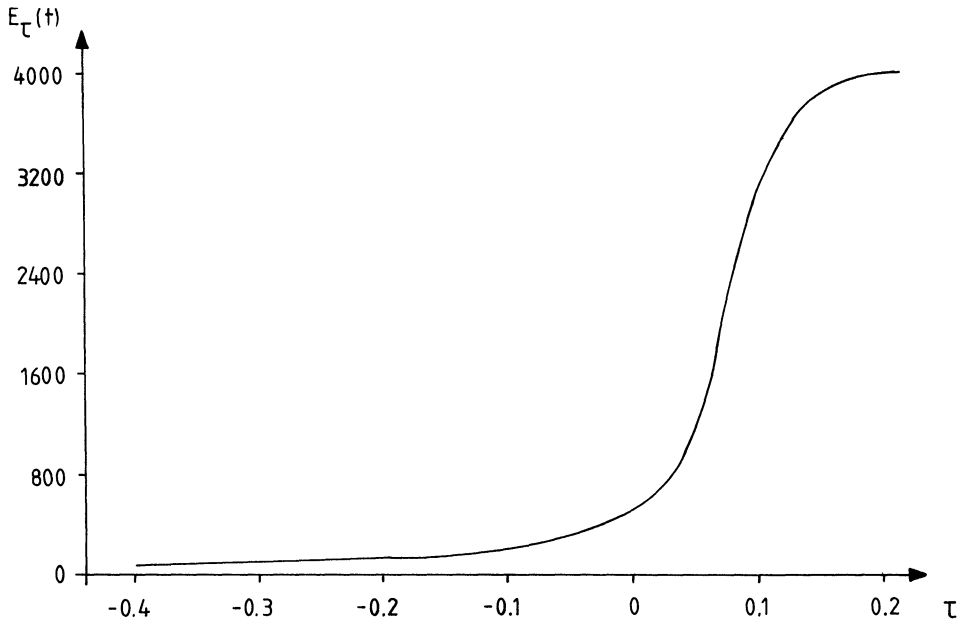

Figure 6. Expected Number of Triangles for a Transitivity Model With n = 30 Vertices.

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

Figure 7. Symmetrized Version of Sampson Monastery Data (indicators of some positive influence between the two members of a pair).

regression to the Sampson monastery data. Following Fienberg et al. (1985) we form an 18 × 18 table, the entries of which indicate the level of positive influence expressed between the pairs of monks. For simplicity we deal with a symmetrized table, where the entries indicate the existence of positive influence in at least one direction between the monks in the pair (see Fig. 7). There are 61 ones and 92 zeros among the $\binom{18}{2} = 153$ pairs.

To fit a general triad model with parameters ρ, σ, and τ we applied the logit regression approach. We computed s_{uv} and t_{uv} for each of the 153 pairs $\{u, v\}$ in Figure 7 and ran the logit regression option of the BMDP package. The results were

$$\tilde{\rho} = -.10, \qquad \tilde{\sigma} = -.23, \qquad \tilde{\tau} = .97. \tag{4.12}$$

The interpretation is complicated by the strong positive correlations between r, s, and t, but the rather large $\tilde{\tau}$ value evidently suggests a tendency toward transitivity of positive influence.

Testing the adequacy of (4.12) is not simple; in particular, likelihood ratio tests are not available because the Y_{uv} values are binary. One way of checking the fit is to see how well the Y_{uv} values are predicted by the discriminant function

$$\tilde{\rho} + \tilde{\sigma}s + \tilde{\tau}t \tag{4.13}$$

with a suitable cut point. We find that 119 of the 153 Y_{uv} values are correctly predicted. This should be compared with the fact that 92 of the 153 Y_{uv} values can be correctly predicted by the constant 0, which is the best prediction if $\sigma = \tau = 0$, that is, if we have a Bernoulli graph. Thus the odds of correct prediction increases from 92/61 = 1.5 to 119/34 = 3.5 by including σ and τ in the model.

