Extended Abstract: Exponential Random Graph Models for Sexual Networks on Likoma Island, Malawi: Implications for Sexual Behavior and HIV Control

James Holland Jones [*]
Department of Anthropological Sciences
Stanford University

Stephane Helleringer
Department of Sociology
University of Pennsylvania

Hans-Peter Kohler
Population Studies Center
University of Pennsylvania

September 23, 2006

1 Introduction

Sexual networks are the primary mechanism through which HIV is spread and transformed in Sub-Saharan Africa (SSA). Theoretical network models have shown that individuals' positions within these sexual networks, and the structural characteristics of the network itself, are important determinants of HIV infection risks and disease dynamics (Kretzschmar and Morris 1996; Ghani and Garnett 2000; Newman 2002). Several features of sexual networks that are predicted by these models to enhance the spread of HIV have been empirically documented in SSA including *concurrency* of sexual partnerships (Morris 1997), *skewed degree distributions* of sexual networks (Anderson and May 1991; Jones and Handcock 2003b), and *large and robust connected components* (Moody et al. 2003; Moody and White 2003).

Despite the clear theoretical significance of structural features of risk in HIV transmission, very little work has been done on characterizing the structure of sexual networks in SSA. This observation is largely attributable to the extremely demanding requirements of network data collection. Using sociocentric sexual network data from Likoma Island, Malawi, we have a rare opportunity to understand the structural features of the risk structure of a real sexual network in an area characterized by a generalized AIDS epidemic.

---

[*]Correspondence Address: Department of Anthropological Sciences, Building 360, Stanford, CA 94305-2117; phone: 650-723-4824, fax: 650-725-9996; email: `jhj1@stanford.edu`

1

1.1 Sexual Networks Violate the Assumptions of Standard Epidemiological Models

All measures to control an epidemic stem from the requirement to bring the basic reproduction number below the epidemic threshold of $\mathcal{R}_0 = 1$. The basic reproduction number, $\mathcal{R}_0$, is the expected number of secondary infections arising from a single, typical infectious individual in a completely susceptible population (Heesterbeek 2002). The most efficient route for achieving this end can be ascertained through analysis of the the transmission model of the infectious process.

The classical models of mathematical epidemiology (e.g., Bailey 1975; Anderson and May 1991) rely on the assumption that sexual partners are randomly selected (i.e., the population is assumed to be well-mixed and unstructured). In this model, two key measures to study epidemics are (1) the basic reproduction number, $\mathcal{R}_0$, and (2) the final size of an epidemic s_∞. In a well-mixed and socially unstructured population, $\mathcal{R}_0$ is the product of three quantities: the transmissibility τ, the rate of contact between susceptible and infectious individuals $\bar{c}$ and the duration of infectiousness δ. Epidemics are nonlinear phenomena and $\mathcal{R}_0$ is a threshold parameter. When $\mathcal{R}_0 > 1$, an epidemic is certain in a deterministic model and has non-zero probability in a stochastic model. Strategies for disease control and eradication are aimed at bringing $\mathcal{R}_0$ below the threshold of unity, i.e, when the average infection generates fewer secondary infections than necessary for replacement and the epidemic fades. In the well-mixed and unstructured case, the final size of the epidemic is given by the implicit equation $\log(s_\infty) = \mathcal{R}_0(s_\infty - 1)$, which has exactly two roots on the interval $[0\ 1]$ when $\mathcal{R}_0 > 1$. The smaller of these roots is the proportion of the population remaining uninfected at the end of an epidemic.

