# Pathways of Infection: Sexual Networks and HIV/AIDS in a sub-Saharan Population

Hans-Peter Kohler\*

Stéphane Helleringer<sup>†</sup>

Valerio Bacak<sup>‡</sup>

September 17, 2010

#### Abstract

The sexual networks connecting members of a population have important consequences for the spread of sexually transmitted diseases including HIV. However, very few datasets currently exist that allow an investigation of the structure of sexual networks, particularly in sub-Saharan Africa where HIV epidemics have become generalized. Using the longitudinal sociocentric sexual network data available in the Likoma Network Study (LNS), this paper can provide new and important insights into the structure of sexual networks in sub-Saharan populations, the relationship between the size and structure of sexual networks and HIV infection risks, and for the first time in a sub-Saharan population, the evolution of the population-level sexual network over time, including the differential changes in network position and structure by LNS Round 1 HIV status, marital/socioeconomic status, and Round 1 network position.

## 1 Introduction

A growing conceptual and empirical attention to social networks and health in recent years has provided a large body of evidence suggesting that, in context where individuals are connected through social relationships, so is their health. A recent book *Connected: The Surprising Power of Our Social Networks and How They Shape our Lives* (Christakis and Fowler 2009), for example, claims that "how we feel, whom we marry, whether we fall ill, and how much money we make and whether we vote—everything hinges on what others around us are doing, thinking and feeling." While this specific claim is rather bold, there is indeed growing evidence that a broad range of health behaviors are associated with, and sometimes even causally affected by, a person's social networks. Health outcomes related to social network structures for instance include smoking, alcohol and substance use, obesity, sexual behaviors and HIV risks, mental health, contraceptive use, risk perceptions, and subjective well-being (e.g. Andrews et al. 2002; Bearman et al. 2004; Behrman et al. 2002; Christakis and Fowler 2007; Fiori et al. 2006; Fowler and Christakis 2008; Helleringer and Kohler 2007; Kaplan et al. 2001; Kohler 1997, 2001; Kohler et al. 2007). Even if many existing studies of the relationship between social networks and health cannot establish causal effects because of the endogeneity and self-selection of networks (for a discussion of these issues,

<sup>\*</sup>Frederick J. Warren Professor of Demography, Department of Sociology, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104; *Email:* hpkohler@pop.upenn.edu.

<sup>&</sup>lt;sup>+</sup>Assistant Professor, Heilbrunn Department of Population and Family Health, Mailman School of Public Health, Columbia University, 60 Haven Avenue, New York, NY 10032; *Email: sh2813@columbia.edu* 

<sup>&</sup>lt;sup>‡</sup>Graduate Student, Department of Sociology, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104; *Email:* vbacak@sas.upenn.edu.

see for instance Behrman et al. 2009), a selected subset aims to do so (e.g., Behrman et al. 2002; Kohler et al. 2007), and there is a growing consensus that social networks are strongly related to health outcomes through a variety of mechanisms, including improved access to health information, better social support, the diffusion of information, peer pressure and the enforcement of behavioral norms (for a review, see Smith and Christakis 2008).

An important limitation of this literature on social networks and health, however, is the use of egocentric (or local) network data in the majority of studies on this topic. These egocentric data provide information about the social ties of a survey respondent, but contain no information about the larger network in which the respondent is embedded. In addition to concerns about establishing causal relationships between networks and health that have been discussed extensively in the literature (Behrman et al. 2009; Manski 2000; Soetevent 2006), therefore, these egocentric network studies are limited because they provide only a very restricted view of a person's social capital and do not allow analyses of how global/local network structures and a person's structural position within a larger community-level social network affect important health outcomes (e.g., Smith and Christakis 2008).

Recognizing the limitations of egocentric network data, a small but growing number of studies has begun to collect sociocentric network data (a.k.a. sociometric, complete, or global networks), in which all or nearly all members of a community or group and their linkages to each other are represented as part of saturation samples. While egocentric data include only the direct links from the focal individuals (the "egos") to other persons, sociocentric networks include both direct and indirect ties and allow mapping community-level network of social relationships. Sociocentric data are less affected by measurement error because all social relationships, and the interactions among network partners, are potentially reported by each of the two members of the relationship. In addition, utilizing the recent advances in network theory (Carrington et al. 2005; Morris 2004), sociocentric network studies—especially when they are longitudinal—can identify the complex patterns of interrelations among persons and their health. Several path-breaking results, for example, have been obtained from analysis of the sociocentric AddHealth network data, including that obese adolescents are often socially isolated within the network among high school youth (Strauss and Pollack 2003), and Bearman et al.'s (2004) finding that reported romantic partnerships over an 18-month period in a Midwestern high school connecting 52% of all romantically involved students were embedded in one very large "spanning tree".

