A SURVEY OF MATHEMATICAL MODELS FOR GRAPHS

Stanley S. Wasserman, University of Minnesota

Introduction

The basic mathematical concept in this paper is the directed graph, or digraph, which is defined as a set V of nodes or "points" and a set L of directed arcs or "lines," connecting pairs of nodes. The set V contains g distinct elements, $v_1, v_2, \ldots, v_g$, and the set L contains C arcs, $\ell_1, \ell_2, \ldots, \ell_C$. We further require that no two distinct lines be in parallel; i.e., there exists at most one line ℓ_i connecting node v_j to node v_k. For convenience and to adhere to the established convention, a loop, a line connecting v_j to v_j, is not allowed in the digraph.

Digraphs differ from the more common undirected graphs because they have the additional characteristic that every line has an orientation or direction. Digraphs in which an arc from v_j to v_k implies the existence of an arc from v_k to v_j are symmetric. Symmetric digraphs are, of course, undirected graphs. When we desire to denote a directed line in terms of its two points, we write $\ell_i = v_j v_k$ for the directed line running from v_j to v_k. We let D_g be a specific digraph on g nodes. Note that D_g is a zero-one, or binary directed graph. The strengths or intensities attached to each arc are irrelevant, since our definition does not allow for the existence of valued lines. In what follows, we discuss mathematical representations for both directed graphs and undirected graphs, although we concentrate on the more general directed graph.

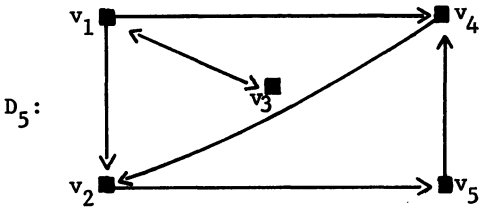
A digraph D_g is easily represented by a (gxg) matrix. We define a matrix $\underset{\sim}{X}$, with elements

$$X_{ij} = \begin{cases} 1, & \text{if } v_i v_j \in L \\ 0, & \text{otherwise.} \end{cases}$$

The matrix $\underset{\sim}{X}$ is called the adjacency matrix of D_g and has one row and one column for every node in V. An adjacency matrix for an undirected graph if, of course, symmetric. A different ordering of the elements in V produces an adjacency matrix that differs from $\underset{\sim}{X}$ by a simultaneous row-column permutation. Two digraphs with g nodes whose adjacency matrices differ by such a row-column rearrangement are called isomorphic. Note that since loops are not allowed, the diagonal elements of $\underset{\sim}{X}$, X_{ii}, i=1, 2,...,g, are set to zero.

Two sets of quantities are particularly interesting. The outdegree of node v_i, written r_i, is the number of arcs originating at node v_i. The indegree of node v_j, written c_j, is the number of arcs terminating at node v_j. Every element in $\underset{\sim}{r}$ and $\underset{\sim}{c}$ takes on a value between 0 and (g-1). Figure 1 shows an example of a digraph and associated adjacency matrix, including indegrees and outdegrees. The standard reference for these concepts is Harary, Norman, and Cartwright [1965].

This discussion of mathematical models for graphs is both a literature review and a collection of future suggestions for graph modelling. We present several models originally developed for processes other than graphs, giving model assumptions and a few derived results. We also comment on the applicability of these models to directed graphs, and in particular, social networks. Quite a few ideas for future research are given. We feel that a thorough understanding of existing models applicable to directed graphs is an essential prerequisite for the development of relevant and encompassing stochastic models for social networks.



$\underset{\sim}{X}$:

	v_1	v_2	v_3	v_4	v_5	r_i
v_1	0	1	1	1	0	3
v_2	0	0	0	0	1	1
v_3	1	0	0	0	0	1
v_4	0	1	0	0	0	1
v_5	0	0	0	1	0	0
c_j	1	2	1	2	1	

Figure 1: Digraph and Associated Adjacency Matrix.