However, because HIV is transmitted by intimate sexual contacts between partners, and because people employ varied and elaborate rules to choose their partners, HIV transmission dynamics in real populations are not well described by the classical epidemiological model. For instance, while African men (and to a lesser extent women) do not report having more sexual partners than men elsewhere, they tend to have more than one on-going long-term relation at any point in time. Halperin and Epstein's summary of research findings shows that close to 20% of males in Eastern African countries report concurrent partnerships, versus less than 5% almost everywhere outside of Africa (Halperin and Epstein 2004). Partnerships in SSA can can overlap for months, maybe years (Lagarde et al. 2001; Morris and Kretzschmar 2000). This pattern of sexual partnerships that overlap rather than follow each other sequentially (Kretzschmar and Morris 1996; Moody 2002), is one of several important characteristics of human sexual networks that violate the classical epidemiological model and importantly affect HIV infection risks and disease dynamics. *Concurrent partnerships* appear to increase the speed at which HIV spreads through a population, and have probably contributed to the rapid take-off of the HIV epidemic in SSA in the 1980s (Morris and Kretzschmar 2000). Other violations of the classical epidemiological model include: (1) *assortative mixing*, i.e., the selection of sexual partners based on their individual characteristics, can structure a network into communities within which the disease spreads rapidly, but across which the spread is slow (Laumann et al. 1994; Laumann and Youm 1999; Morris 1993); (2) *small worlds*, i.e., networks characterized by bridges joining otherwise disjoint clusters (Watts and Strogatz 1998; Watts 1999), can lead to thresholds and rapid disease diffusion to distant subpopulations; (3)*robust networks*, i.e., groups of persons

tied together by more than one path in the sexual network, can decrease the ability to control the spread of HIV because redundant connections continue to transmit HIV even after some transmission paths are broken or eliminated (Moody et al. 2003; Potterat et al. 2002); (4) *skewed degree distributions*, i.e., networks containing individuals with very a high number of partners (high degree network members), can result in epidemics driven by promiscuous individuals (e.g., Liljeros et al. 2001; for a critical perspective, see Jones and Handcock 2003a; Handcock and Jones 2004).

Because of the strong violations of the basic epidemic model inherent in sexual transmission of HIV, we currently lack the ability to reliably estimate $\mathcal{R}_0$, determine the optimal interventions for bringing $\mathcal{R}_0$ below threshold, or understanding how many people will ultimately be infected during the course of the HIV epidemic. Using the detailed information on the sexual networks of Likoma Island, we undertake to remedy this shortcoming.

1.2 Data

Ongoing work conducted in Likoma island (a small island with high HIV prevalence located in the northern region of Lake Malawi) allowed us to construct detailed images of sociocentric sexual networks. Complete sexual histories were obtained for the last five sexual partners in a census of adults in 7 villages of the island. Using a combination of name generators, attribute matching, and GIS-assisted spatial matching, we were able to construct a sociomatrix of sexual partnerships from the egocentric data generated by the sexual history surveys.

We extracted data on the sexual relations having taken place over the last 12 months. The sexual network derived from these relationships is composed of 1463 adults. Of these, 1085 reside in Likoma and 378 reside on the mainland, either in Malawi, Mozambique or elsewhere. Table 1 presents the distribution of components.

Component Size	2	3	4	5	6	7	8	9	12	13	16	18	21	600
Count	183	63	20	10	7	2	1	2	1	1	2	1	1	1

Table 1: Empirical component size distribution of the one-year sexual network.

The one-year sexual network is extremely sparse with 1208 ties and a corresponding tie density of 0.0011. Figure 3 plots the one-year network. The large connected component of 600 individuals is clearly visible in the center of the plot. This large component clearly represents a major structural risk for HIV transmission and a key goal of statistical modeling is understanding what forces act to form it. If we understand the behavioral mechanisms through which such large connected components are formed, we can design interventions to address the risk.

Two network statistics are of particular interest for understanding the formation of large connected components in sexual networks: tie density and dyad-wise shared partnerships (DSPs). These network features have been hypothesized to play a strong role in the diffusion of infections over networks (Morris 2003). Tie density represents the level of sexual activity (i.e., the number of partners per actor), whereas DSPs represent the degree of network transitivity. Well understood properties of random graphs lead to the formation of giant components in random networks at threshold tie densities. However, comparison of random networks with empirically

network stat	estimate	s.e.	p-value	MCMC s.e.
edges	-5.6750	1.5980	0.000383	0.38233
dsp1	-0.5186	0.1378	0.000168	0.02598
dsp2	0.5861	0.4147	0.157557	0.07793
nodematch.sex	-4.2113	1.5702	0.007317	0.35206

