An Assessment of Preferential Attachment as a Mechanism for Human Sexual Network Formation
Author(s): James Holland Jones and Mark S. Handcock
Source: *Proceedings: Biological Sciences*, Vol. 270, No. 1520 (Jun. 7, 2003), pp. 1123-1128
Published by: The Royal Society
Stable URL: http://www.jstor.org/stable/3558837
Accessed: 03/10/2009 01:52

THE ROYAL SOCIETY

Received 22 November 2002
Accepted 3 February 2003
Published online 11 April 2003

An assessment of preferential attachment as a mechanism for human sexual network formation

James Holland Jones[1,2,3*] **and Mark S. Handcock**[1,2]

[1]*Center for Statistics and the Social Sciences,* [2]*Center for Studies in Demography and Ecology, and* [3]*Center for AIDS and Sexually Transmitted Diseases, University of Washington, Seattle, WA 98195, USA*

Recent research into the properties of human sexual-contact networks has suggested that the degree distribution of the contact graph exhibits power-law scaling. One notable property of this power-law scaling is that the epidemic threshold for the population disappears when the scaling exponent ρ is in the range $2 < \rho \leq 3$. This property is of fundamental significance for the control of sexually transmitted diseases (STDs) such as HIV/AIDS since it implies that an STD can persist regardless of its transmissibility. A stochastic process, known as preferential attachment, that yields one form of power-law scaling has been suggested to underlie the scaling of sexual degree distributions. The limiting distribution of this preferential attachment process is the Yule distribution, which we fit using maximum likelihood to local network data from samples of three populations: (i) the Rakai district, Uganda; (ii) Sweden; and (iii) the USA. For all local networks but one, our interval estimates of the scaling parameters are in the range where epidemic thresholds exist. The estimate of the exponent for male networks in the USA is close to 3, but the preferential attachment model is a very poor fit to these data. We conclude that the epidemic thresholds implied by this model exist in both single-sex and two-sex epidemic model formulations. A strong conclusion that we derive from these results is that public health interventions aimed at reducing the transmissibility of STD pathogens, such as implementing condom use or high-activity anti-retroviral therapy, have the potential to bring a population below the epidemic transition, even in populations exhibiting large degrees of behavioural heterogeneity.

Keywords: network scaling; HIV; sexually transmitted disease; epidemiology; Yule distribution

1. INTRODUCTION

The course of an epidemic of an infectious disease is governed by a threshold parameter, R_0, the basic reproductive number (Anderson & May 1991). R_0 is the expected number of secondary cases produced by a single index case in a population of susceptibles. In a stylized formulation, R_0 is a product of the transmissibility of the infectious agent, the duration of the infection and some measure of the contact rate between susceptible and infected individuals. Public health strategies for control and eradication are based on reducing transmissibility, shortening the duration of infection and reducing the contact rate between susceptible and infected individuals. A puzzle in sexually transmitted disease (STD) epidemiology has been how epidemics are maintained given the relatively small number of sexual contacts that people have (relative to the number of contacts for non sexually transmitted infections such as measles or influenza). The answer to this puzzle is that heterogeneity in sexual activity can drive an STD epidemic (Hethcote & Yorke 1984).

In single-sex models with heterogeneous levels of sexual activity, R_0 increases approximately linearly with the variance in the number of sexual partners (Anderson & May 1991). Analogous results have been derived for two-sex models (Newman 2002). Heterogeneity in sexual activity is typically estimated from local network data (Morris 1997) gathered in sexual history surveys. In sexual network analysis, sexual-contact networks are represented as random graphs, where the nodes of the graph represent individual people and the edges represent sexual contact. The number of edges adjacent to a particular node is its degree, and the collection of nodal degrees is the degree distribution of the population (Wasserman & Faust 1994). It is the variance of this degree distribution that plays such an important role in determining the threshold reproduction number for an STD. An understanding of the degree distribution of a sexually active population and of the micro forces that generate this distribution is an important step toward designing public health interventions to eradicate STDs.

