The spread of a disease through human populations can be significantly altered by patterned networks of social contact. Largely in response to AIDS, the way in which social networks channel disease has recently become the focus of a sustained modeling effort in epidemiology. The challenge has been to develop a general framework capable of representing both simple and arbitrarily complicated mixing structures, and of solving the matching problem in a non-equilibrium multi-group population. This work is reviewed here. A recent contribution from network analysis links log-linear models of contact structure to diffusion equations for transmission. This framework is described in detail, and some applications to the spread of AIDS are described. The results show that careful analysis of structured mixing can reveal significant features of an epidemic that would otherwise be missed.

Epidemiology and Social Networks:
Modeling Structured Diffusion

MARTINA MORRIS
Columbia University

Social networks play an important role in the spread of infectious disease, but one that has been largely ignored. Perhaps due to the long and spectacular list of successes produced by medical biology, from vaccines to treatments to cures, disease has come to be framed almost entirely in biological terms. Biological factors clearly play a critical role in the spread of disease, regulating the transmissibility and natural history of the infectious agent. Social factors, however, are also important, as these regulate the patterns of interpersonal contact and thus the structure within which transmission is channeled.

In the field of epidemiology, attention has historically been restricted to biological factors. Modeling efforts have focused almost exclusively on natural history of the infectious agent: the period of infectiousness, the duration of latency or incubation, the duration of

AUTHOR'S NOTE: *Supported in part by grant SES-91-10798 from the National Science Foundation. Address reprint requests to M. Morris, Department of Sociology, Columbia University, New York, NY 10027.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 22, No. 1, August 1993 99-126

the illness, the rate of mortality, the existence of natural or exposure-related immunity, the role of disease-spreading vectors like insects. This tradition of research has produced both a well integrated framework of statistical estimation from epidemiological and experimental data, and simulation models for projection. The standard reference texts are Bailey (1975), Anderson (1982), and Anderson and May (1991). The dominant assumption made regarding social factors has been the assumption of random mixing (also referred to as homogeneous mixing). While this choice of focus might be expected in epidemiology, it is somewhat surprising that sociologists have not really challenged the biologically driven analysis of disease. It is as though we, too, are convinced that disease is not a proper topic for sociological inquiry.

Most people recognize that infectious disease has played an important role in social history. The most infamous cases include the disastrous effects of introducing European diseases, especially smallpox, into the Native American population, and the sequence of plagues in Europe (cf. McNeill 1976).

The flip side of this coin is that various structural aspects of social relations have always served to channel these diseases. Like other exchangeable goods, for example, information, wealth, occupational access, and norms, the diffusion of disease through a human population traces the structure of social networks. Different kinds of diseases travel along different structural routes. The plague, for example, is spread by a mobile vector of rats and fleas. This makes for an efficient, long-lasting infectious vehicle and opens up the potential for spread via long-distance transportation routes; in 14th-century Europe, this meant trade routes. In such cases, macro-economic structures dominated the diffusion path. With the measles, by contrast, there is a relatively short infectious period, and transmission involves either casual or indirect contact. Such diseases are more likely to evolve as geographically localized processes structured by centers of collective activity like schools and supermarkets. Finally, there are diseases spread only by intimate or prolonged contact, such as sexually transmitted diseases. These diseases travel along the most selective forms of social networks, operating on what is comparatively a microstructure.

From a modeling perspective, epidemiologists have made some progress incorporating the first two forms of spatially structured diffusion, but very little on the last. It is in this last area—the patterns of differential association—that the tools of social network analysis are beginning to be used.

STRUCTURED DIFFUSION IN EPIDEMIOLOGY

The foundations of modern mathematical epidemiology were laid in the early part of this century. Though, perhaps, never explicitly recognized as such, sociological dynamics have played a part in these models from the beginning. In one of the earliest contributions to formal modeling, the role of social contact in the spread of disease was expressed as a "mass action principle" (Hamer 1906). This principle defined the population dynamics of an epidemic in a simple and intuitive way, proportional to the product of three factors: the probability that one member of the contact is susceptible, the probability that the other is infected, and the number of effective contacts made between individuals per unit time. The first two factors are conventionally defined as the relative fraction of each group in the population. The third factor is a function of the number of contacts and the probability of transmission per contact. Despite the proliferation of additional terms in modern models of disease dynamics, this mass action principle remains central.

Some of the most important later contributions include the threshold theorem of Kermack and McKendrick (1927), which ties the outbreak of an epidemic to the density of susceptibles, the development of stochastic formulations to mimic the role of chance and variation (Bartlett 1955, 1956; Bailey 1975), and the parameterization of models in terms of a "reproductive rate," R, a measure of the number of new cases generated by each infective that serves as a practical index for measuring the rate of spread or for tracking the effectiveness of immunization efforts (Anderson 1982). Both stochastic and deterministic formulations are used in the field. Mollison (1977, 1991) provides an excellent, though non-social-science-oriented, review of the relative strengths, weaknesses, and comparative predictions of deterministic and stochastic models of diffusion. Analytic solutions for the dependence of the epidemic on key parameters is often impossible. With stochastic models, for example, analytic solutions are

typically only available for the pure birth versions of the models (known as simple epidemic models); birth-death processes (known as general epidemic models), even the simplest kind, are intractable (Bailey 1975, 88ff.). For this reason, much of the work involves either strong simplifying assumptions, asymptotic approximations, or simulation.

The stochastic-deterministic split in epidemic modeling is analogous to the sociometric-egocentric split in network analysis. Stochastic formulations, like sociometric analyses, model each individual node in a system and the links among them. In network analysis, such data are typically represented as a sociomatrix or sociogram. Deterministic formulations, like egocentric analyses, divide the population into relatively homogeneous groups and map the average rates of contact among them. These models are sometimes called "compartmental models" in epidemiology. In network analysis, such data are typically represented in the form of a contact or mixing matrix.

The most systematic work on structured diffusion has been in the area of spatial models (for a social science oriented review, cf. Bartholomew 1973, pp. 323-34). This work is typically based on stochastic models. Selection bias here operates as a simple function of physical proximity: the closer two individuals are, the higher the probability that they will come into contact. The position of individuals is modeled either in continuous space, using distributions from the exponential family (Fisher 1937; Van den Bosch, Metz, and Diekmann 1990), or in discrete space, with individuals arrayed on a lattice (Harris 1974; Cox and Durrett 1988). Because of their simplicity and ease of simulation, lattices are the most common framework for modeling two-dimensional space, and contacts are often restricted to the "nearest-neighbor" in the lattice (Harris 1974). The nearest-neighbor epidemic models are equivalent to percolation models under certain conditions (Mollison 1977, p. 302). The simplicity of the lattice comes at a price, however, as the shape of the grid (e.g., square with four neighbors, square with eight neighbors, or hexagonal) has been shown to affect the type of epidemic patterns observed in simulation (Lloyd 1991). Continuous diffusion approximations have also been used, and work is beginning to be done on modeling epidemics on random graphs (e.g., Mollison and Barbour 1989). Models for spatially distributed populations have been considered in Cliff and Ord (1975), Hethcote (1976, 1978), Mollison (1977), and May and Anderson (1984a, 1984b). Other references may be found in these articles.