5. DEPENDENCE STRUCTURES FOR DIRECTED RANDOM GRAPHS

We now consider directed graphs and give characterizations that are analogous to those for undirected graphs given in Section 3. This will show what kinds of sufficient statistics are useful and will provide the formalism needed for embedding dyad independence models and simple Markov models in larger classes of models that might be useful for testing goodness of fit. We comment briefly on some natural embeddings of this kind but present no inferential methods here. Such work is in progress. A particularly interesting consequence of our results is that it is possible to state what model assumptions are required to legitimate inference methods based on triad counts in digraphs. Triad count statistics have been much used with no recourse to an appropriate theoretical framework.

Let G be a directed graph on a set $N = \{1, \ldots, n\}$ of n individuals; that is, $G \subseteq N^{(2)}$. The adjacency matrix $Y = (Y_{uv})$ is defined by

$$\begin{aligned} Y_{uv} &= 1 \text{ if } (u, v) \in G \\ &= 0 \text{ otherwise} \end{aligned} \tag{5.1}$$

for $(u, v) \in N^{(2)}$.

The dependence structure of a random directed graph G is given by a graph D on $M = N^{(2)}$, which has an edge between (s, t) and (u, v) in M iff Y_{st} and Y_{uv} are dependent conditional on the rest of Y. Thus the vertices of D are all of the possible arcs in G, and the edges in D correspond to the conditionally dependent pairs of arcs in G. According to the Hammersley–Clifford theorem we have the following characterization of the probability structure of G.

Theorem 4. Any random directed graph G on N with dependence structure D has probability

$$\Pr(G) = c^{-1}\exp \sum_{A \subseteq G} \alpha_A, \tag{5.2}$$

where c is a normalizing constant and α_A is an arbitrary constant if A is a clique of D and $\alpha_A = 0$ otherwise.

This log-linear model has parameters α_A and sufficient statistics

$$\prod_{(u,v) \in A} Y_{uv} \tag{5.3}$$

for all A that are cliques of D. The sufficient statistics are indicators of subgraphs in G that are called *sufficient subgraphs.*

A directed graph G is said to be a *Markov graph* if its dependence graph D has edges only between these pairs (s, t) and (u, v) with an individual in common. Thus there is conditional dependence only between incident arcs of G. The cliques of D are now found to be the subsets of M that correspond to arcs, mutual arcs (2-cycles), stars of order 3 or more, and triangles of various kinds. The sufficient subgraphs of G are displayed in Figure 8, for $n = 4$.

Set $R^{(1)}_{uv} = \{(u, v)\}$, $R^{(2)}_{uv} = \{(u, v), (v, u)\}$, and let $T^{(k)}_{uvw}$ for $k = 1, \ldots, 7$ denote the seven nonisomorphic triangles on $\{u, v, w\}$. Let $S^{(i,j)}_{v_0 \cdots v_k}$ denote a star with center v_0 having in-arcs from the first i vertices and out-arcs to the last j vertices in $(v_1, \ldots, v_k)$. If $k < i + j$ there are $i + j - k$ mutual arcs between v_0 and each one of $v_{k-j+1}, \ldots, v_i$. The star $S^{(i,j)}_{v_0 \cdots v_k}$ is invariant against permutations of $(v_1, \ldots, v_{k-j})$, permutations of $(v_{i+1}, \ldots, v_k)$, and, if $k < i + j$, permutations of $(v_{k-j+1}, \ldots, v_i)$. Let π_k be