Categorizing Mathematical Models

In addition to reviewing mathematical models specifically developed for graphs, we examine, in some detail, mathematical models originally constructed for entities other than directed graphs. The structure considered are models for processes from such natural science fields as statistical physics, organic chemistry and biology, and biophysics, easily modified to become models for graphs. The models from the literature for graphs themselves, are, by and large, from the social sciences, and postulate mathematical representations for social networks, specific sets of social relations linking members of well-defined groups. Mathematically, a social network may be defined as a binary directed graph with nodes for individual group members and arcs for the relational links.

Before we present the various models, we note that all mathematical models (for many kinds of processes) can be dichotomized twice: one by a deterministic vs. stochastic division and once by a static vs. time-dependent split. The first

dichotomy is a function of whether or not the model under consideration incorporates probabilistic assumptions. Deterministic models allow no opportunity for the graph to deviate from a prescribed pattern, usually given by a system of differential equations, or substantive theory. Stochastic models by definition do not allow the current or future structure of a graph to be predicted with certainty. The second dichotomy is solely a function of what the model postulates about time. Does the model in question assume that a graph evolves over time (time-dependence) or not (stasis)?

Deterministic, Static Models

Until very recently, the analysis of directed graphs, particularly in the social science context, was static and deterministic. Networks were not explicitly assumed to evolve over time, and conclusions drawn from a single network were deterministic, or precisely defined. The sociological version of Heider's balance theory (see Heider [1958]) was the prevalent paradigm. Leinhardt [1977] discusses the beginnings of network analysis, focusing on Heider's contribution. Heider's research was generalized by Cartwright and Harary [1956] in a paper where formal graph theory was introduced to social network research.

The predictions of Cartwright and Harary's structural balance theory did not accord with reality. Davis [1967], referring to the lack of empirical support for the dichotomous cliquing of groups predicted by Cartwright and Harary's theorem, further elaborated on the balance paradigm, extending it to multiple clusters of individual. However, the deterministic nature of the theory was retained, and consequently, the model's fit to empirical data remained poor. What was needed was a model incorporating probabilistic assumptions on the relations among group members. In a series of papers, Davis, Leinhardt, and Holland built a stochastic component into the paradigm.

Stochastic, Static Models

The Davis-Holland-Leinhardt methodology involves computing conditional uniform distributions on the space of all directed graphs. The most highly conditioned distribution controls for the dyad census, or the number of mutual, asymmetric, and null arcs in a digraph (see Holland and Leinhardt [1975]). Essentially, one computes the first two moments of the 16 component triad census, a count of the isomorphism classes of the $\binom{g}{3}$ triads in a digraph, and compares the empirically determined triad census with its expectation. Davis [1977] reviews this line of research, and Wasserman [1977a] discusses other random directed graph distributions.

This approach is static in time, since it concentrates on only one adjacency matrix. It is, however, stochastic. The analysis can even be compared to current methodology on stochastic processes. Holland and Leinhardt essentially compute equilibrium distributions for digraphs, and assume that data on the digraph process follow these distributions. One outstanding question is whether any of these "equilibrium" distributions are true equilibrium distributions obtained from some stochastic process. Further research may clarify this issue.

Deterministic, Time-Dependent Model

Several deterministic models for directed graphs have been proposed. Differential equations are the driving forces of such models, in which the effect of any change in the system can be predicted with certainty. However, in the social sciences, and to a lesser extent in the natural sciences, changes in a system cannot be predicted with certainty, usually because of the unpredictable nature of the objects being modelled. This uncertainty is best modelled through the use of probability distributions on random variables instead of the "controlling" mathematical variables of a system of differential equations. (A blend of the two approaches would be promising, but no such model has been developed.) We prefer to concentrate on the more realistic set of stochastic models, and we merely refer the reader to Bernard and Killworth [1977] for a recent review of deterministic models.

Stochastic, Time-Dependent Models

For the remainder of this paper, we discuss stochastic, time-dependent models, first from the social sciences, and then from the natural sciences.

The first model is the "Dynamic Model" of Holland and Heinhardt [1977a]. The Holland-Leinhardt stochastic model is actually an encompassing framework for the modelling of graphs, more general than an explicit statement on the evolution of digraphs through time. The framework operates on the individual arcs in L, the most elementary and basic level of a digraph. In Wasserman [1977b], we develop this modelling system theoretically, and discuss several simple parameterizations and estimation of structural parameters.