For most social processes, however, factors other than physical distance contribute to biased mixing. Spatial models can be adapted to make use of some more socially relevant metrics. Hagerstrand (1967), a geographer interested in the spread of innovations, modified a simple spatial model by mapping the frequency of telephone contacts among regions and using these as the basis for probabilities of contact in his simulations. A similar approach has been used more recently in a simulation study of the global spread of influenza (cf. Longini, Fine, and Thacker 1986 for a review). Here, the volume of airline travel was used as the metric of distance among regions. A social distance metric could potentially be constructed by several commonly used methods (e.g., multidimensional scaling, factor analysis, log-linear models, and latent class analysis). The potential for spatial models to represent biased mixing is there and, as yet, largely untapped.

Outside of spatial models, few attempts were made to model nonrandom mixing until the late 1970s and early 1980s when, despite the relative availability of penicillin and other antibiotics, gonorrhea rates rose precipitously (see Figure 1). To understand and potentially control the population dynamics of this sexually transmitted disease, it was recognized that specific forms of selective mixing would have to be modeled.

Compartmental models were typically adopted for modeling selective mixing. In almost all cases, the models were built around contact or mixing matrices, a cross-tabulation of partnerships by the category of the subject and the category of the partner. Typically, both the group definitions and the mixing functions were very simple. Activity level was the most common characteristic used to form groups, distinguishing those with high levels of sexual activity (e.g., prostitutes, highly active homosexuals) from those with low levels. This approach was pioneered by Lajmanovich and Yorke (1976) and popularized by Hethcote and Yorke (1984) in their modeling of core populations in gonorrhea dynamics. In this early work, the rates of contact between groups were assumed to be governed by proportionate mixing (see also Barbour 1978; Nold 1980; Hethcote and Van Ark 1987). Under this assumption, one group's exposure to another is proportional to the total number of contacts made by each group rather than the total

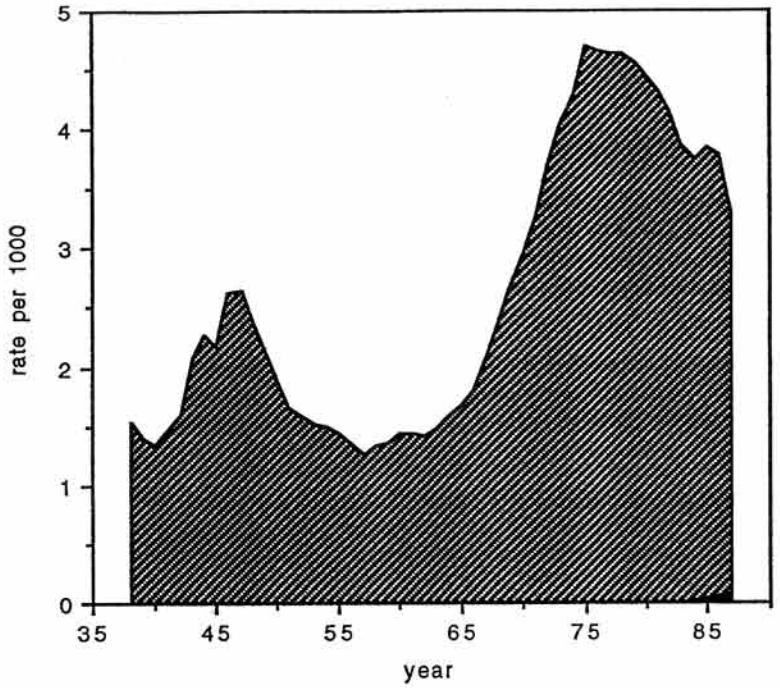

Figure 1: U.S. Gonorrhea Rate, 1938-1987
Source: *Morbidity and Mortality Weekly Review, Summary of Notifiable Diseases*, 1987.

number of people, a small variation on the original random mixing formulation. Letting π_{ij} denote the proportion of group *i*s contacts that are with *j*s,

$$\pi_{ij} = \frac{c_j N_j}{\sum_k c_k N_k} \qquad (1)$$

$$= \pi_j,$$

where $c_{(.)}$ is the group-specific average number of partners per unit time, and $N_{(.)}$ is the number in the group. The π_{ij}s are conditional probabilities that sum to one across the index *j*. This formulation, while

it recognizes population heterogeneity, still assumes homogeneous mixing; every group has the same fraction of their contacts with group *j*. Proportionate mixing is equivalent to the statistical assumption of independence in the contact matrix. Under this assumption, for example, female prostitutes and highly active male homosexuals, by virtue of their high relative activity levels, would have a great deal of interaction. What is still ignored is social selectivity.[1]

The first real model for selection bias in epidemiology, the "like-with-like" or "preferred mixing" model, introduced a single parameter to reflect the tendency to choose a partner from one's own group (Nold 1980; Hethcote, Yorke, and Nold 1982; Hethcote and Yorke 1984; Hethcote and Van Ark 1987; Jacquez, Simon, Koopman, Sattenspiel, and Perry 1988; Anderson, Gupta, and Ng 1990). The bias parameter in this model is similar in function to the homophily bias parameter used in network analysis (Fararo and Sunshine 1964). Under this model, the fraction of contacts between group *i* and group *j* is given by

$$\pi_{ij} = \begin{cases} \rho + (1-\rho)\pi_j & i=j \\ (1-\rho)\pi_j & i \neq j, \end{cases} \qquad (2)$$

where ρ is the homophily bias parameter, the fraction of contacts reserved for partners from the same group and π_j is the proportional mixing fraction defined in equation (1) above, the fraction of all contacts that are made by group *j*. As the homophily bias varies from 0 to 1, mixing varies from proportional to restricted (within-group contacts only). One of the earliest papers to use this approach was concerned with the role of selection bias in the relative effectiveness of three contact-tracing methods in the control of gonorrhea (Hethcote, York, and Nold 1982). Using an eight-group model defined by gender, symptomatic status, and high or low activity level, and a single parameter to measure homophily, they found that the relative efficiency of different contact-tracing methods was directly related to the degree of homophily. In simulation models for the spread of AIDS, the introduction of a homophily parameter was found to slow the rise of incidence of the disease (De Gruttola and Mayer 1988) and to result in the generation of sequential waves of infection as the virus moves through the different groups in the population (**Hyman** and Stanley

1988). This approach has also been generalized to allow for differential homophily, group-specific biases with a parameter ρ_i for each group (Jacquez, Simon, and Koopman 1989). While homophily bias models, uniform and differential, introduce an important form of selective mixing, they do so only on the diagonal of the contact matrix; outgroup contacts continue to be allocated proportionally.

Before the early 1980s, there are virtually no models in which preference orderings for outgroup contacts can be introduced. One exception is a paper by Lajmanovich and Yorke (1976). Here, no constraints are placed on the variability in contact rates among groups, and there are few specific findings. They do prove that if the disease does not die out, it will persist at levels in each group that are independent of the levels of initial infection. Another exception is the simulation study by Elveback, Ackerman, Gatewood, and Fox (1971). This study examined the effects of structuring population contact by families, neighborhoods, school, and play groups, with contact rates assumed to vary in each group. School and play groups are shown here to function as bridges between families and neighborhoods, increasing the size of epidemics. The mixing patterns in these studies are not really modeled; they are more a collection of ad hoc assumptions. The attention given to the role of social factors in structuring the mixing process, in contrast to activity levels and disease stages, marks a more sociologically grounded formulation than previous work had shown.