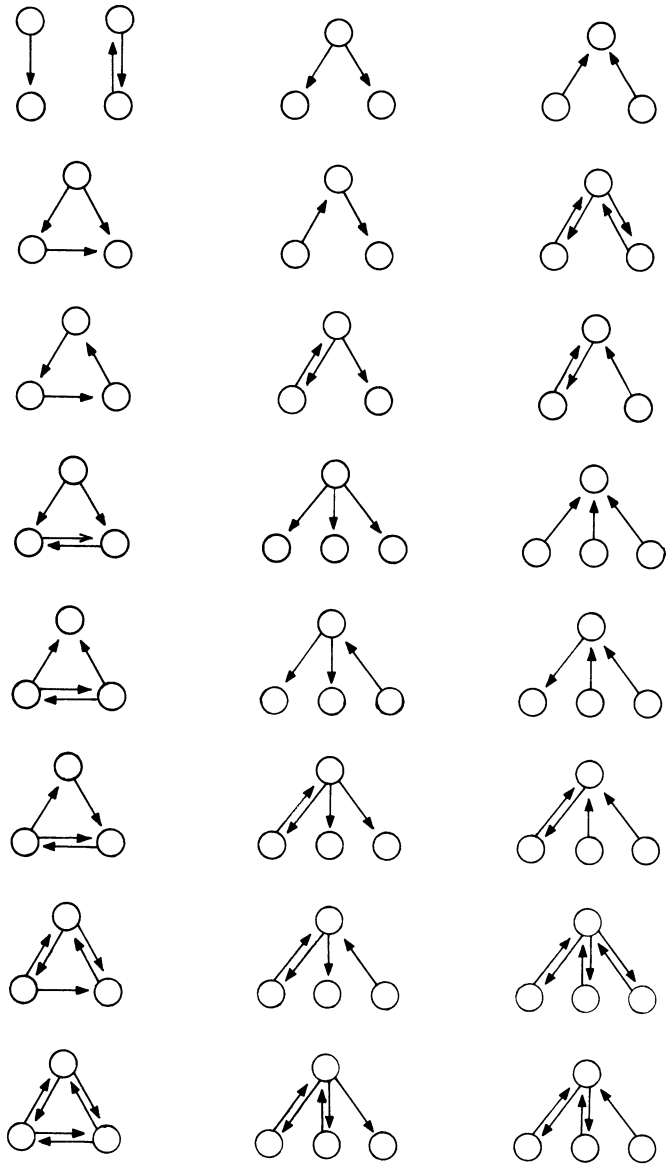

Figure 8. Sufficient Subgraphs of Directed Markov Graphs of Order n = 4.

the number of permutations of (u, v, w), leaving $T^{(k)}_{uvw}$ invariant.

There are $n^{(2)}$ arcs, $\binom{n}{2}$ mutual arcs, $n\binom{n-1}{k}3^k$ stars of order $k + 1$ for $k = 2, \ldots, n - 1$, and $27\binom{n}{3}$ triangles. In total the number of sufficient subgraphs is equal to

$$3\binom{n}{2} + 27\binom{n}{3} + \sum_{k=2}^{n-1} n\binom{n-1}{k}3^k = n(4^{n-1} - 1) + 27\binom{n}{3} - 3\binom{n}{2}. \quad (5.4)$$

The parameters corresponding to these subgraphs are denoted by

$$\begin{aligned} \rho^{(k)}_{uv} &= \alpha(R^{(k)}_{uv}), & k &= 1, 2, \\ \sigma^{(i,j)}_{v_0 \cdots v_k} &= \alpha(S^{(i,j)}_{v_0 \cdots v_k}), & k &= 2, \ldots, n - 1, \\ \tau^{(k)}_{uvw} &= \alpha(T^{(k)}_{uvw}), & k &= 1, \ldots, 7. \end{aligned} \quad (5.5)$$

We obtain the following result from Theorem 4.