In a pioneering study of networks and health in a poor developing country, Helleringer and Kohler created the first complete sexual network data for a large population on Likoma Island, Lake Malawi. Combining a census of the population with a sociocentric sexual network survey of *all* adults aged 18–35, Helleringer and Kohler (2007) document the existence of a large and robust sexual network (Figure 1). Half of all sexually active respondents were linked together in a *giant network component*, and more than one quarter were connected together through multiple independent chains of sexual relations. Such structural features of sexual networks have been associated with epidemic spread of STIs in high-risk groups, but prior to this study had never been documented among the general population. This unique design of the *Likoma Network Study* (NLS) has provided important new findings on the role of concurrency of sexual partners

and HIV risk, the role of migrants and the contribution to HIV risks within rural populations, the uneven distribution of HIV risks within large sexual networks, data quality and misreporting of sexual relationships, and the determinants of sexual relationships and patterns of homophily and clustering within sexual networks (Clark et al. 2008; Helleringer and Kohler 2008; Helleringer et al. 2007, 2009a,b,c,d).

An *limitation* of the previous analyses of the Likoma Network Study, and any similar analyses conducted in sub-Saharan Africa, has been there cross-sectional nature. The key innovation of this paper the first extensive use of the second round of data that was collected as part of the Likoma Network Study collection was conducted during January-August 2007, resulting in the *first* longitudinal study of complete sexual networks in sub-Saharan Africa. Using this longitudinal feature of the LNS, we can therefore dynamically investigate in this paper for-the first time in a sub-Saharan African context-the population-level structure of sexual networks, the spread of HIV through these networks, the interac-





Figure 1: Components of the Likoma sexual networks of size six and larger

Zoom graph in pdf version of this application for a more detailed view of this network.

tion between network structures and infection risks, and the interactions between individual's HIV-status, sexual behaviors, and aggregate network structures. This paper will therefore provide new insights to important policy-relevant questions, including: How do sexual networks change over time in populations heavily affected by HIV? How do HIV-positive respondents change their sexual networks once they learn their HIV status? What are the structural positions in sexual networks of individuals who are HIV-positive or who become infected with HIV? Are policy interventions that target "high risk individuals" efficient?

# 2 The Likoma Network Study

The analyses in this paper are based on the *Likoma Network Study* (LNS), an innovative project that conducted a *complete sexual network survey* in Likoma, a small island on Lake Malawi with high HIV prevalence. The data collected as part of the LNS (*i*) two rounds of data on (quasi) complete sexual networks covering the young adult population in seven villages of Likoma collected in 2005/6 and 2007, (*ii*) detailed data on the socioeconomic and demographic situation, subjective

health, and HIV/AIDS related behaviors, attitudes and risk-perceptions of individuals and their sexual network partners, (*iii*) HIV status of respondents and their sexual/social network partners, (*iv*) geographic locations (GPS data) of respondents and their network partners, and (*v*) limited data on migration to and from the island. By choosing Likoma, our study takes advantage—similar to other epidemiological island studies—of the limited range of mobility and the well-defined population boundaries of insular communities (Cliff et al. 2000; Whittaker 1999). These features imply that a high proportion of the islanders' sexual partners reside on the island, thereby increasing the probability of tracing sexual partners.

Studies using a design similar to that of the LNS have been conducted in different contexts, (e.g, Bearman et al. 2004; Klovdahl et al. 1994) but were lacking for African populations with generalized HIV epidemics. The *Likoma Network Study* (LNS) thus constitutes—to our best knowledge the first *longitudinal* sociocentric study of sexual networks among a general population of SSA. While several papers using the first round of LNS data have been published (Helleringer and Kohler 2008; Helleringer et al. 2007, 2009a,b,c,d), this paper will provide the first set of analyses featuring the longitudinal LNS data and the second round of LNS data.

In both rounds of the LNS survey (2005/6 and 2007), the data collection for our sociocentric study of sexual partnerships occurred in two stages. First, we conducted a census of every individual on Likoma island to obtain a roster of potential partners. Second, we conducted a sexual network survey with *all* inhabitants of the individuals aged 18–35 (LNS Round 1) or 18–50 (LNS Round 2) in the study villages (Figure 2), asking respondents for information about their romantic and sexual partners. The saturated sampling frame used in this study then allowed us to construct the population-level sexual network by matching the reported sexual partners with the census roster, and then linking the data of all young adults residing in the sample villages. The context and methodology of this survey are summarized below, and additional details are provided in Helleringer et al. (2009a).