With the emergence of AIDS in the early 1980s, the role of selective mixing became the focus of a sustained modeling effort. Some of this work assumes a particular structure of mixing, but the goal has been to develop more general representations of arbitrarily complicated structures. Sattenspiel (1987a, 1987b), an anthropologist, extended the earlier work of Elveback et al. (1971), systematizing the informal mixing assumptions of the earlier work into migration matrix models. She simulated the spread of hepatitis through a child population with a two-level mixing structure: children mixed within neighborhoods, and some children from each neighborhood mixed in day-care centers. The children who attended day-care centers thus functioned as bridges among the neighborhoods. In these simulations, the size of the epidemic was found to depend on (a) the origin of the outbreak, and (b) the density of intragroup relations. In Gail, Preston, and Piantadosi (1989), the contact matrix is constructed by using a mixture of available data or knowledge, and an assumption of proportionate mixing where data are absent. The mixture is determined by minimizing the discrimination information (Kullback-Liebler distance) from proportionate mixing matrix to the matrix of known constraints. Again through simulation, Gail found that the degree of heterogeneity in mixing directly affects the optimal testing strategy for the control of AIDS. In the most general formulations (Busenberg and Castillo-Chavez 1989; Jacquez et al. 1989; Koopman, Simon, Jacquez, and Park 1989; Blythe, Castillo-Chavez, Palmer, and Cheng 1991), arbitrarily complicated mixing structures can be represented. These models, however, tend to be highly over-parameterized, with little attention paid to either interpretation or estimation of the parameters.

A simpler model has recently been proposed by Gupta and Anderson (1989). As with the homophily bias approach, the level of assortative mixing is summarized here by one parameter, but the parameter can index disassortative as well as assortative mixing. The parameter, 'Q' is given by

$$Q = \frac{(\sum_{i=1}^{N} w_i - 1)}{(N - 1)}$$

where the w_i are the eigenvalues of an $N \times N$ mixing matrix. At the assortative extreme, this measure tends to unity, at the disassortative extreme to $-\frac{1}{N}$, and for proportional mixing, it takes the value 0. Simulations show that assortative mixing generates the most rapid initial growth, while disassortative generates larger epidemics over the long term. The model has been applied to a sociometric data set from a male homosexual community in Iceland (Haraldsdottir, Gupta, and Anderson 1992) where activity levels were used to define subgroups. The observed mixing pattern was found to be disassortative and generated a larger epidemic than would be the case under proportional or assortative mixing. While this approach is more tractable than some of the over-parameterized models described above, the single parameter it employs is more of a mathematical convenience than a substantive summary. The interpretation of the value of 'Q' is meaningful only with respect to 0 and the extreme values. The parameter also does not

specify how selection patterns change as the population profile changes (an issue that will be dealt with in detail below). These dynamics thus remain outside the model and must be separately addressed in an ad hoc fashion.

A slightly different focus can be found in models of pair formation and dissolution (Dietz 1987, 1988; Dietz and Hadeler 1988; Waldstätter 1989). Where the classical models of infectious disease assume a constant mixing between males and females, pair-formation models address the issue of temporary monogamous coupling. Such coupling can introduce significant transient effects at the early stage of an epidemic. Couples in which both members are infection-free are effectively "immune" for the duration of the partnership, and a significant number of such couples will slow the spread of the epidemic relative to a noncoupled population. Technically, pairing is introduced by creating new categories for the population reflecting single or paired status, and new parameters for the flows between categories reflecting the pairing and separation rates. Dietz and Hadeler's (1988) primary focus is on deriving analytic solutions for the dependence of the reproductive rate, R, on the pairing parameters. This becomes fairly difficult due to the matching constraints introduced by multigroup pairing, so they divide the mixing population only by gender. The findings from this work suggest that the persistence of the disease (endemic equilibrium) is directly linked to the separation rate and the number of partnerships during the life course. Given a low probability of disease transmission (.001 - .002 per contact), the minimum number of lifetime partners needed for persistence of the disease is about 12. For extremely short partnerships (3 - 4 days) more than 1,000 partners over the lifetime would be needed for persistence.

There are thus several different traditions in epidemiology for modeling structured mixing. An important distinction is between the spatial models, which enumerate all of the nodes and links in a network, and the contact matrix models (of which the pair-formation models are a special case), which operate on aggregate units linked by average rates of contact. As noted above, this distinction is essentially equivalent to the sociometric and egocentric traditions in network analysis. The remainder of this article will focus largely on the latter approach. Part of the motivation for this choice of focus is the availability of data. The necessity for completely enumerating the

nodes and links in a sociometric analysis makes the collection of such data feasible only for small bounded communities. In addition, when the "links" involve sensitive information (as with sexually transmitted diseases), contact tracing is intrusive, subject to significant refusal bias, and raises serious confidentiality issues. Egocentric data, while subject to other forms of reporting bias, can be collected using standard survey methods and raises fewer confidentiality issues. While the egocentric approach introduces some heterogeneity within groups and loses information on triad or transitivity bias, it tends to be more appropriate for large-scale epidemic modeling.

In moving from simple to general formulations of a mixing structure, two problems arise. The first is complexity. For an $n \times n$ contact matrix, there are potentially n^2 parameters that must be identified and integrated into the diffusion equation; more if there are changes over time. For example, when the subpopulation sizes are not stable or when activity levels are allowed to change, additional parameters are needed to model the process of selection in a nonequilibrium structure. Thus, in some of the general models cited above, the total number of parameters used to model the process is as much as four times the number of cells in the matrix.

This raises the second problem, that of selecting and justifying the parameter settings. The logical choice would be to base these settings on empirical data, though the lack of network data often makes this impossible. Empirically based parameters require a framework for estimation, however, and a statistically sound estimation framework is noticeably lacking from all of the papers which propose selective mixing models.

These issues, then, provide criteria for judging among competing models. There is no unique representation for a given contact matrix. Some formulations are inherently limited, like proportional or preferred mixing, but even these can be parameterized in different ways. Arbitrary mixing functions, which have proportional or preferred mixing as special cases, can also be represented in many ways. A particular representation, then, can not claim superiority on the basis of being general. A claim to superiority can, however, be based on a model's ability to solve the two problems identified above: manageable complexity, and a framework for estimation. Log-linear models can make this claim. These models make it possible to parsimoniously

describe a mixing structure, they generate parameters that have clear and relatively intuitive interpretations, and they provide a sound statistical basis for estimating these parameters from data.

The remainder of this article reviews a log-linear based framework for analyzing the effects of networks on diffusion (Morris 1991a). While the immediate application is to epidemiology, this framework can be used for any socially structured diffusion process. In sociology, as in epidemiology, the issue of structured diffusion has been explored (Rapoport 1953, 1957, 1979, 1980; Coleman 1964, pp. 492-514; Coleman, Katz, and Menzel 1966; Skvoretz 1985; Granovetter 1978, 1983; Granovetter and Soong 1988), but a general analytic framework has been lacking.

A GENERAL FRAMEWORK FOR SELECTIVE MIXING

In almost all models for diffusion that rely on interpersonal contact, transmission is represented by a term resembling

$$\text{Transmission rate} = \beta(t)S(t)I(t), \tag{3}$$

where S and I are the number of susceptibles and infecteds, and β is the "coefficient of conversion" or transmission coefficient.[2] In contagion models, $\beta(t)$ can be decomposed into two components, one representing activity levels or contacts and the other representing the conditional probability of transmission:

$$\beta(t) = \frac{c(t)\tau}{N(t)} \tag{4}$$

and

$c(t)$ = the number of contacts per person per unit time,

τ = the probability of transmission given contact,

$N(t) = S(t) + I(t)$, the total active population.