We next present three other models for social networks. These are a model in discrete time by Katz and Proctor [1959], a model based on learning theory of Rainio [1966], and a more recent model of Sørenson and Hallinan [1976].

Following the social science models, we discuss three models from the natural sciences. The first model that we shall discuss is for percolation processes of the flow of fluid through a medium. Broadbent and Hammersley [1957] give a mathematical formulation of percolation theory as it applies to crystals and mazes. Frisch and Hammersley [1963] present a thorough review of the theory, giving definitions and listing some of the results available at the time and unsolved problems.

Secondly, we shall describe a stochastic model for polymerization, or the evolution of polymers in organic chemistry. Polymers are "units" (or atoms) which associate into clusters and are also capable of disassociation. The model is Whittle's [1965a, 1965b], and is based on both the Gibbs equilibrium distribution for an ensemble of particles, and the deterministic kinetic equations of thermodynamics. The blend of these two approaches produces a unique set of stochastic kinetic equations as a model for polymerization.

Next, we discuss a model for neural networks of biophysics proposed by Rapoport. Rapoport's models of random and biased nets are not stochastic in nature; however, we include them here because the various types of biased nets are parallel to the simple stochastic models discussed by Wasserman [1977b]. Rapoport's notion of

"biases" may even be considered as the theoretical forerunner of the structural parameters of the Holland-Leinhardt framework. These models are presented in a group of papers written in the 1950's by Rapoport, in the Bulletin of Mathematical Biophysics. Rapoport [1957] reviews the contributions to the theory of random and biased nets, and Rapoport [1963] discusses the importance of nets to the theory of social interaction.

Throughout this section, we let $\underset{\sim}{X}(t)$ be the adjacency matrix representing the state of the digraph at time t. The binary-valued matrix $\underset{\sim}{X}(t)$ has elements $(X_{ij}(t))$ where

$$X_{ij}(t) = \begin{cases} 1, \text{ if } v_i v_j \in L \text{ at time } t \\ 0, \text{ otherwise.} \end{cases} \quad (1)$$

The time parameter, t, is assumed continuous, $t \geq 0$. Throughout, we set the g diagonal terms, $(X_{ii}(t))$, to 0 for all i and t.

We let $\underset{\sim}{x}$ be a single state of the continuous time stochastic process $\underset{\sim}{X}(t)$. The process has a finite state space $\underset{\sim}{S}$ of all possible $2^{g(g-1)}$ binary-valued (gxg) matrices with zero diagonal. We shall let $\underset{\sim}{w}$, $\underset{\sim}{x}$, $\underset{\sim}{y}$, $\underset{\sim}{z}$, ... denote elements of the state space.

Social Science Models

1. Holland-Leinhardt Framework

The Holland-Leinhardt framework is merely two simple assumptions regarding the stochastic nature of the arcs $X_{ij}(t)$. The first assumption is that $\underset{\sim}{X}(t)$ is Markov chain. Thus by Assumption 1, the current state of the process is all that we need to predict future behavior of the process.

We make an additional assumption regarding conditional independence of the elements of $\underset{\sim}{X}(t+h)$ given $\underset{\sim}{X}(t)$ for small h, conditional choice independence.

This is a critical assumption and in unique to this framework. It states that for very small intervals of time, the changes in a digraph are statistically independent. Consequently, the probability that any two arcs change simultaneously is essentially zero. This assumption is crucial for theoretical results, since it greatly simplifies the mathematics.

The infinitesimal transition rates depend on the entire adjacency matrix at time t, and may imply complex interrelations among the elements of $\underset{\sim}{x}$. Holland and Leinhardt [1977a, 1977b] and Wasserman [1977b] give examples of various functions specifically postulating that the infinitesimal transition rates of the digraph process are linear functions of various graph-theoretic quantities.