Table 2: Results of model fit using edges, dsp(1:2), and sex-homophily. Edges, dsp(1), and sex-homophily are highly significant.

measured networks indicate substantial departures from the simple random structure. Forces leading to the formation of DSPs are potentially responsible for such departures. Figure 1 represents graphically a dsp(1) and dsp(2) in a heterosexual network. Transitivity is a fundamental property of disease transmission networks as it describes the propensity of any two individuals in the networks to share partners. Moderate to high degrees of transitivity will lead to clustering (Jones and Handcock 2003b) and localization effects in which local prevalence is greatly amplified but transmission out of the local cluster is low. Such processes have the potential to explain some of the heterogeneity in epidemic outcomes across localities.

1.3   Statistical Estimation

To investigate these processes of clustering, we will fit exponential random graph (ERGM) models using STATNET (Handcock et al. 2003), statistical software developed by Mark Handcock and colleagues for this very purpose. STATNET fits models of the form

$$P_\eta(Y = y|X) = c^{-1} \exp\{\eta^\mathsf{T} g(y, X)\},$$

where Y is a sociomatrix and $P(Y = y|X)$ indicates the probability of a particular tie in sociomatrix Y conditional on X, a matrix of covariates, η is a q-vector of coefficients, $g(y, X)$ is a q-vector of network statistics, c is a normalizing constant $c = c(\eta) = \sum_{y \in \mathcal{Y}} \exp\{\eta^\mathsf{T} g(y, X)\}$, and $\mathcal{Y}$ is the sample space of possible sociomatrices. This space is astronomically large for even modestly sized networks. This fact has been the major stumbling block for likelihood-based inference for social networks. However, recent developments in statistical computing now allow approximate maximum likelihood estimates of the coefficients η to be calculated using Markov Chain Monte Carlo (MCMC) simulation (Handcock et al. 2003).

1.4   Preliminary Results

Three parameters were highly statistically significant from the model fit. Edges showed a log odds of -5.6750, reflecting the low density of ties in the model. Dyad-wise shared partnerships also showed a negative coefficient, indicating that a tie added to the Likoma network is less likely to result in a DSP than in a random graph. Sex-homophily also had a highly negative log-odds, reflecting the fact that this is a heterosexual network. Neither sex nor age homophily had significant effects on the probability of observing a tie in the one-year network.

References

Anderson, R. M. and R. M. May (1991). *Infectious diseases of humans: Dynamics and control.* Oxford: Oxford University Press.

Bailey, N. (1975). *The mathematical theory of infectious disease.* New York: Hafner Press.

Ghani, A. C. and G. P. Garnett (2000). Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. *Sexually Transmitted Diseases 27*, 579–587.

Halperin, D. and H. Epstein (2004). Concurrent sexual partnerships help to explain Africa's high HIV prevalence: implications for preventio. *Lancet 363*, 4–6.

Handcock, M. S., D. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris (2003). `statnet`: An R package for the statistical modeling of social networks. funding support from NIH grants R01DA012831 and R01HD041877.

Handcock, M. S. and J. Jones (2004). Likelihood-based inference for stochastic models of sexual network evolution. *Theoretical Population Biology 65*, 413–422.

Heesterbeek, J. A. P. (2002). A brief history of r_0 and a recipe for its calculation. *Acta Biotheoretica 50*(3), 189–204.

Jones, J. and M. S. Handcock (2003a). Sexual contacts and epidemic thresholds. *Nature 425*, 605–606.

Jones, J. H. and M. S. Handcock (2003b). An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society of London Series B-Biological Sciences 270*, 1123–1128.

Kretzschmar, M. and M. Morris (1996). Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences 133*(2), 165–195.

Lagarde, E., B. Auvert, M. Carael, M. Laourou, B. Ferry, E. Akam, T. Sukwa, L. Morison, B. Maury, J. Chege, I. N'Doye, and A. Buve (2001). Concurrent sexual partnerships and HIV prevalence in five urban communities of sub-saharan Africa. *AIDS 15*, 877–884.