As formulated here, this term reflects the assumption of random mixing. Susceptibles and infecteds are two otherwise undifferentiated groups that mix in proportion to their numbers in the population. They

mix randomly with respect to their disease status and with respect to other social characteristics, such as age, race, and gender. For simplicity, the transmission coefficient, τ, is represented as constant over time. As with other biological assumptions, simplicity is chosen here in order to focus attention on the mixing issues. These assumptions may be easily relaxed.

To relax the assumption of random mixing, the infection rate term must be modified to reflect the subgroups in the population and the mixing structure among them. A model for selective mixing replaces $\beta(t)S(t)I(t)$ with the term $\sum_{ij} \beta_{ij}(t)S_i(t)I_j(t)$, where the subscripts i and j index the subgroups in the population and $\beta_{ij}(t)$ the transmission coefficient between them:

$$\beta_{ij}(t) = \frac{c_i(t)\pi_{ij}(t)\tau_{ji}}{N_j(t)}, \tag{5}$$

$c_i(t)$ = the average number of contacts per unit time for members of group i,

$\pi_{ij}(t)$ = the conditional probability of a partner from group j given a subject from group i,

τ_{ji} = the per-partner probability of transmission from j to i,

and

Nj(t) = the total number in group j.

Note that the transmission coefficient τ_{ji} need not be symmetric. In the case of AIDS, for example, the transmission coefficient from male to female is believed to be higher than the coefficient from female to male. Status asymmetries may be present in other transmission contexts. There is, however, an implicit symmetry constraint in the number of contacts between groups, and this constraint will be addressed at length below.

The level of grouping is flexible. It can be used to index individuals or to index broader categories, such as age, race, gender, religion, and cross-tabulations of these characteristics. Ideally, the level is chosen to ensure homogeneity within the groups created, but the choice of level is also constrained by the type of data available. The partnerships

can be displayed as a mixing matrix, with the rows representing subject-groups and the columns representing partner-groups. The cell entries of this matrix represent dyads rather than individuals, and the subject-group classification is, by convention, identical to the partner-group classification, so the mixing matrix is square. Where the index refers to individuals, this will be a sociomatrix.

Log-linear models are an appropriate method for summarizing the structure of this mixing matrix (Holland and Leinhardt 1981; Fienberg and Wasserman 1981). When the data are in the form of a sociomatrix, the models can be used to investigate the effects of in-degree, out-degree, and attribute bias (Wasserman and Faust 1993 provide a systematic review); to estimate block models for node-group bias (cf. Wang and Wong 1987); and to summarize forms of dyad dependence that lead to transitivity bias (Frank and Strauss 1986; Strauss and Ikeda 1990). When the data are in the form of a group-mixing matrix, log-linear models can be used to investigate various patterns of within- and between-group selection bias.[3] For nominal parameters, patterns of homophily can be represented by one or more factor levels for diagonal cells (cf. Marsden 1988, for an example). Such homophily bias models represent the mixing continuum for the hypotheses most commonly used in epidemiology: proportional, preferred, and restricted mixing. Models for nominal factors also generalize to arbitrary mixing functions, however, so that ingroup and outgroup preferences of all kinds may be examined. Ordinal models can be used when the location of a cell relative to the main diagonal can be summarized and interpreted as the distance between the two groups. A wide range of ordinal models exists, from a simple one-parameter model to the log-multiplicative models for estimating row and column scores (Goodman 1984, pp. 66-111).

The estimated parameters are also easily integrated into the diffusion model. Because the cell counts of the contact matrix, x_{ij} can be expressed as a function of the row total and the conditional row probabilities,

$$x_{ij}(t) = N_i(t)c_i(t)\pi_{ij}(t), \tag{6}$$

the transmission coefficient $\beta_{ij}(t)$ can now be expressed in terms of the mixing matrix cell counts

$$\beta_{ij}(t) = \frac{x_{ij}(t)\tau_{ji}}{N_i(t)N_j(t)}. \tag{7}$$

The mixing matrix cell counts, in turn, can be expressed in terms of an appropriate log-linear model, thus linking the summary parameters of the mixing structure to the diffusion process.

The dynamic nature of these processes, however, raises an additional problem. For many reasons, including demographic cycles, changes in behavior, and differential rates of removal from the active population, the relative sizes of subgroups in the population may not be stable over time. The model must, therefore, specify how people alter their mixing patterns in response to changes in the population profile.

What do people do when their preferred partners become more (or less) available? This is an interesting substantive question, and the answer will vary from one context (say, rumor spreading) to the next (say, sexual behavior). For any contact process, however, it is still possible to define the range of possible answers. As a preferred partner group becomes less available, an individual can either substitute with partners from other, less preferred groups, or reduce the number of partners they have.

This choice is constrained by the behavior of other groups, that is, it must obey a contact consistency constraint: the number of contacts between *i*s and *j*s must equal the number of contacts between *j*s and *i*s,[4] and this implies symmetry in the contact matrix, $x_{ji}(t)$ must equal $x_{ji}(t)$. In practice, any empirical contact matrix will display some departure from symmetry due to a combination of sampling variability and reporting error. In theory, however, the constraint must hold, and empirical matrices will have to be adjusted accordingly. Given equation (6) above, this means

$$N_i(t)c_i(t)\pi_{ij}(t) = N_j(t)c_j(t)\pi_{ji}(t). \tag{8}$$

Group sizes $[N_{(.)}(t)]$ change independently over time. In order to ensure that this equation remains satisfied over time, corresponding changes must be made to either or both of the remaining parameters, $c_{(.)}(t)$ and $\pi_{(..)}(t)$. In substantive terms, as the relative availability of groups changes over time, subjects must either adjust their contact rates or their selection patterns or both.

Fixed contact rates = "pure drive"

As the parameterization of the constraint suggests, the possible behavioral responses can be arrayed along a continuum. At one end, the contact rates can be fixed, with all compensation operating through selection patterns [i.e., $c_{(.)}(t) = c_{(.)}$]. This can be seen as a "pure drive" model of contacts: each individual has an inherent drive or quota to achieve, and as one group becomes less available, partners are simply chosen from more available groups in order to fill the quota.

The way in which selection patterns change in this model is constrained but not determined by the population changes. This can be seen by manipulating the contact equation above. Letting α_i be the proportional change in group size and β_{ij} be the proportional change in selection patterns, substitution into equation (8) gives

$$\alpha_i N_i(t)\ c_i\ \beta_{ij}\pi_{ij}(t) = \alpha_j N_j(t)\ c_j\ \beta_{ji}\pi_{ji}(t),$$

and, with simplification, the dependence of selection on population changes can be made explicit,

$$\beta_{ij} = \frac{\alpha_j}{\alpha_i}\beta_{ji}.$$

degrees of freedom

At any point in time, two things are fixed: population changes, $\alpha_{(.)}$s, and the margins of the contact matrix (by virtue of the fixed contact rates). The compensating changes in selection patterns thus have $\frac{n(n-1)}{2}$ degrees of freedom, the degrees of freedom in a square, symmetric, fixed-margin matrix. This flexibility means that the model can generate massive changes in partnership patterns as the population structure changes, a feature that may not be appropriate when modeling sexual (or other) partnerships. In fact, the notion of selection "preference" here is almost misleading, as the model makes selection primarily a function of drive and availability.