Representative surveys of sexual behaviour reveal that the typical person has very few sexual partners in the course of a year (Serwadda *et al.* 1995; Laumann *et al.* 1994; Lewin 1996). Given this observation, concern clearly focuses on the statistical properties of the tails of the degree distribution. Recent work on the properties of human sexual-contact networks has suggested that they are characterized by power-law decay of their tails (Liljeros *et al.* 2001). These networks are described as 'scale free' in the recent network literature, to reflect their extreme skewness. However, the key scientific question that arises in this work is not whether a network is scale free, but whether the network's idealized degree distribution has infinite variance, a phenomenon occurring in a specific range of the scaling exponent ρ of the power law. A distribution characterized by a scaling parameter in this range places significant probability on the occurrence of very large degrees. Consequently, it can be shown that there is no epidemic threshold in a population characterized by an idealized infinite-variance degree distribution

*Author for correspondence (jameshj@stat.washington.edu).

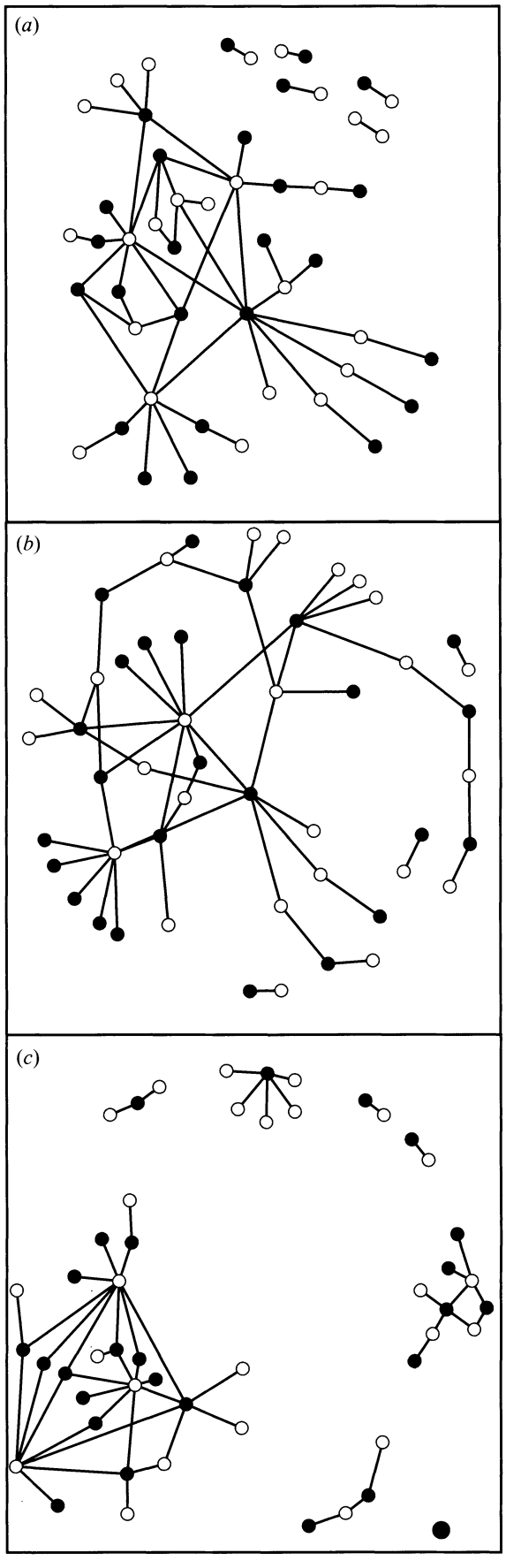

Figure 1. (*a*) Simulated 50-actor predominantly heterosexual network with infinite-variance degree distribution. The underlying distribution is Yule with parameter $\rho = 2.5$. The algorithm producing this network is essentially that of Molloy & Reed (1995). (*b*,*c*) Simulated 50-actor predominantly heterosexual network with infinite-variance degree distribution and a (*b*) maximum or (*c*) minimum propensity to form short paths. The realized degree distribution in (*b*) and (*c*) is identical to that in (*a*). Filled circles, males; open circles, females.

(Lloyd & May 2001; Pastor-Satorras & Vespignani 2001; Newman 2002), allowing a pathogen of arbitrarily small transmissibility to be maintained (Lloyd & May 2001).

The intuition underlying this surprising result is that a network that is simultaneously consistent with (i) the low mean degree characteristic of human sexual behaviour and (ii) the power-law decay of the tail of the degree distribution will exhibit large connected components. Randomly infecting a node in such a network is therefore likely to yield a large epidemic. Figure 1*a* illustrates this idea with a simulated 50-actor (mostly) heterosexual network with an infinite-variance degree distribution (explicitly equation (2.1) with $\rho = 2.5$). The large connected component suggests that the expected size of an epidemic started by randomly infecting a single node would be large.