Incomplete-network bias affecting sociocentric network designs implies that the obtained networks are merely "quasi-complete" (Doherty et al. 2005; Ghani et al. 1998). This bias usually occurs because (1) some network members reside outside of the research area and cannot be enrolled, (2) adequate information is not always available to link all reported network partners, and (3) respondents often do not report all of their sexual partnerships (Doherty et al. 2005). We used ACASI to minimize misreporting of sexual behaviors (Bloom 1998; Mensch et al. 2003). The island setting of Likoma was chosen to reduce the potential biases from the other two sources. Indeed, Likoma extends over only 18 square kilometers, has limited transportation to the mainland, and its population is small with just over 7,000 persons living in a dozen villages (Figure 2). As a result, and essential for the data quality of this project, a limited set of identifying information allowed the matching of nominated partners with the census rosters.

#### 2.1 Round 1 data collection (2005/6)

*Roster of potential sexual partners:* All households of Likoma and Chizumulu (a neighboring island distant of 5 kilometers) were enumerated for the Round 1 of the LNS data collection to establish a detailed roster of potential sexual partners. Household informants were asked to provide



**Figure 2: Geographic location of the sampled villages and village-specific participation rates.** Each circle represents a dwelling unit. Dark circles represent dwelling units in the villages that were included in the sexual network survey. Empty circles represent dwelling units in the villages that were not included in this sampling frame. Denominators of the survey participation rates are the total number of eligible respondents (aged 18–35 and their spouses) in a given village, based on the initial household census. Denominators of the HIV testing participation rates are the total number of respondents who completed the sexual network survey in a given village. Island boundaries and location of dwelling units are approximate.

names, maiden names (for married women), nicknames, and sociodemographic characteristics of all household residents. More than 1,300 households were listed in Likoma, and roughly 500 in Chizumulu. The enumeration also included (1) temporary migrants (i.e., household residents who were temporarily absent), (2) family members who moved permanently during the last 5 years, and (3) family members who died during the last 5 years.

Sexual network survey: We subsequently conducted an ACASI sexual network survey with all inhabitants aged 18–35 (and their spouses) of the sampled villages. A total of 923 participants (501 women and 422 men) were interviewed. The age range 18-35 was selected because most HIV infections occur during that span (Heuveline 2003). The location of sampled villages, as well as participation rates, are displayed in Figure 2. Initially, the two villages with the highest proportion of extra-marital births were selected as seed villages (villages 13 and 9), and the remaining five villages were selected based on contiguity to these seeds. The names of up to five partners with whom a respondent had been in a romantic relation with within three years prior to the survey were recorded using headsets. The recall period of three years was chosen because of the long duration of HIV infectivity. A total of 2,040 reports of partnerships were collected during this network survey, among which 1,858 (91.0%) involved sexual intercourse. If a nominated partner was living in Likoma (currently or at the time of the relationship), full names were collected (see (Helleringer et al. 2009a) for a further discussion of this name generator) and respondents were asked to specify his/her residence within the island (village, proximity to a landmark). Using this information, we attempted to identify each nominated partner within the roster of potential partners. This partner tracing was initially conducted using phonetic name-matching algorithms in STATA

(Blasnick 2001), and the links were then verified and/or completed by manual inspection. More than 80% of sexual relations with a partner residing in Likoma were successfully linked to a record in the roster obtained from the island census (Helleringer et al. 2009a). A sexual relationship was assumed to exist if it was reported by *at least one partner*.

The sexual network survey instrument included for each reported relationship the starting and ending dates, the location and occasion of original encounter, the ways in which partners knew each other before the start of the relationship, whether the relationship involved sexual intercourse, the frequency of sexual intercourse, condom use during the relationship, and other measures of relationship quality. All answers to these questions were categorical. In particular, dates of relationships were recorded using the following categories: more than a year ago/within last year/within last month/ongoing. Non-marital partnerships were classified as either stable relationships, infrequent partnerships or one-off casual encounters. Condom use was defined as "consistent" when the respondent reported always using condoms with all partners. Respondents were also asked about their own health and health-care utilization, including the presence of STI symptoms (indicating of the presence of either ulceration of the genital area, discharge or inflamation/irritation during urination), having ever been tested for HIV, and having received an injection during the year prior to the survey. Finally, respondents in six villages were tested for HIV (N = 597; the 7<sup>th</sup> village—village 14 in Figure 2—could not be included due to funding and timing constraints). HIV serostatus was determined using two rapid test assays (Determine HIV/1-2<sup>™</sup> and UniGold HIV<sup>™</sup>), with participation rates varying across villages from 54–82% (p < 0.01, Figure 2).