Fixed selection patterns = "pure selection"

At the other end, the selection patterns can be fixed with all compensation operating through contact rates [i.e., $\pi_{(..)}(t) = \pi_{(..)}$]. This can be seen as a "pure selection" model of behavior and might reflect some notion of personal identity. Here, people choose partners in fixed ratios, the mix rather than the quantity being the important aspect. Using the approach above, and letting γ_i represent the proportional change in contact rates (the $c_{(.)}$s) this time, substitution into equation (8) and simplification gives

$$\gamma_i = \frac{\alpha_j}{\alpha_i}\gamma_j.$$

One solution to this will always be $\gamma_i = \alpha_i^{-1}$, the group changes its contact rate in exact inverse proportion to its change in size, but the general solution above implies that for any set of population changes, choosing one of the contact rate adjustments (γ_j) sets the rest. In other words, while there may be different contact rates for every group, there is only one degree of freedom under the pure selection model for setting them all. It is as if the group with the ability to enforce its sexual activity level drives the rest of the population.

degrees of freedom

Neither of these models is very satisfying or realistic. In between the two extremes that they define lie what might be called "modified selection" models. In these models, individuals accommodate changes in relative group size by adjusting both contact rates and selection patterns. The way in which this mix is obtained is open to modeling, though, in general, this flexibility entails a loss of direct control over the basic behavioral components, $c_{(.)}$ and $\pi_{(..)}$. In return, it becomes possible to separate and directly model the effects of population composition and preferences.

modified selection models

The population composition can be regarded as an opportunity structure or market, a constraint within which people negotiate their preferences. This suggests a different decomposition of the contact matrix cell counts. Rather than the traditional margin × conditional probability equation (6), cell counts can be modeled as a function of population structure and mixing preferences. Population structure is then free to change over time, and mixing preferences may be treated as either stable or changing, depending on the substantive context. In simple form, this decomposition might look like

$$x_{ij}(t) = \frac{N_i(t)N_j(t)}{N(t)}\alpha_{ij}, \qquad \alpha_{ij} = \alpha_{ji} \tag{9}$$

$$= K_{ij}(t)\alpha_{ij}$$

where $K_{ij}(t)$ represents the population or opportunity structure, and the mixing preference term, α_{ij}, could reflect a mutual signaling dynamic between the two potential partners, for example,

$$\alpha_{ij} = cs_{ij}s_{ji}.$$

The s_{ij} terms can be interpreted as the probability that a member of group i signals yes to a member of group j, a function of activity level preference and selection preferences for members of group i. These are not required to be symmetric. The scale factor c represents the number of contact opportunities, that is, the average number of new people met per unit time. Under this model, behavior now varies over time in response to relative group sizes, and the preferences of both partners contributes to the probability of contact being made. These two features enhance the substantive credibility of the model.

The α_{ij} preference term may be treated as either constant or time-dependent. In the absence of longitudinal data, it would seem reasonable to assume, as a first approximation, that the preferences are constant. The reasonableness of this assumption clearly varies with context. Sexual preferences, for example, are likely to be less volatile than brand loyalties. The assumption of constant preferences, in any case, does not constrain behavior to be constant, as behavior is also a function of changing opportunities, $K_{ij}(t)$.

Perhaps most importantly, the α_{ij} term from equation (9) can be estimated from data using standard techniques under some reasonable conditions (cf. Morris 1991a, Appendix). Assuming constant preferences, for example, one can estimate the α_{ij} term from contact matrix data at any time point using a log-linear model. Letting m_{ij} represent the expected cell counts under the model,

$$\hat{m}_{ij}(0) = K_{ij}(0)\hat{\alpha}_{ij} \tag{10}$$

I don't really understand this

$$= \exp\left\{\log K_{ij}(0) + \hat{u} + \hat{u}_{1(i)} + \hat{u}_{2(j)} + \sum_{l} \hat{u}_{l(ij)}\right\}.$$

The first three u terms represent the reference category and the main row and column effects, respectively. The last term, $u_{l(ij)}(0)$, represents the l'th interaction term, either the standard row × column interaction, or a specific contrast, for example a diagonal parameter. These provide a summary of the mixing preference structure, an estimate of $\hat{\alpha}_{ij}$. The $K_{ij}(0)$ term, which represents only population size, is assumed

to be known and to have a coefficient of 1, so it can be fit as an offset (cf. Aitken 1989, p. 127).

The estimates of $\hat{\alpha}_{ij}$ can then be used to incorporate selective mixing into a diffusion model. In a simulation setting, the new subgroup sizes computed at each time step can be used to calculate $K_{ij}(t)$, the population structure. The mixing matrix cell counts can then be updated using the new population structure and the estimates of the preferences obtained by the log-linear model,

$$\hat{m}_{ij}(t) = K_{ij}(t)\hat{\alpha}_{ij}.$$

These estimates can then be substituted back into the infection rate term of the diffusion model (Equation 7), completing the updating algorithm and establishing the final link in the modeling framework:

$$\hat{\beta}_{ij}(t) = \frac{\hat{m}_{ij}(t)\tau_{ji}}{N_i(t)N_j(t)}, \tag{11}$$

$$= \frac{\hat{\alpha}_{ij}\tau_{ji}}{N(t)}.$$

In short, log-linear models can be used to estimate the underlying selection preference structure. These estimates can then be used to update the contact patterns and thus the infection rate, as the population profile changes over time.

The framework reviewed here provides a simple way to model the widest possible variety of mixing patterns and incorporate these patterns into dynamic models of diffusion. Log-linear methods make it possible to systematically explore and summarize network structure. The conditional likelihood ratio statistic can be used to monitor the accuracy-parsimony trade-off, and the size and standard error of the parameter estimates can be used to identify the major sources of differential association. The modified selection model and the updating algorithm allow the estimated mixing parameters to drive a diffusion simulation even as the population profile changes. It is, in short, an integrated and general framework for exploring the role of social structure in the dynamic of social diffusion.

SOME RESULTS IN PROJECTING THE SPREAD OF AIDS

For diseases that require only casual or unintentional contact, like influenza, the assumption of random mixing may not introduce serious distortion. For a sexually transmitted disease like AIDS, however, this assumption implies that sexual partners are chosen at random and is clearly untenable. The sexual opportunities available to an individual, and the partners deemed appropriate, will vary systematically from one social group to the next. Demographic attributes like age, race, ethnicity, gender, sexual preference, and marital status play an overwhelming role in defining the boundaries of appropriate sexual partner groups. Other attributes, like occupation, education, social status, and religion play a similarly important role. The evidence for such selective association in friendship (e.g. Laumann 1973; Verbrugge 1977; Fischer 1982), and marriage formation (Hunt 1940; Centers 1949; Pagnini and Morgan 1990) is overwhelming. In projecting the sexual spread of AIDS, therefore, selective mixing cannot be ignored.

While the log-linear framework proposed above solves the methodological challenge posed by selective mixing, the virtual absence of empirical data remains a problem. Reliable data on sexual behavior are rare, sexual network data (either egocentric or sociometric) even more so. A limited range of data are available from various sources, however, including census data on age matching for first marriages, General Social Survey data on contact rates in the general adult population, and one network-oriented AIDS survey. These data have made it possible to use the log-linear modeling framework to explore how selective mixing might potentially affect the spread of the epidemic.