Modelling a finite population with an infinite-variance degree distribution is clearly an idealization. Note, however, that the probability under such models that the populations considered here contain one or more individuals with more contacts than the population size is infinitesimally small.

In recent work on the scaling of a variety of systems with possible power-law distributions, the scaling exponent has been inferred from the plot of the empirical cumulative distribution against degree (or frequency) on double-logarithmic axes. A theoretical curve is then fitted to the apparently linear region of this empirical plot, either 'by eye' or using a curve-fitting algorithm such as least-squares regression (Axtell 2001). The scaling exponent is estimated from the slope of the line. If least-squares regression is used, the standard error of the slope estimate is used as a measure of the uncertainty of the scaling exponent.

This is a very poor statistical approach to the estimation of the scaling exponent as the assumptions justifying least-squares regression do not hold. First, the empirical values are highly correlated (typical sequential correlations are 0.7 or higher). This is especially true for the values for higher degrees where the sequential correlation approaches unity. The additional information in the latter points is very small, and visual trends are as likely to be caused by the high correlations as to be real. For this reason, considering only the upper tail of the distribution and inferring a pattern is a very dubious practice. Second, the statistical variation in the values is not constant but increases rapidly with the degree (typically by an order of magnitude). This is caused by the logarithmic nature of the plot and the decreasing probabilities. Third, it is usually the procedure to exclude values from the plot that correspond to zero frequencies (e.g. fig. 2 in Liljeros *et al.* 2001). These points contain a great deal of information on the degree distribution, and their exclusion introduces bias into the estimates. Fourth, the high-degree frequencies are sensitive to misreporting and population heterogeneity (e.g. Morris 1993). While these can be adjusted for stat-

istically, the least-squares and regression approaches are overly influenced by them. Finally, accessible statistical methodology, such as the likelihood approach applied here, exists that does not suffer from these defects.

The only empirical estimates of scaling parameters in human sexual networks currently available come from an analysis of sexual history survey data carried out in Sweden, a country with an HIV/AIDS prevalence of less than 1%, and this analysis was subject to the methodological problems described above. A critical test of the adequacy of the current formulation of sexual network scaling models will therefore come from estimating the scaling parameters using robust unbiased methodology in a variety of populations, including some with a clear epidemic. We estimate the scaling parameters of the heterosexual-contact network in three populations: (i) Rakai district, Uganda; (ii) Sweden; and (iii) the USA.

2. MATERIAL AND METHODS

(a) *Data*

We used local network data gathered from men and women as part of three large representative surveys of sexual behaviour. The Rakai district is an administrative unit of southern Uganda with a mature AIDS epidemic and an HIV/AIDS prevalence of *ca.* 16%. The primary mode of HIV transmission in Rakai is believed to be heterosexual. Data were collected as part of the Rakai Project Sexual Network Survey (Serwadda *et al.* 1995). Data for Sweden come from the 1996 'Sex in Sweden' survey based on a nationwide probability sample and financed by the (Swedish) National Board of Health (Lewin 1996). Data for the USA come from the National Health and Social Life Survey (NHSLS) (Laumann *et al.* 1994). Neither Sweden nor the USA is characterized by a generalized HIV/AIDS epidemic, with a national prevalence for both countries of less than 1%. For all surveys, we used the reported number of sexual partners in the last year as the estimate of individual network degree. Sample sizes are given in table 1.

The degree distributions for the three samples are plotted in figure 2. In the recent network scaling literature, it is customary to plot sample degree distributions as (apparently) continuous survival plots on logarithmic axes (Amaral *et al.* 2000; Liljeros *et al.* 2001). That is, the log-probability of degree at least k is plotted against $\log(k)$. Such plots can be visually misleading on several fronts: (i) they suggest a continuous distribution of network degree; (ii) they obscure the fact that there are frequently zero-frequency degrees for $k < k_{\max}$ in the sample; (iii) they do not represent those individuals who have been sexually active but had $k = 0$ for the sample interval; and (iv) they overemphasize log-linearity because of the extreme autocorrelation in the tail. Figure 2 presents the degree distributions as raw frequency histograms, emphasizing their discrete nature and the occurrence of zero-frequency degrees.