Theorem 5. Any directed Markov graph has probability

$$\Pr(G) = c^{-1}\exp\Bigg[\sum_{k=1}^{2}\sum \rho^{(k)}_{uv}/k + \sum_{k=2}^{n-1}\sum_{i=0}^{k}\sum_{j=k-i}^{k}\sum \sigma^{(i,j)}_{v_0 \cdots v_k} \div (k - i)!(k - j)!(i + j - k)! + \sum_{k=1}^{7}\sum \tau^{(k)}_{uvw}/\pi_k\Bigg], \quad (5.6)$$

where the first unspecified sum is over $(u, v) \in N^{(2)}$ with $R^{(k)}_{uv} \subseteq G$, the second unspecified sum is over $(v_0, \ldots, v_k) \in N^{(k+1)}$ with $S^{(i,j)}_{v_0 \cdots v_k} \subseteq G$, and the last unspecified sum is over $(u, v, w) \in N^{(3)}$ with $T^{(k)}_{uvw} \subseteq G$. This log-linear model has

$$n(4^{n-1} - 1) + 27\binom{n}{3} - 3\binom{n}{2} \quad (5.7)$$

parameters and sufficient statistics corresponding to arcs, mutual arcs, stars of order 3 or more, and triangles of seven kinds.

We note that the Holland–Leinhardt model is obtained by setting all σ and τ parameters equal to 0, $\rho^{(2)}_{uv} = \rho$, and $\rho^{(1)}_{uv} = \alpha_u + \beta_v + \theta$ with $\sum \alpha_u = \sum \beta_v = 0$, so that there are $2n$ free parameters and sufficient statistics corresponding to out-degrees, in-degrees, and total number of mutuals.

By assuming homogeneity and putting

$$\begin{aligned} &\rho_k = \rho^{(k)}_{uv}, \qquad k = 1, 2, \\ &\sigma_{ijk} = \sigma^{(i,j)}_{v_0 \cdots v_k}, \\ &\quad 0 \le i \le k,\ 0 \le j \le k,\ k \le i + j,\ k = 2, \ldots, n - 1, \\ &\tau_k = \tau^{(k)}_{uvw}, \qquad k = 1, \ldots, 7, \end{aligned} \quad (5.8)$$

we see that the number of parameters is reduced to

$$9 + \sum_{k=2}^{n-1}\binom{k+2}{2} = 5 + \binom{n+2}{3}, \quad (5.9)$$

and we obtain the following result.

Theorem 6. Any homogeneous directed Markov graph has probability

$$\Pr(G) = c^{-1}\exp\Bigg[\sum_{k=1}^{2}\rho_k r_k + \sum_{k=2}^{n-1}\sum_{i=0}^{k}\sum_{j=k-i}^{k}\sigma_{ijk}s_{ijk} + \sum_{k=1}^{7}\tau_k t_k\Bigg], \quad (5.10)$$

where r_1 is the number of arcs in G, r_2 is the number of mutual arcs in G, s_{ijk} is the number of stars of order $k + 1$ having i in-arcs and j out-arcs in G, and t_k is the number of triangles of type k in G. The number of parameters and sufficient statistics is $5 + \binom{n+2}{3}$.

If we assume $\sigma_{ijk} = 0$ for $k > 2$, so that only 15 parameters remain, we see that the sufficient statistics are the numbers of arcs, mutual arcs, stars of different kinds of