One of the most important findings is that the type of attribute used to define partner selection rules makes a difference. When these attributes are relatively stable over an individual's lifetime, like gender, race, or sexual preference, stable mixing groups are likely to form. The potential for spread between two groups that are not directly connected by sexual interaction then (e.g., gay men and heterosexual women) depends on the existence and size of a bridge population. By contrast, when the attributes that define partner selection rules change over an individual's lifetime (e.g., age or marital status), more fluid mixing groups are formed and the potential for spread is much higher. Here the mixing structure is characterized by a dual transmission

regime, with some infections carried into a group by individuals whose group membership changes, and other infections passed between groups by sexual contact. This kind of mixing structure makes the epidemic much more likely to spread, even with lower rates of sexual contact. The two attribute types thus generate two distinct transmission regimes. The implication is that the amount of behavioral change required, and the most strategically important subgroups from an intervention standpoint, depend on the type of mixing structure in which people are embedded (Morris 1989, 1990).

Another significant finding is that the effects of selective mixing interact strongly with other characteristics of the transmission process, most importantly, variations in initial seroprevalence (fraction of the population infected) and subgroup activity levels. Using one of the few available data sets on networks of sexual partners, the AIDS in Multi-Ethnic Neighborhoods Survey (AMEN; Catania, Coates, Kegels, and Fullilove 1992) the effects of selection patterns based on racial and ethnic boundaries was examined (Morris 1991b). These effects were shown to be fairly complex.

The initial effects of mixing depend strongly on initial prevalence: Mixing has little effect if there is uniform prevalence across all groups, but fairly large effects if initial prevalence varies. Where initial prevalence in a subgroup is high, positive assortative mixing acts to intensify within-group spread, increasing rates of infection relative to proportional mixing. Where initial seroprevalence is low, on the other hand, positive assortative mixing helps to keep it low. These initial effects may be transient, however. If the group has a relatively lower contact rate, the lower rate will be amplified by within-group selection and eventually lead to lower levels of infection under selective mixing. If the group has a relatively higher contact rate, on the other hand, within-group selection amplifies the rate of transmission, leading to higher levels of infection under selective mixing. Where both initial prevalence and activity level are low, and the homophily bias is relatively strong, selective mixing always leads to lower levels of infection.

Finally, these simulations have also shown that even when the effects of selective mixing are small at the aggregate level, they may still be quite pronounced at the subgroup level. In the AMEN-based analyses described above, the effects of selective mixing on the total

number infected at each point in the epidemic path was fairly small: Selective mixing led to about 1% to 2% less infection over time. The effects at the subgroup level, however, tended to be much larger and more variable, with selective mixing generating as much as 15% to 20% more (or less) infection in each subgroup. Such subgroup effects are also critical from an intervention standpoint. They become visible more quickly, and they help to pinpoint where intervention is most needed and most likely to be effective.

Overall, the simulations suggest that there is some potential for a heterosexual epidemic under the conditions examined. These models suggest, however, that the first evidence of such an epidemic will not appear at the aggregate level, but rather at the subpopulation level. If the infection is just beginning to spread in the non-drug-using heterosexual population, noticeable growth in the overall number of AIDS cases will not occur for decades. The overall incidence of AIDS may actually even decline for some period of time, but in the presence of selective mixing, such overall trends will be misleading. Much attention is currently being given to the recent reduction in the *overall* rate of growth in new AIDS cases (cf. Centers for Disease Control 1991a, Figure 5). The rate of growth in heterosexual cases, however, while still small, continues to rise (cf. Centers for Disease Control 1991b, Figures 1a and 1b). This type of pattern is consistent with the evolution of an epidemic in a selectively mixing population. To interpret the overall decline as meaning that the epidemic has run its course, therefore, is probably premature. Careful monitoring of sentinel groups, guided by selective mixing models, can help to distinguish between short- and long-term dynamics in this epidemic.

At the same time, it is important to keep in mind the limitations of the simulation approach. The characteristics used to define the mixing groups in these simulations—age, race, ethnicity, and sexual preference—are clearly not the only relevant characteristics used for partner selection. Other characteristics, such as marital status and socioeconomic status, need to be considered, and variation in activity levels within groups is also likely to be significant. Theoretically, this poses no problem for the modeling framework, as one can easily expand the mixing matrix to permit more specific mixing group definitions. Practically, however, increasing the number of mixing groups exponentially raises the number of cells in the mixing matrix that have to

be filled, requiring exponential increases in sample size. In addition, the reliability and validity of respondent reports on partner characteristics become problematic when one moves from asking the respondents to report their partner's race/ethnicity to the number of partners their partner has had in the last year.

As a general rule, simulations of the sort presented here are best used for gaining insight into the dynamics of diffusion through typical mixing structures. Demands for realism or accuracy will seldom be satisfied, but understanding may still be enhanced. Finally, it is worth noting, again, that the compartmental model approach adopted here precludes the investigation of certain types of network biases, most importantly triad or transitivity bias. It can be expected that triads would function much as the couples in the pair-formation models reviewed above and would reduce the speed of within-group spread. This is an area that should be addressed in future research.

CONCLUSION

A decade ago, a paper on epidemiology and social networks would have seemed like a strange idea. Things have changed quickly. In large part, the change has been "disease driven," but this should not obscure the methodological and theoretical gains that both fields now stand to gain from collaboration. At this point, there is great deal of interest among epidemiological modelers in developing methods for analyzing how social networks channel epidemics. Among sociologists, by contrast, the response has been small; practically none have become involved in the modeling effort (Klovdahl 1985; Klovdahl et al. 1991; Laumann, Gagnon, Michaels, Michael, and Coleman 1989; and Morris 1990, 1991a, 1991b are the exceptions). This is unfortunate for two reasons. From a practical perspective, projections for the spread of diseases like AIDS may be very misleading if the effects of selective mixing are ignored. From a theoretical perspective, diffusion processes offer an interesting vantage point for research into the dynamic aspects of social structure. The challenge is to formally represent the way that social structure transforms individual behavior into an often unintended collective outcome.

Both sociology and epidemiology have something to gain from collaboration here, as their strengths and weaknesses are complementary. Where sociology is strong on the theory of selective mixing, and weak on the mathematical depiction of diffusion under these conditions, epidemiology is strong on the mathematics of diffusion, and weak on the modeling of selective mixing. A major barrier between the fields has been the absence of a general analytic framework that bridges the mixing and diffusion processes. The framework reviewed in this article provides such a bridge and offers a general method for modeling diffusion through social structures.

NOTES

1. Proportionate mixing is assumed by another popular method of incorporating heterogeneity in activity levels, the coefficient of variation approach developed by Anderson, Medley, May, and Johnson (1986).

2. When conversion is a function of both contagion and innovation (a noncontagious process, like an external source), the rate of change in adopters (infecteds) reflects both components: $(\alpha + \beta I)S$. For a discussion of the stochastic version of this process, see Coleman (1964, pp. 343-53). This kind of model is commonly used in market research (Bass 1969, pp. 217-18) and is equivalent to the simple contagion model when $\alpha = \beta I(0)$ (Bartholomew 1973, p. 299).