(b) *Stochastic model*

The underlying stochastic model motivating the partnership distributions is essentially that of Simon (1955). It is based on two assumptions: (i) a constant probability $(\rho - 2)/(\rho - 1)$ that the $r + 1$th partnership in the population is initiated with a previously sexually inactive person; and (ii) the probability that the $r + 1$th partnership will be with a person with exactly k partners is proportional to $k f(k \mid r)$, where $f(k \mid r)$ is the frequency of nodes with exactly k partnerships out of the r total partnerships in the population. Simon called the limiting partnership distribution of this process the Yule distribution, following the pioneering work of Yule (1924). Recent authors have referred to this process as 'preferential attachment' (Barabási & Albert 1999). The probability mass function (PMF) of the Yule distribution (Johnson *et al.* 1992) is

$$p(K = k) = \frac{(\rho - 1)\Gamma(k)\Gamma(\rho)}{\Gamma(k + \rho)}, \ \rho > 1, \quad \text{for } k = 1, 2, \ldots, \tag{2.1}$$

where $\Gamma(\rho)$ is the gamma function of ρ. The Yule distribution has power-law behaviour in the sense that $p(K = k)/k^{-\rho}$ is approximately constant for large k. The stochastic formulation requires $\rho > 2$, so the mean of the Yule distribution is $(\rho - 1)/(\rho - 2)$. For $\rho \leq 3$ the variance of the Yule distribution is infinite. Any value of ρ greater than 2 corresponds to a preferential attachment model.

(c) *Statistical inference*

Consider fitting a PMF $p_\theta(K = k)$ to survey information where θ is the parameter. For example, for the Yule model the parameter is ρ, the scaling exponent. We adopt a likelihood framework to estimate the model parameters and compare the different models against each other. The likelihood framework provides a set of powerful tools for inference. Given a random sample of n individuals with reported degrees $K_1, \ldots, K_n$, the likelihood of the model is

$$\mathcal{L}(\theta, k_{\min} \mid K_1 = k_1, \ldots, K_n = k_n) \equiv \sum_{i=1}^{n} P_\theta(K = k_i \mid K > k_{\min}) \ k_{\min} = 0, 1, \ldots. \tag{2.2}$$

A maximum-likelihood estimator (MLE) for θ is a value $\hat{\theta}$ that maximizes equation (2.2) as a function of θ. Formulae for the full data likelihoods are given in Handcock & Jones (2002).

Although the statistical properties of the MLE can be analysed asymptotically, we employ bootstrap methods to quantify the small-sample properties of MLEs and to calculate confidence intervals (Efron & Tibshirani 1993).

We adapt the model to allow for the possibility that the tail behaviour (i.e. $k > 1$) of the degree distribution may differ fundamentally from the majority of the observations for which $k = 0$ or 1 (May & Lloyd 2001). We generalize the Yule model to be able to include parameters to fit the probabilities of lower degree (Handcock & Jones 2002). To choose the best-fitting Yule model for the observed data, we employed a Bayesian information criterion (BIC) approach to model selection (Raftery 1995). The BIC represents the integrated likelihood of a model and takes into account both the number of parameters a model uses and the sample size. Given a random sample of size n, $(K_1, \ldots, K_n)$, the BIC is given by

$$\mathrm{BIC} = -2\mathcal{L}_f(\hat{\theta}, k_{\min} \mid K_1, \ldots, K_n) + \log(n)(d + k_{\min} + 1),$$

where d is the dimension of θ, and $\mathcal{L}_f$ is the generalization of $\mathcal{L}$ to include $k = 0, 1, \ldots, k_{\min}$.

3. RESULTS

The results of the Yule model fits are given in table 1. For all populations apart from Rakai women, the best-fitting model fitted the proportions with degree zero and one separately.