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

Laumann, E. and Y. Youm (1999). Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. *Sexually Transmitted Diseases 26*(5), 250–261.

Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg (2001). The web of human sexual contacts. *Nature 411*(6840), 907–908.

Moody, J. (2002). The importance of relationship timing for diffusion. *Social Forces 81*, 25–46.

Moody, J., M. Morris, J. Adams, and M. Handcock (2003). Epidemic potential in human sexual networks. Unpublished working paper.

Moody, J. and D. White (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review 68*, 103–127.

Morris, M. (1993). Epidemiology and social networks: Modeling structured diffusion. *Sociological Methods & Research 22*(1), 99–126.

Morris, M. (1997). Sexual networks and HIV. *AIDS 11*, S209–S216.

Morris, M. (2003). Local rules and global propertis: Modeling the emergence of network structure. In R. L. Breiger, K. M. Carley, and P. Pattison (Eds.), *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, pp. 174–186. Washington, DC: National Academy Press.

Morris, M. and M. Kretzschmar (2000). A microsimulation study of the effect of concurrent partnerships on the spread of HIV in Uganda. *Mathematical Population Studies 8*(2), 109–133.

Newman, M. E. J. (2002). Spread of epidemic disease on networks. *Physical Review E 66*(1), art. no.–016128.

Potterat, J., L. Phillips-Plummer, S. Muth, R. Rothenberg, D. Woodhouse, T. Maldonado-Long, H. Zimmerman, and J. Muth (2002). Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sexually Transmitted Infections 78*(Supplement 1), 159–163.

Watts, D. and S. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature 393*, 440–443.

Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton: Princeton Studies in Complexity.
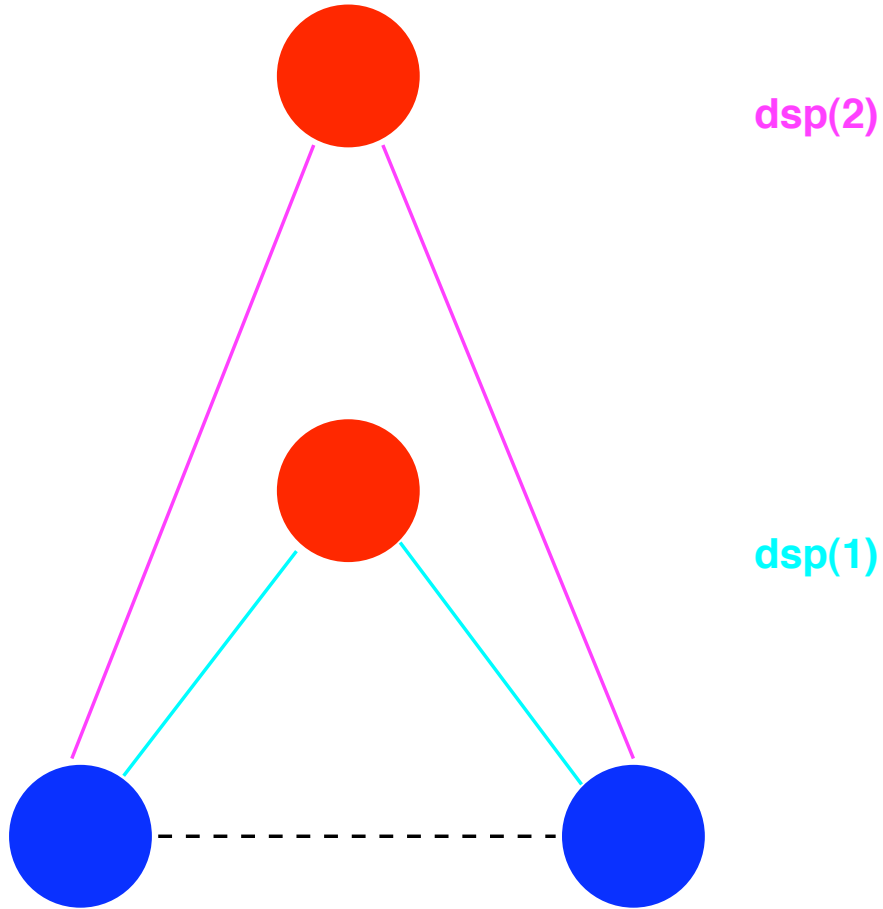
Figure 1: Dyad-wise shared partnerships. The dyad is the two blue nodes and their shared partners are the red nodes, with the relationships defining dsp(1) in turquoise and the dsp(2) relationships in magenta.
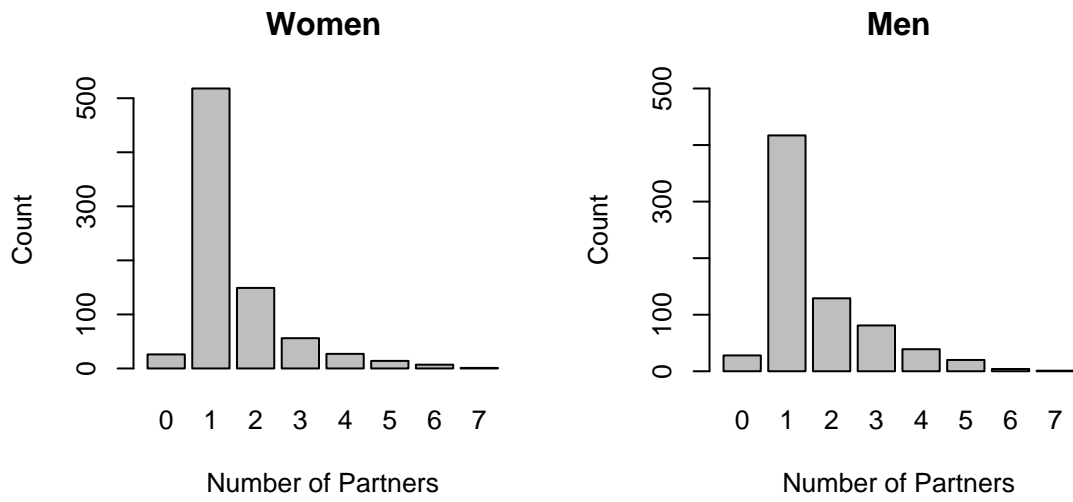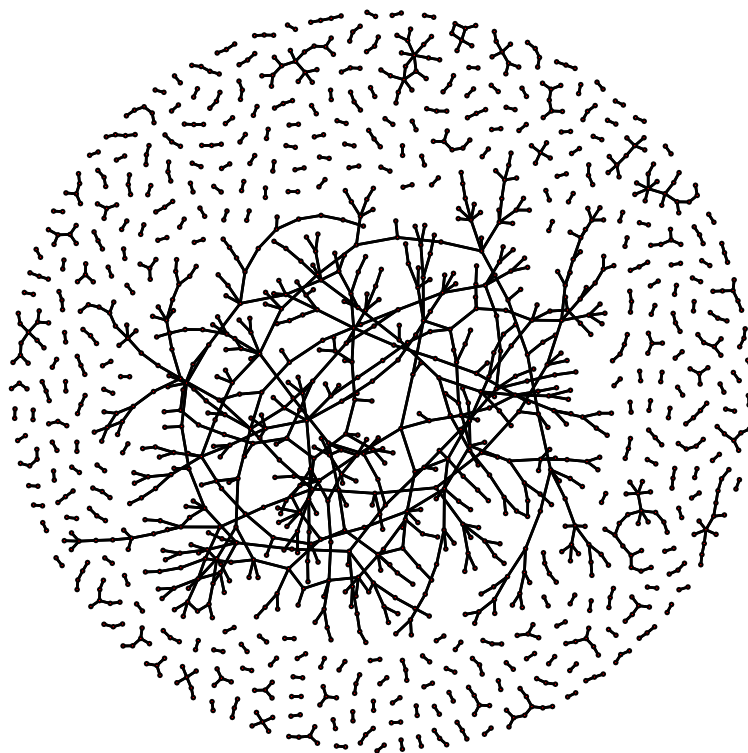
Figure 2: Female and male degree distributions.

Figure 3: One-year sexual network for Likoma Island.