3. When subjects have more than one partner (a form of clustered sampling), this may introduce some correlation among contacts within persons, raising questions about the statistical independence of the observed dyads. Dependence would result if respondents filled rigid quotas of partner types; the type of partner for the later contacts, then, would not be independent of the type of partner for the earlier contacts. Given the low average number of partners (per year) for most of the population, this problem may not be a serious one. Yamaguchi (1990) presents methods that can be used to assess this assumption.

4. This is essentially equivalent to the "two-sex" problem in demographic modeling (Pollard 1973; Pollack 1986). Implicit in the model that will be reviewed here is an alternative solution to the two-sex problem. Note that the transmission coefficient τ_{ij} does not have to be symmetric, so transmission may be more likely in one direction than the other. Only the contact process must be symmetric.

REFERENCES

Aitkin, M., D. Anderson, B. Francis, and B. Hind. 1989. *Statistical Modeling in GLIM*. Oxford: Oxford University Press.

Anderson, R. M. 1982. *The Population Dynamics of Infectious Diseases*. London: Chapman Hall.

Anderson, R. M., S. Gupta, and W. Ng. 1990. "The Significance of Sexual Partner Contact Networks for the Transmission Dynamics of HIV." *Journal of Acquired Immune Deficiency Syndromes* 3:417-29.

Anderson, R. M. and R. M. May. 1991. *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press.

Anderson, R. M., G. F. Medley, R. M. May, and A. M. Johnson. 1986. "A Preliminary Study of the Transmission Dynamics of the Human Immunodeficiency Virus (HIV), the Causative Agent of AIDS." *IMA Journal of Mathematics Applied in Medicine and Biology* 3:229-63.

Bailey, N.T.J. 1975. *The Mathematical Theory of Infectious Diseases*. New York: Hafner Press.

Barbour, A. 1978. "Macdonald's Model and the Transmission of Bilharzia." *Transactions of the Royal Society for Tropical Medicine and Hygiene* 72:6-15.

Bartholomew, D. J. 1973. *Stochastic Models for Social Processes*. New York: Wiley.

Bartlett, M. S. 1955. *Stochastic Processes*. Cambridge: Cambridge University Press.

———. 1956. "Deterministic and Stochastic Models for Recurrent Epidemics." Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability 4:81-109.

Bass, F. M. 1969. "A New Product Growth Model for Consumer Durables." *Management Science* 15:215-27.

Blythe, S., C. Castillo-Chavez, J. S. Palmer, and M. Cheng. 1991. "Toward a Unified Theory of Sexual Mixing and Pair Formation." *Mathematical Biosciences* 107:379-407.

Busenberg, S. and C. Castillo-Chavez. 1989. "Interaction, Pair Formation and Force of Infection Terms in Sexually Transmitted Diseases." Pp. 289-300 in *Mathematical and Statistical Approaches to AIDS Epidemiology*, edited by C. Castillo-Chavez. Berlin: Springer-Verlag.

Catania, J. A., T. J. Coates, S. Kegels, and M. T. Fullilove. 1992. "Condom Use in Multi-Ethnic Neighborhoods of San Francisco: The Population-Based AMEN (AIDS in Multi-Ethnic Neighborhoods) Study." *American Journal of Public Health* 82:287.

Centers for Disease Control. 1991a. *HIV/AIDS Surveillance Report*. January:1-22.

———. 1991b. *Morbidity and Mortality Weekly Report (MMWR)* 40(June 7):357-69.

Centers, R. 1949. "Marital Selection and Occupational Strata." *American Journal of Sociology* 54:530-35.

Cliff, A. D. and J. K. Ord. 1975. "Model Building and the Analysis of Spatial Pattern in Geography." *Journal of the Royal Statistical Society* Series B 37:297-348.

Coleman, J. S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.

Coleman, J. S., E. Katz, and H. Menzel. 1966. *Medical Innovation*. New York: Bobbs-Merrill.

Cox, J. T. and R. Durrett. 1988. "Limit Theorems for the Spread of Epidemics and Forest Fires." *Stochastic Process and Applications* 30:171-91.

De Gruttola, V. and K. H. Mayer. 1988. "Assessing and Modeling Heterosexual Spread of the Human Immunodeficiency Virus in the United States." *Review of Infectious Diseases* 10:138-50.

Dietz, K. 1987. "Epidemiological Models for Sexually Transmitted Infections." Proceedings of the First World Congress Bernoulli Society, Tashkent, 1986.

———. 1988. "On the Transmission Dynamics of HIV." *Mathematical Biosciences* 90:397-414.

Dietz, K. and K. P. Hadeler. 1988. "Epidemiological Models of Sexually Transmitted Diseases." *Journal of Mathematical Biology* 26:1-25.

Elveback, L. R., E. Ackerman, L. Gatewood, and J. P. Fox. 1971. "Stochastic Two-Agent Epidemic Simulation Models for a Community of Families." *American Journal of Epidemiology* 93:267-80.

Fararo, T. J. and M. H. Sunshine. 1964. *A Study of a Biased Friendship Net*. Syracuse, NY: Syracuse University Press.

Fienberg, S. and S. Wasserman. 1981. "Analyzing Data from Multivariate Directed Graphs: An Application to Social Networks." Pp. 289-306 in *Interpreting Multivariate Data*, edited by V. Barnett. London: Wiley.

Fischer, C. S. 1982. *To Dwell Among Friends: Personal Networks in Town and City*. Chicago: University of Chicago Press.

Fisher, R. A. 1937. "The Wave of Advance of Advantageous Genes." *Annals of Eugenics* 7:355-69.

Frank, O. and D. Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81:832-42.

Gail, M., D. Preston, and S. Piantadosi. 1989. "Disease Prevention Models of Voluntary Confidential Screening for Human Immunodeficiency Virus (HIV) in Isolated Low Risk and High Risk Populations and in Mixed Gay/Heterosexual Populations." *Statistics in Medicine* 8:59-81.

Goodman, L. A. 1984. *The Analysis of Cross-Classified Data Having Ordered Categories*. Cambridge: Harvard University Press.

Granovetter, M. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83:1420-43.

———. 1983. "Threshold Models of Diffusion and Collective Behavior." *Journal of Mathematical Sociology* 9:165-79.

Granovetter, M. and R. Soong. 1988. "Threshold Models of Diversity: Chinese Restaurants, Residential Segregation, and the Spiral of Silence." *Sociological Methodology* 69-104.

Gupta, S. and R. Anderson. 1989. "Networks of Sexual Contacts: Implications for the Pattern of Spread of HIV." *AIDS* 3:807-17.

Hagerstrand, T. 1967. *Innovation Diffusion as a Spatial Process*. Chicago: University of Chicago Press.

Hamer, W. H. 1906. "Epidemic Disease in England." *Lancet* 1:733-39.

Haraldsdottir, S., S. Gupta, and R. Anderson. 1992. "Preliminary Studies of Sexual Networks in a Male Homosexual Community in Iceland." *Journal of AIDS* 5:374-81.

Harris, T. E. 1974. "Contact Interactions on a Lattice." *Annals of Probability* 2:969-88.

Hethcote, H. W. 1976. "Qualitative Analysis for Communicable Disease Models." *Mathematical Biosciences* 28:335-56.

———. 1978. "An Immunization Model for a Heterogeneous Population." *Theoretical Population Biology* 14:338-49.

Hethcote, H. W. and J. W. Van Ark. 1987. "Epidemiological Models for Heterogeneous Populations: Proportionate Mixing, Parameter Estimation, and Immunization Programs." *Mathematical Biosciences* 84:85-118.