Table 1. Results of statistical inference for Yule model. Estimates of ρ are given with 95% bootstrap confidence intervals. $k_{\min}$ is the lowest degree to which the parametric model is fit for all $k < k_{\min}$.

country	sex	n	$k_{\min}$	BIC	ρ (95% CI)
Uganda	women	803	0	1070.45	17.04 (12.58, 25.19)
	men	621	1	1587.79	5.43 (4.32, 6.53)
Sweden	women	1335	1	2158.64	4.23 (3.60, 5.21)
	men	1476	1	3041.55	3.25 (3.01, 3.63)
USA	women	1919	1	3224.03	3.84 (3.34, 4.55)
	men	1506	1	3267.56	3.03 (2.80, 3.32)

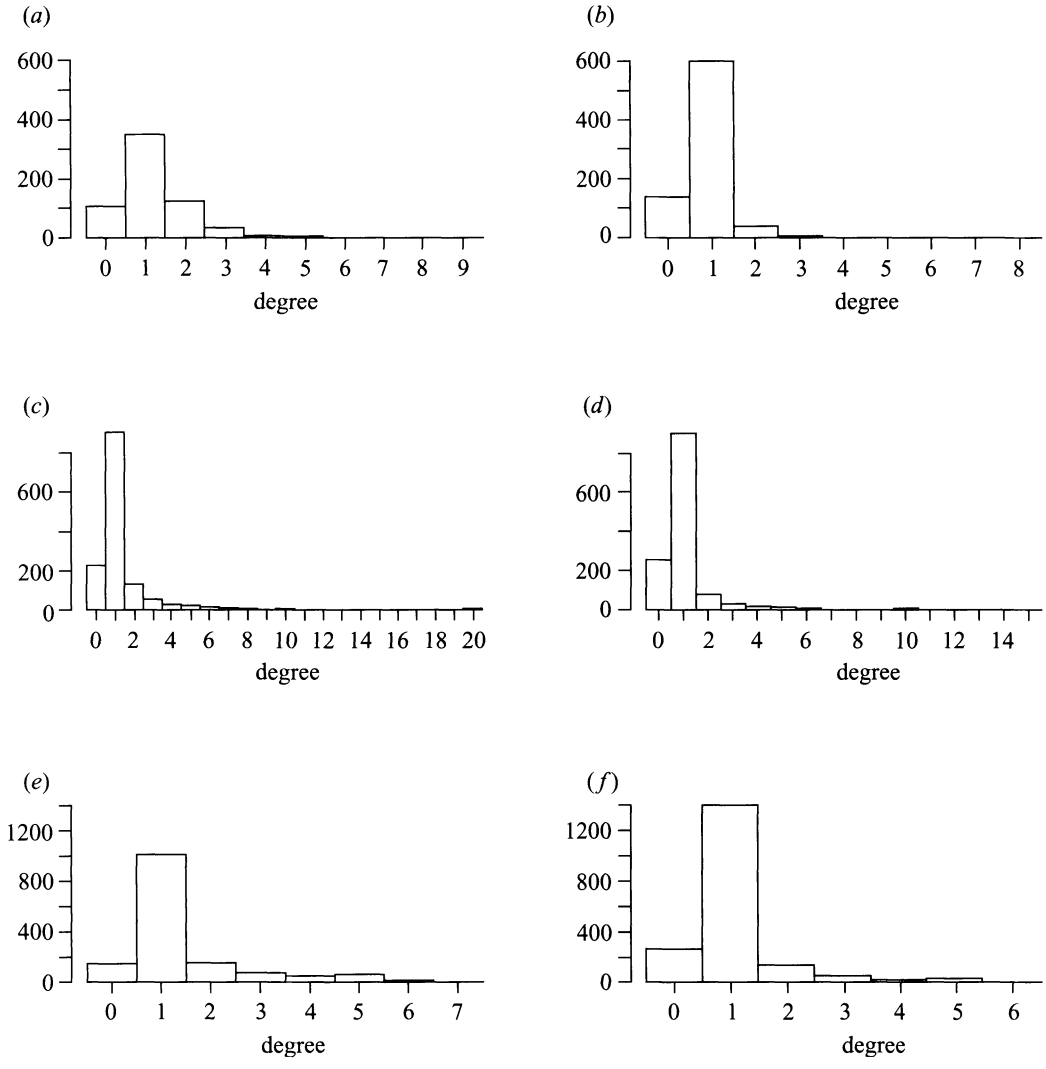

Figure 2. Reported degree distributions for (*a*,*c*,*e*) men and (*b*,*d*,*f*) women from the three population samples (*a*,*b*: Uganda; *c*,*d*: Sweden; *e*,*f*: USA). The plots are histograms showing the absolute number of observed degree k (including zeros).

For all models but one, the interval estimate of the scaling parameter falls above the range in which the Yule distribution has infinite variance (i.e. $\rho > 3$). The 95% confidence interval for ρ for men from the USA NHSLS sample includes values within the infinite-variance region.