This line of inquiry into the nature of social structure and evolution of social networks is unique because of the proposed framework for parameterization. By assuming that, for small intervals of time, the arcs of a digraph operate in a statistically independent manner, we are able to assume various functional forms for the infinitesimal transition rates of the process. Thus, a researcher may define "social structure" by a set of graph-theoretic quantities, and combine these in a linear fashion to form the change rates of the process. This aspect of the framework is an important contribution to mathematical sociology, being an explicit statement on the evolution of a digraph as a continuous time Markov chain and providing a "wide open" framework for quantifying social structure.

2. Other Models from Social Science

There have been other attempts at modelling social networks as stochastic processes. In an early paper, Katz and Proctor [1959] analyze a sociomatrix at the level of dyads, or pairs of nodes in the digraph. The authors assume that the $\binom{g}{2}$ dyads are independent observations on a time discrete Markov chain, and therefore test whether a specific data set is compatible with the assumptions of a Markov chain. Unfortunately, no explicit structural model is developed for the evolution of a network over time.

Rainio [1966] develops a stochastic theory of social interaction. He posits a vector of probabilities, summing to unity, that regulates the frequency of interaction between individual i and the remaining (g-1) individuals in the group. These g vectors, one for each individual, evolve over time. This model is applied to a group of twelve girls, and the individual learning parameters (α,β) varied to provide the best fit of the model to the data. The model is very similar to the learning theory models developed by Bush and Mosteller [1955], and although its discrete time nature is a great simplification, it is an important contribution.

More recently, Sørenson and Hallinan [1976] hypothesize that each triad, or triple of nodes, in a network is a continuous time Markov chain. However, unlike the $\binom{g}{2}$ dyads in a network, the $\binom{g}{3}$ triads are not independent, and the assumption that the set of triads are independent observations on a basic Markov chain is incorrect. Unlike the model of Sørenson and Hallinan, the Holland-Leinhardt framework operates at the level of individual choices, the most elementary and basic level of a network. Placing a stochastic mechanism on the dyads or triads and ignoring subgraphs of lesser order is indeed less accurate in describing the operational behavior of a group.

Natural Science Models

1. Percolation Processes

Percolation theory seeks to describe the spread of a fluid throughout a medium. The random mechanism can either be attributed to the fluid or the medium: the former alternative is easily recognized as a diffusion process, while the latter is a percolation process. By its nature, percolation theory is more deterministic than diffusion theory, being subject to more restrictive assumptions, and certainly less widely known. The examples of percolation processes are many, ranging from fractures of crystals, and water absorption in a porous solid, to spread of blight in an orchard. Percolation theory stands apart from general epidemics (see Bartholomew [1973], Chapter 9 and 10) in that the medium under consideration is constrained by a particular geographic structure.

Percolation theory considers the following problem:

Let C be a connected graph with a countable set of nodes $\{X_i\}_{i=0}^{\infty}$ and arcs $\{L_{ij}\}$ joining X_i to X_j. Each arc L_{ij} is blocked, so that no fluid may traverse it, with probability $1-p_{ij}$, and unblocked, with probability p_{ij}. We then supply fluid to a random set of arcs, and study the flow of fluid through the system. This is the simplest case in percolation theory. More complicated situations are given in Frisch and Hammersley [1963].

As one can see, percolation theory might be quite important to the study of diffusion of innovations through a social network. Unfortunately, because of the complicated mathematics, it is virtually inaccessible to social scientists, and very rarely referenced.

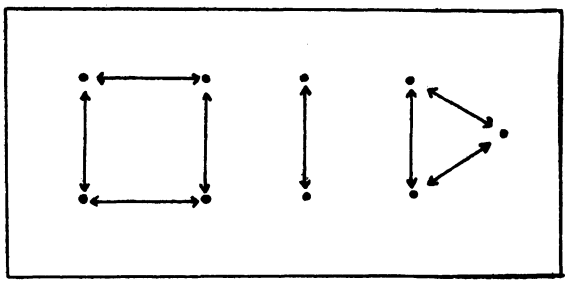
2. Polymerization Processes

We shall now consider Whittle's model for polymerization. The polymerization process has a state space of all symmetric graphs whose adjacency matrices can be permuted to block diagonal form, a space much smaller than S. An example of polymerization will be illustrated.