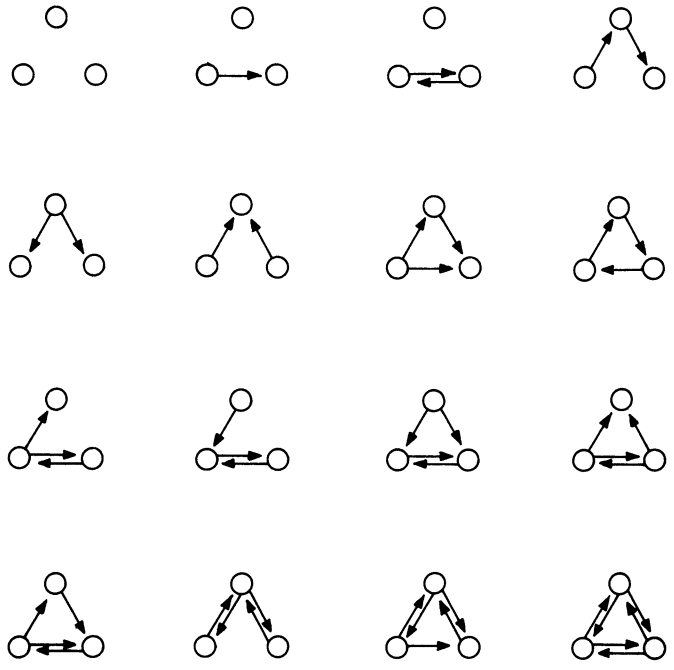

Figure 9. Triads in Directed Graphs.

order 3, and triangles of different kinds. An equivalent set of statistics is the set of triad counts, that is, the numbers of induced subgraphs of order 3 of all possible kinds. There are 16 different kinds of triads (displayed in Figure 9) and their counts sum to $\binom{n}{3}$, so there is need for only 15 of them as sufficient statistics. Triad count statistics in directed graphs have been used by Frank (1978), Holland and Leinhardt (1981), and others, and it is interesting to note that these statistics are sufficient for a directed homogeneous Markov graph having no parameters for stars of order 4 or more. If parameters are included also for stars of order 4, then the number of parameters increases from 15 to 25, and this should provide a natural extension of the triad count model within the class of Markov graphs.

6. DEPENDENCE STRUCTURES FOR COLORED RANDOM GRAPHS

The results in Sections 3 and 5 can be extended to cover the case of a matrix $Y = (Y_{uv})$ of not necessarily binary random variables. We will sketch some results without going into details of proof. We can think of Y as a network data matrix giving multiplicities, capacities, flows, or some other quantities defined on the arcs in a directed graph. We will assume that there are finitely many distinct values on Y_{uv}, and we will call these values colors and label them $0, \ldots, r$. The matrix Y gives a random coloring of a complete graph on N. For short we will call Y a colored graph on N. The conditional dependence structure in Y will be given by a dependence graph D on $M = N^{(2)}$. According to the Hammersley–Clifford theorem we obtain the following theorem.

Theorem 7. Any randomly colored graph Y on N with dependence structure D has probability

$$\Pr(Y = y) = c^{-1}\exp \sum_{A \subseteq M} \lambda_A(y_A), \tag{6.1}$$

where c is a normalizing constant given by

$$c = \sum_{y} \exp \sum_{A \subseteq M} \lambda_A(y_A) \tag{6.2}$$

and λ_A are functions satisfying

$$\sum_{y_{uv}=0}^{r} \lambda_A(y_A) = 0 \tag{6.3}$$

for any $(u, v) \in A$ if A is a clique of D and $\lambda_A(y_A) = 0$ otherwise.

The random coloring is said to be a Markov coloring if the colors of any two nonincident arcs are independent conditional on the colors of the other arcs in M. Thus the dependence graph D of a Markov coloring is identical to a dependence graph of a directed Markov graph as defined in the previous section and we have the same cliques as there. The number of parameters needed is found to be

$$n[(r+1)^{2(n-1)} - 1] + \binom{n}{3} r^3(r+2)^3 - \binom{n}{2} r(r+2). \tag{6.4}$$

Under homogeneity the number of parameters can be shown to be reduced to

$$f(r) + g(r) + \sum_{k=2}^{n-1} h_k(r), \tag{6.5}$$

where

$$f(r) = (r^2 + 3r)/2 \quad \text{and}$$

$$g(r) = (r^6 + 6r^5 + 12r^4 + 11r^3 + 8r^2 + 4r)/6 \tag{6.6}$$

are the numbers of ρ and τ parameters and

$$h_k(r) = \sum_{i=0}^{k} \sum_{j=k-i}^{k} \binom{r-1+k-i}{k-i} \times \binom{r-1+k-j}{k-j} \binom{r^2-1+i+j-k}{i+j-k} \tag{6.7}$$

is the number of σ parameters for k-stars.