Elsewhere, we have shown the effect of conditioning on higher degree on the confidence intervals of the scaling parameter estimates for Swedish males (Handcock & Jones 2002). However, it is worth noting here that, in addition to reducing the 'goodness-of-fit' substantially and increasing the BIC, estimates based on high $k_{\min}$ (e.g. 4 or 5 as in Liljeros *et al.* (2001)) yield wildly increasing confidence intervals.

4. DISCUSSION

Using methods appropriate to the inference problem, we have estimated the scaling parameter of the Yule distribution, the limiting distribution for the preferential attachment process, for local sexual network data from three large datasets. The scaling parameter estimates indicate that the variances of the idealized degree distributions for

both sexes are finite in two out of the three populations, with the plausibility that $\rho < 3$ only for American men.

The estimate of the Yule scaling parameter for American men was 3.03 and the confidence interval overlaps the region of infinite variance. However, the Yule distribution was not the globally best-fitting model for the American data. In a separate paper (Handcock & Jones 2002), we have developed a variety of stochastic models for sexual network growth and estimated the models using the same data analysed here. The best-fitting model for American men does not have a power-law tail, and therefore, has finite variance.

The predictions of the model depend on the form of the population degree distribution. The intuition underlying power-law scaling models is that the tails of the degree distributions in human sexual networks are long and decrease relatively slowly. However, the extremely high values of the scaling exponents of the Yule model for most of the local networks indicate that the observed degree, in fact, falls off rapidly within the range of the data. Models with power-law tails fit the observed data because of the essentially L-shaped nature of degree distributions, where the great majority of people have low degree and a very small fraction have high degree. This observation suggests that a unitary behavioural process, such as preferential attachment, is unlikely to underlie empirical sexual network degree distributions.

This, of course, does not preclude more complex forms of preferential-attachment-like processes from contributing to the formation of sexual networks. One model that has been suggested is the so-called truncated power law distribution, which displays power-law behaviour below some characteristic ceiling κ (Newman 2002). In a separate work (Handcock & Jones 2002), we compare a variety of models motivated by different stochastic processes using likelihood-based techniques for multi-model inference (Burnham & Anderson 2002).

While the language of recent work may be novel in epidemiology, the interventions suggested by the putative power-law behaviour of sexual networks are not particularly radical, as has been suggested (Liljeros *et al.* 2001; Dezsö & Barabási 2002). Behavioural heterogeneity was recognized as an important contributor at an early stage in the HIV/AIDS epidemic (Anderson *et al.* 1986) and degree-based interventions were proposed (Woolhouse *et al.* 1997). Targeting at-risk populations such as commercial sex workers (Ford & Koetsawang 1999), truck drivers (Morris *et al.* 2000), army recruits (Nelson *et al.* 2002) and injection drug users (Neaigus 1999) has a proven record in reducing disease incidence.

Our results suggest that efforts to reduce pathogen transmissibility are not wasted. A sexual network with finite variance will have an epidemic threshold for positive transmissibility. Indeed, public health efforts aimed at reducing the transmissibility of HIV have met with great success. Recently, Velesco-Hernandez *et al.* (2002) have argued that the use of high-activity anti-retroviral therapy and other public health interventions in San Francisco have brought the R_0 value for HIV in homosexual men below the threshold and, all things being equal, a slow endemic fade-out can be expected. Thailand's 100% condom use intervention for commercial sex workers and army recruits has been a spectacular success in curbing an incipient generalized AIDS epidemic (Ford & Koetsawang 1999; Nelson *et al.* 2002).

Much of the recent interest in the scaling of networks has focused exclusively on the behaviour of the degree distribution (Pastor-Satorras & Vespignani 2001; Liljeros *et al.* 2001; Dezsö & Barabási 2002; Newman 2002), and some of this work proposes policy recommendations based on the inferred properties of the degree distribution (Liljeros *et al.* 2001; Dezsö & Barabási 2002). However, there are other features of networks that could have a substantial impact on epidemic processes. Two structural properties of networks that have received some attention are concurrency and local clustering. Morris & Kretzschmar (1995, 1997) have documented the impact of concurrency in sexual networks on the speed and final size of epidemics. Networks characterized by moderate amounts of concurrency (holding degree distribution constant) produce larger epidemics faster. Watts (1999) has popularized the concept of 'small world' networks, namely, those networks with high clustering and short minimum path length (relative to the Bernoulli graph). The joint effect of high clustering and short path length means that an epidemic could spread rapidly through a small-world network. Amaral *et al.* (2000) note that power-law networks can be small world networks, but power-law scaling of the degree distribution is not a necessary condition for the small world phenomenon.