Hethcote, H. W. and J. A. Yorke. 1984. *Gonorrhea Transmission Dynamics and Control*. Berlin: Springer-Verlag.

Hethcote, H. W., J. A. Yorke, and A. Nold. 1982. "Gonorrhea Modeling: A Comparison of Control Methods." *Mathematical Biosciences* 58:93-109.

Holland, P. and S. Leinhardt. 1981. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 77:33-50.

Hunt, T. C. 1940. "Occupational Status and Marriage Selection." *American Sociological Review* 5:495-504.

Hyman, J. M. and E. A. Stanley. 1988. "Using Mathematical Models to Understand the AIDS Epidemic." *Mathematical Biosciences* 90:415-74.

Jacquez, J. A., C. P. Simon, and J. Koopman. (1989). "Structured Mixing: Heterogeneous Mixing by the Definition of Activity Groups." Pp. 301-15 in *Mathematical and Statistical Approaches to AIDS Epidemiology*, edited by C. Castillo-Chavez. Berlin: Springer-Verlag.

Jacquez, J. A., C. P. Simon, J. Koopman, L. Sattenspiel, and T. Perry. 1988. "Modelling and Analyzing HIV Transmission: The Effect of Contact Patterns." *Mathematical Biosciences* 92:119-99.

Kermack, W. O. and A. G. McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London* Series A 115:700-21.

Klovdahl, A. S. 1985. "Social Networks and the Spread of Infectious Diseases The AIDS Example." *Social Science and Medicine* 21(11):1203-16.

Klovdahl, A. S., J. Potterat, D. Woodhouse, J. Muth, S. Muth, and W. W. Darrow. 1991. "HIV Infection in an Urban Social Network: A Progress Report." Sunbelt Social Network Conference, Tampa, FL.

Koopman, J., C. P. Simon, J. A. Jacquez, and T. S. Park. 1989. "Selective Contact Within Structured Mixing with an Application to HIV Transmission Risk from Oral and Anal Sex." Pp. 316-49 in *Mathematical and Statistical Approaches to AIDS Epidemiology*, edited by C. Castillo-Chavez. Berlin: Springer-Verlag.

Lajmanovich, A. and J. A. Yorke. 1976. "A Deterministic Model for Gonorrhea in a Nonhomogeneous Population." *Mathematical Biosciences* 28:221-36.

Laumann, E. O. 1973. *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.

Laumann, E. O., J. H. Gagnon, S. Michaels, R. T. Michael, and J. S. Coleman. 1989. "Monitoring the AIDS Epidemic in the United States: A Network Approach." *Science* 244:1186-89.

Lloyd, M. 1991. "The Effect of Grid Shape on an Epidemic Growth Model." Unpublished manuscript.

Longini, I. M., P.E.M. Fine, and S. B. Thacker. 1986. "Predicting the Global Spread of New Infectious Agents." *American Journal of Epidemiology* 123(3):383-91.

Marsden, P. V. 1988. "Homogeneity in Confiding Relations." *Social Networks* 10:57-76.

May, R. M. and R. M. Anderson. 1984a. "Spatial Heterogeneity and the Design of Immunization Programs." *Mathematical Biosciences* 72:83-111.

———. 1984b. "Spatial, Temporal and Genetic Heterogeneity in Host Populations and the Design of Immunization Programmes." *IMA J. Math. Appl. Med. Biol.* 1:233-66.

McNeill, W. H. 1976. *Plagues and Peoples*. New York: Anchor.

Mollison, D. M. 1977. "Spatial Contact Models for Ecological and Epidemic Spread." *Journal of the Royal Statistical Society* Series B 39:283-326.

———. 1991. "Dependence of Epidemic and Population Velocities on Basic Parameters." *Mathematical Biosciences* 170:255-87.

Mollison, D. M. and A. Barbour. 1989. "Epidemics and Random Graphs." Pp. 86-89 in *Stochastic Processes in Epidemic Theory*, edited by J. P. Gabriel, C. Lefevre, and P. H. Picard. Berlin: Springer-Verlag.

Morris, M. 1989. "Networks and Diffusion: An Application of Log-Linear Models to the Population Dynamics of Disease." Ph.D. thesis, University of Chicago.

———. 1990. "Networks and Diffusion: Modeling the Effects of Selective Mixing on the Spread of Disease." *American Journal of Sociology* (under revision).

———. 1991a. "A Log-Linear Modeling Framework for Selective Mixing." *Mathematical Biosciences* 170:349-77.

———. 1991b. "Racial and Ethnic Boundaries in the Spread of AIDS." Sunbelt Social Network Conference, Tampa, FL.

Nold, A. 1980. "Heterogeneity in Disease-Transmission Modeling." *Mathematical Biosciences* 52:227-40.

Pagnini, D. L. and S. P. Morgan. 1990. "Intermarriage and Social Distance Among U.S. Immigrants at the Turn of the Century." *American Journal of Sociology* 96:405-32.

Pollak, R. 1986. "A Reformulation of the Two-Sex Problem." *Demography* 23:247-59.

Pollard, J. H. 1973. *Mathematical Models for the Growth of Human Populations*. Cambridge: Cambridge University Press.

Rapoport, A. 1953. "Spread of Information Through a Population with Socio-Structural Bias: I and II." *Bulletin of Mathematical Biophysics* 15:523-33, 535-43.

———. 1957. "Contribution to the Theory of Random and Biased Nets." *Bulletin of Mathematical Biophysics* 19:257-77.

———. 1979. "Some Problems Relating to Randomly Constructed Biased Networks." Pp. 119-136 in *Perspectives on Social Network Research*, edited by P. W. Holland and S. Leinhardt. New York: Academic Press.

———. 1980. "A Probabilistic Approach to Networks." *Social Networks* 2:1-18.

Sattenspiel, L. 1987a. "Epidemics in Nonrandomly Mixing Populations: A Simulation." *American Journal of Physical Anthropology* 73:251-65.

———. 1987b. "Population Structure and the Spread of Disease." *Human Biology* 59(3):411-38.

Skvoretz, J. 1985. "Random and Biased Networks: Simulations and Approximations." *Social Networks* 7:225-61.

Strauss, D. and M. Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85:204-12.

Van den Bosch, F., J.A.J. Metz, and O. Diekmann. 1990. "The Velocity of Spatial Population Expansion." *Journal of Mathematical Biology* 28:529-65.

Verbrugge, L. M. 1977. "The Structure of Adult Friendship Choices." *Social Forces* 56:1286-1309.

Waldstatter, R. 1989. "Pair Formation in Sexually-Transmitted Diseases." Pp. 260-274 in *Mathematical and Statistical Approaches to AIDS Epidemiology*, edited by C. Castillo-Chavez. Berlin: Springer-Verlag.

Wang, Y. Y. and G. Y. Wong. 1987. "Stochastic Blockmodels for Directed Graphs." *Journal of the American Statistical Association* 82:8-19.

Wasserman, S. and K. Faust. 1993. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Yamaguchi, K. 1990. "Homophily and Social Distance in the Choice of Multiple Friends." *Journal of the American Statistical Association* 85:356-66.

Martina Morris is an assistant professor of sociology at Columbia University. Her research examines the effects of network structures on epidemiological transmission processes. She is currently investigating the role of sexual networks in the spread of AIDS in the United States, Thailand, and Uganda, using egocentric network data.