Suppose at time t, we have a group of size g that is composed of $k \leq g$ distinct clusters of nodes or cliques or polymers, such that no arcs exist between cliques, and within each clique, all arcs are present. Thus, each clique is strongly connected, and the adjacency matrix for the digraph can be permuted to a matrix with blocks of ones along the diagonal, one block per clique, and zeros elsewhere. Figure 2 depicts a situation with 9 nodes and 3 clusters.

When two-polymers come together, all arcs between the two come into existence, so that the new clique is also strongly connected. When a polymer disassociates into two new polymers, all arcs between the two smaller polymers disappear. Thus we always have a symmetric digraph with block-diagonal-permutable adjacency matrix. Note that the second assumption of the Holland-Leinhardt framework does not apply to Whittle's polymerization process, since, in general, a large number of arcs change simultaneously whenever polymers associate or disassociate.

Recently, in sociology, there has been renewed interest in clique formation. One of the proposed models for cliques in social networks (Breiger, Boorman, and Arabie [1975] and White, Boorman, and Breiger [1976]), the blockmodel, is not stochastic. Merging Whittle's stochastic model of group structure with the blockmodel proposed by White, et al, would be a substantial contribution to the analysis of social networks. There are other, simpler stochastic models for group changes. Morgan [1976] gives a review of these models, in addition to extending Whittle's results by proving the polymerization model to be reversible.



0	1	1	1	0	0	0	0	0
1	0	1	1	0	0	0	0	0
1	1	0	1	0	0	0	0	0
1	1	1	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	1	1	0

Figure 2: Polymerization Digraph and Permuted Adjacency Matrix.

3. Random and Biased Nets

Next, we discuss Rapoport's models for random and biased nets. Rapoport defines a random net as a binary directed graph with each node assigned a fixed outdegree a. This "fixed choice" adjacency matrix is only random in the conditional sense that every node in the group is equally likely to receive one of the a arcs of node i. The adjacency matrix is conditioned to have a fixed vector of outdegrees $(a,a,\ldots,a)'$.

Rapoport assumes a single adjacency matrix, with some fixed outdegree, and examines a tracing of the network. A tracing is merely a path through the net, beginning at an arbitrary number of nodes. We then record these nodes as the initial set and all new nodes that are chosen by the initial set are termed first remove,. This tracing is continued, all the while recording the fraction of the population present in the initial set, first remove, second remove, etc. By examining these fractions, Rapoport estimates a by α, the apparent choice or axone density. This choice density was found to deviate from a empirically.

In an attempt to explain this deviation, Rapoport defines certain biases, operating in nets, that could cause the discrepancy. These biases include distance bias, symmetry bias, and transitivity biases. Distance bias decreases the chance that two individuals far apart from one another in the constructed "social space" will interact. Symmetry bias increases the chance of a choice $j \to i$, if the choice $i \to j$ is present, and as defined, is identical to the μ_0 and μ_1 terms in the mutuality

model of Wasserman [1977b]. Transitivity biases have a similar interpretation as τ_0, τ_1, and τ_2 in the model presented in Holland and Leinhardt [1977a].

Unfortunately, Rapoport is able to do little with these biases mathematically, except to estimate gross statistical features of the graph. There is a strong relation between Rapoport's work and the models of mutuality, popularity, and expansiveness discussed in Wasserman [1977b] utilizing the modelling framework. A further study of his relation would be useful.

Concluding Remarks

We have discussed several models for graphs and compared the stochastic, time-dependent models to the new modelling framework proposed by Holland and Leinhardt [1977a]. Directions for future research are indicated throughout this paper, specific ideas concerning how existing models could be accurately represented by this new framework. This research should prove both promising and exciting, with additional insight into the evolution of social networks over time as an added benefit.

Acknowledgements

Support was provided by NSF Grant SOC73-05489 to Carnegie-Mellon University. I thank Stephen Fienberg, Maureen Hallinan, Kenneth Land, Samuel Leinhardt, and Anatol Rapoport for helpful suggestions.

References

Bartholomew, D.G. [1973] Stochastic Models for Social Processes, London: John Wiley & Sons.