The limitations of the exclusively degree-based perspective of Liljeros *et al.* (2001) or Dezsö & Barabási (2002), for example, are highlighted by the fact that infinite-variance networks can have dramatically different structures depending on the values of other network parameters, and that these different structures are expected to produce qualitatively different epidemic behaviour. In figure 1*b*,*c*, we present simulated networks with the same degree distributions as that of figure 1*a*. However, in both these networks, we further specified the propensity to form short paths between actors, a measure of clustering in networks (Wasserman & Faust 1994). Networks were simulated conditional upon the degree distribution used in figure 1*a* using a Markov Chain Monte Carlo algorithm (Handcock 2003). These are based on an exponential random-graph model for the network structure, and not simply on the marginal degree distribution (Frank & Strauss 1986). Figure 1*b* shows an infinite-variance network with a high propensity for forming short paths, whereas figure 1*c* presents an infinite-variance network with a low propensity. It seems highly likely that the epidemic behaviours on these networks, nonetheless characterized by the same infinite-variance degree distribution, would be qualitatively different. The network with low propensity for short paths yields isolated cliques of high connectivity, in contrast to the connected giant component of the high-propensity network.

The simple model of Newman (2002), assuming a power-law degree distribution and random mixing conditional on the degree distribution, predicts that, for some of the observed values of ρ, an epidemic would be impossible. Possible explanations for this potentially counter-intuitive result include: (i) a power-law does not actually describe the degree distribution; (ii) mixing is not random, but is a function of actor attributes; (iii) geographical- or social-locality effects segment the network; and (iv) there are substantial changes in network structure over time.

These observations highlight the need to exercise caution in developing public health policy from information on the degree distribution alone (Liljeros *et al.* 2001; Dezsö & Barabási 2002), regardless of the inferential procedures employed to characterize the network.

The analysis we have provided here indicates that interventions aimed at reducing transmissibility still have the potential to eradicate STDs. Though sexual degree distributions may have long tails, the models analysed here are characterized by finite variance. Both degree-based and transmissibility-reducing interventions have the possibility of lowering the reproductive rate of STD agents below the epidemic threshold and should continue to be pursued in the quest for STD elimination.

The authors gratefully acknowledge the critical feedback and support they have received from Martina Morris, King Holmes, Julian Besag, Adrian Raftery, Steve Goodreau, Mark Newman, Richard Hayes, Roy Anderson, Bob May and two anonymous reviewers. Frederik Liljeros generously made available the Swedish data used in Liljeros *et al.* (2001). The authors especially thank Dr Bo Lewin, Professor of Sociology, Uppsala University and head of the research team responsible for the 'Sex in Sweden' study, for providing the data used in this study. This research was supported by grant 7R29HD034957 from NICHD and grant 7R01DA012831 from NIDA.

REFERENCES

Amaral, L. A. N., Scala, A., Barthelemy, M. & Stanley, H. E. 2000 Classes of small-world networks. *Proc. Natl Acad. Sci. USA* **97**, 11 149–11 152.

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control.* Oxford University Press.

Anderson, R. M., Medley, G., May, R. M. & Johnson, A. 1986 A preliminary study of the transmission dynamics of the Human Immunodeficiency Virus (HIV), the causative agent of AIDS. *IMA J. Math. Appl. Med. Biol.* **3**, 229–263.

Axtell, R. L. 2001 Zipf distribution of US firm sizes. *Science* **293**, 1818–1820.

Barabási, A. L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512.

Burnham, K. & Anderson, D. 2002 *Model selection and inference: a practical information-theoretic approach*, 2nd edn. New York: Springer.

Dezsö, Z. & Barabási, A. L. 2002 Halting viruses in scale-free networks. *Phys. Rev.* E **65**. (DOI 10.1103/PhysRevE.65.055103.)

Efron, B. & Tibshirani, R. J. 1993 *An introduction to the bootstrap*. New York: Chapman & Hall.

Ford, N. & Koetsawang, S. 1999 A pragmatic intervention to promote condom use by female sex workers in Thailand. *Bull. WHO* 77, 888–894.