If we restrict the model by assuming no σ parameters for $k > 2$, then the number of parameters is

$$f(r) + g(r) + h_2(r) = (r^6 + 6r^5 + 15r^4 + 23r^3 + 26r^2 + 19r)/6. \tag{6.8}$$

In particular, for $r = 1$ we retain 15 parameters corresponding to the 16 triads in Figure 8. For $r = 2$ we obtain 137 parameters, which can be shown to correspond to 138 different triads in directed graphs with 3 arc colors. Thus it is obvious that homogeneity alone is not sufficient to provide convenient models for valued graphs.

[*Received October 1983. Revised November 1985.*]

REFERENCES

Besag, J. E. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 36, 192–236.

Fienberg, S., Meyer, M., and Wasserman, S. (1985), "Statistical Analysis of Multiple Sociometric Relations," *Journal of the American Statistical Association,* 80, 51–67.
Frank, O. (1978), "Sampling and Estimation in Large Social Networks," *Social Networks,* 1, 91–101.
——— (1979), "Estimating a Graph From Triad Counts," *Journal of Statistical Computation and Simulation,* 9, 31–46.
Frank, O., and Harary, F. (1982), "Cluster Inference by Using Transitivity Indices in Empirical Graphs," *Journal of the American Statistical Association,* 77, 835–840.
Freeman, L. C., Romney, A. K., and White, D. R. (in press), "Research Methods in Network Analysis," *Social Networks.*
Hammersley, J. M., and Handscomb, D. C. (1964), *Monte Carlo Methods,* London: Methuen.
Holland, P. W., and Leinhardt, S. (1981), "An Exponential Family of Probability Densities for Directed Graphs," *Journal of the American Statistical Association,* 76, 33–51.
Karoński, M. (1982), "A Review of Random Graphs," *Journal of Graph Theory,* 6, 349–389.
Ogata, Y., and Tanemura, M. (1984), "Likelihood Analysis of Spatial Point Patterns," *Journal of the Royal Statistical Society,* Ser. B, 46, 496–518.
Pickard, D. K. (1977), "Asymptotic Inference for an Ising Lattice II," *Advances in Applied Probability,* 9, 476–501.
Rapoport, A. (1949), "Outline of a Probabilistic Approach to Animal Sociology II," *Bulletin of Mathematical Biophysics,* 2, 273–281.
——— (1950), "Outline of a Probabilistic Approach to Animal Sociology III," *Bulletin of Mathematical Biophysics,* 12, 7–17.
——— (1957), "Contribution to Theory of Random and Biased Nets," *Bulletin of Mathematical Biophysics,* 19, 257–277.
——— (1979), "A Probabilistic Approach to Networks," *Social Networks,* 2, 1–18.
Ripley, B. (1981), *Spatial Statistics,* New York: John Wiley.
Ruelle, D. (1969), *Statistical Mechanics: Rigorous Results,* New York: Benjamin.
Speed, T. P. (1978), "Relations Between Models for Spatial Data, Contingency Tables and Markov Fields on Graphs," *Supplement Advances in Applied Probability,* 10, 111–122.
Strauss, D. J. (1975), "A Model for Clustering," *Biometrika,* 62, 467–475.
——— (1977), "Clustering on Colored Lattices," *Journal of Applied Probability,* 14, 135–143.
——— (in press), "On a General Class of Models for Interaction," *SIAM Review Journal.*
Strauss, D. J., and Freeman, L. C. (in press), "Stochastic Modelling and the Analysis of Structural Data," *Social Networks.*
Wasserman, S. (1980), "Analyzing Social Networks as Stochastic Processes," *Journal of the American Statistical Association,* 75, 280–294.