Bernard, H.R. and P.D. Killworth [1977] "Deterministic models of social networks" Social Networks: Surveys, Advances, and Commentaries, edited by P.W. Holland and S. Leinhardt, New York: Academic Press, to appear.

Breiger, R.L., S.A. Boorman, and P. Arabie [1975] "An algorithm for clustering relational data with application to social network analysis and comparison with multidimensional scaling," J. of Math. Psych. 12, pp. 328-383.

Broadbent, S.R. and J.M. Hammersley [1957] "Percolation processes. Crystals and mazes," Proc. Cambridge Phil. Soc. 53, pp. 629-641.

Bush, R.R. and F. Mosteller [1955] Stochastic Models for Learning, New York: John Wiley & Sons.

Cartwright, D. and F. Harary [1956] "Structural balance: A generalization of Heider's theory," Psychological Review 63, pp. 277-293. Also in Social Networks: A Developing Paradigm, edited by S. Leinhardt. New York: Academic Press, 1977.

Chung, K.L. [1967] Markov Chains with Stationary Transition Probabilities, 2nd Edition, New York: Springer-Verlag.

Davis, J.A. [1967] "Clustering and structural balance in graphs," Human Relations 20, pp. 181-187. Also in Social Networks: A Developing Paradigm, edited by S. Leinhardt, New York: Academic Press, 1977.

Frisch, H.L. and J.M. Hammersley [1963] "Percolation processes and related topics," J. of Society for Indust. and Applied Math. 11, pp. 894-918.

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

Heider, F. [1958] The Psychology of Interpersonal Relations, New York: John Wiley & Sons.

Holland, P.W. and S. Leinhardt [1975] "Local structure in social networks," Sociological Methodology, 1976, edited by D.R. Heise, San Francisco: Jossey-Bass.

Holland, P.W. and S. Leinhardt [1977a] "A dynamic model for social networks," J. of Math. Soc. 5, pp. 5-20.

Holland, P.W. and S. Leinhardt [1977b] "Social structure as a network process," paper presented at the International Conference on "Mathematical Approaches in Social Network Analysis," Bad Homburg, West Germany.

Katz, L. and C.H. Proctor [1959] "The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process," Psychometrika 24, pp. 317-327.

Leinhardt, S. [1977] "Social networks: A developing paradigm," in S. Leinhardt, editor, Social Networks: A Developing Paradigm, New York: Academic Press.

Morgan, B.J.T. [1976] "Stochastic models of grouping changes," Adv. in Applied Prob. 8, pp. 30-57.

Rainio, K. [1966] "A study on sociometric group structure: An application of a stochastic theory of social interaction," Sociological Theories in Progress, Vol. 1, edited by J. Berger, M. Zelditch, and B. Anderson, Boston: Houghton-Mifflin.

Rapoport, A. [1957] "Contribution to the theory of random and biased nets," Bull. of Math. Biophy. 19, pp. 257-277. Also in Social Networks: A Developing Paradigm, edited by S. Leinhardt, New York: Academic Press, 1977.

Rapoport, A. [1963] "Mathematical models of social interaction," Handbook of Math. Psych. Vol. II, edited by R.D. Luce, R.R. Bush, and E. Galanter, New York: John Wiley & Sons.

Sørenson, A.B. and M. Hallinan [1976] "A stochastic model for change in group structure," Social Science Research 5, pp. 43-61.

Wasserman, S.S. [1977a] "Random directed graph distributions and the triad census in social networks," J. of Math. Soc. 5, pp. 61-86.

Wasserman, S.S. [1977b] "Stochastic models for directed graphs," Ph.D. Dissertation, Harvard University.

White, H.C., S.A. Boorman, and R.L. Breiger [1976] "Social structure from multiple networks. I. Block models of roles and positions," Amer. J. of Soc. 81, pp. 730-780.

Whittle, P. [1965a] "Statistical processes of aggregation of a clustering process in the uncondensed phase," Proc. of R. Soc., Series A 285, pp. 501-519.

Whittle, P. [1965b] "The equilibrium statistics of a clustering process in the uncondensed phase," Proc. of R. Soc., Series A 285, pp. 501-519.