Frank, O. & Strauss, D. 1986 Markov graphs. *J. Am. Stat. Assoc.* **81**, 832–842.

Handcock, M. S. 2003 *Statistical models for social networks: inference and degeneracy in social network modelling and analysis: workshop summary and papers*. National Academy of Sciences Press. (Submitted.)

Handcock, M. S. & Jones, J. H. 2002 Likelihood-based inference for stochastic models of sexual network formation. Working paper #29. Center for Statistics and the Social Sciences, University of Washington.

Hethcote, H. W. & Yorke, J. A. 1984 *Gonorrhea: transmission dynamics and control.* Springer Lecture Notes in Biomathematics. Berlin, Springer.

Johnson, N., Kotz, S. & Kemp, A. 1992 *Univariate discrete distributions*, 2nd edn. New York: Wiley.

Laumann, E., Gagnon, J., Michael, T. & Michaels, S. 1994 *The social organization of sexuality: sexual practices in the United States.* University of Chicago Press.

Lewin, B. (ed.) 1996 *Sex in Sweden.* Stockholm: National Institute of Public Health.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Åberg, Y. 2001 The web of human sexual contacts. *Nature* **411**, 907–908.

Lloyd, A. L. & May, R. M. 2001 Epidemiology—how viruses spread among computers and people. *Science* **292**, 1316–1317.

May, R. M. & Lloyd, A. L. 2001 Infection dynamics on scale-free networks. *Phys. Rev.* E **64**. (DOI 10.1103/PhysRevE.64.066112.)

Molloy, M. & Reed, B. 1995 A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6**, 161–179.

Morris, M. 1993 Telling tails explain the discrepancy in sexual partner reports. *Nature* **365**, 437–440.

Morris, M. 1997 Sexual networks and HIV. *AIDS* **11**, 5209–5216.

Morris, M. & Kretzschmar, M. 1995 Concurrent partnerships and transmission dynamics in networks. *Social Networks* **17**, 299–318.

Morris, M. & Kretzschmar, M. 1997 Concurrent partnerships and the spread of HIV. *AIDS* **11**, 641–648.

Morris, M., Wawer, M. J., Makumbi, F., Zavisca, J. R. & Sewankambo, N. 2000 Condom acceptance is higher among travelers in Uganda. *AIDS* **14**, 733–741.

Neaigus, A. 1999 The network approach and interventions to prevent HIV among injection drug users. *Public Health* **113**, 140–150.

Nelson, K. E., Eiumtrakul, S., Celentano, D. D., Beyrer, C., Galai, N., Kawichai, S. & Khamboonruang, C. 2002 HIV infection in young men in northern Thailand, 1991–1998: increasing role of injection drug use. *J. Acquir. Immune Defic. Syndr.* **29**, 62–68.

Newman, M. 2002 Spread of epidemic disease on networks. *Phys. Rev.* E **66**. (DOI 10.1103/PhysRevE.66.016128.)

Pastor-Satorras, R. & Vespignani, A. 2001 Immunization of complex networks. *Phys. Rev. Lett* **86**, 3200–3203.

Raftery, A. E. 1995 Bayesian model selection in social research. *Sociol. Methodol.* **25**, 111–163.

Serwadda, D., Gray, R. H., Wawer, M. J., Stallings, R. Y., Sewankambo, N. K., Kondelule, J. K., Lainjo, B. & Kelly, R. 1995 The social dynamics of HIV transmission as reflected through discordant couples in rural Uganda. *AIDS* **9**, 745–750.

Simon, H. 1955 On a class of skew distribution functions. *Biometrika* **42**, 435–440.

Velesco-Hernandez, J., Gershengorn, H. & Blower, S. 2002 Could widespread use of combination antiretroviral therapy eradicate HIV epidemics? *Lancet Infect. Dis.* **2**, 487–493.

Wasserman, S. & Faust, K. 1994 *Social network analysis: methods and applications.* Cambridge University Press.

Watts, D. J. 1999 *Small worlds: the dynamics of networks between order and randomness.* Princeton University Press.

Woolhouse, M. E. J. (and 12 others) 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342.

Yule, G. 1924 A mathematical theory of evolution based on the conclusions of Dr J. C. Willis FRS. *Phil. Trans. R. Soc. Lond.* B **213**, 21–87.