#### 2.2 Round 2 data collection (2007)

The data collection procedures for the LNS Round 1 were essentially repeated in 2007 to collect a second round of LNS data that is longitudinally linked to the LNS Round 1. However, two important changes in the data collection enhance the second round of LNS data and strengthen the analyses in this paper: *First*, the age range for eligible respondents was changed from 18–35 years to 18–50 years, thereby including a much larger fraction of the sexually active population. Moreover, this extended age range implies that older partners of young adults are included in the sexual network survey, thereby allowing us to identify the extent to which older sexual partners contribute to the connectivity of the sexual network on Likoma Island. In addition to the extended age range, the LNS Round 2 also covered all villages of Likoma Island, thereby eliminating the restriction to 7 villages that were included in Round 1.

In total, 2,281 individuals were interviewed in the sexual network survey as part of the LNS Round 2, 1,007 of whom are longitudionally linked to the LNS Round 1. A total of 2,827 *unique* sexual relationships were reported by respondents in the LNS Round 2. All round 2 data have been completely entered and are available for analyses.

## 3 Methods

The analyses proposed for this 2011 PAA paper will include the following:

- 1. An analyses of the *sexual network on Likoma Island*, following the approach of our earlier analyses of the LNS Round 1 data in Helleringer and Kohler (2007). In particular, these analyses will show if the characteristic features of the sexual network identified by our earlier analyses—including most importantly the high connectivity of the network, the high prevalence of concurrent partnerships, and the existence of a "giant component"—continue to exist in the LNS Round 2 data (2007/8) that includes a broader age range of respondents and additional villages that were not included in our earlier analyses.
- 2. Analyses of the *difference in HIV prevalence* in different components of the sexual network (e.g., as defined by the size or connectivity of a sexual network cluster). In the LNS Round 1 data, for example, the prevalence of HIV also varies significantly across the network, with sparser regions having a higher HIV prevalence than densely connected components, with these differences being partially explained by differences in the several risk factors related to sexual mixing patterns (Helleringer and Kohler 2007).
- 3. An analyses of the *HIV infection risk* as measured by an respondent's distance to the nearest HIV-positive network members. Preliminary analyses of the 2005/6 data, for example, have found that the median distance to the nearest HIV-positive person is substantially shorter in the highly connected core of the sexual network, despite the fact that HIV prevalence in this core of the sexual network is considerably lower than in less connected components of the sexual network.
- 4. Longitudinal analyses of the *changes in the network structure*, including both analyses of the overall network (including measures such as connectivity, size of the largest connected components, etc.) and on the individual level (including measures such as in/out degree, centrality, the existence of multiple concurrent partnerships, etc) between the two rounds of the LNS data, including in particular also analyses about the extent to which the change in sexual networks over time differs by
  - (a) HIV status of respondents at LNS Round 1, and whether respondents knew their HIV status at Round 1;
  - (b) age, sex and socioeconomic and marital status of respondents at LNS Round 1;
  - (c) the structural position of respondents within the sexual network at LNS Round 1; and
  - (d) the involvement of respondents in multiple concurrent sexual partnerships at LNS Round 1.
- 5. Longitudinal analyses of the exposure to HIV infection risk, separately by age, sex and structural network position at LNS Round 1, using measures of exposure to HIV infection such as the distance in the sexual network to the closest HIV-positive network member.

# 4 Conclusion

The sexual networks connecting members of a population have important consequences for the spread of sexually transmitted diseases including HIV. However, very few datasets currently exist that allow an investigation of the structure of sexual networks, particularly in sub-Saharan Africa where HIV epidemics have become generalized. Using the longitudinal sociocentric sex-

ual network data available in the Likoma Network Study (LNS), this paper can provide new and important insights into the structure of sexual networks in sub-Saharan populations, the relationship between the size and structure of sexual networks and HIV infection risks, and for the first time in a sub-Saharan population, the evolution of the population-level sexual network over time, including the differential changes in network position and structure by LNS Round 1 HIV status, marital/socioeconomic status, and Round 1 network position.

# References

- Andrews, J. A., Tildesley, E., Hops, H. and Li, F. (2002). The influence of peers on young adult substance use. *Health Psychology* 21(4): 349–357.
- Bearman, P. S., Moody, J. and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology* 110(1): 44–91.
- Behrman, J. R., Kohler, H.-P. and Watkins, S. C. (2002). Social networks and changes in contraceptive use over time: Evidence from a longitudinal study in rural Kenya. *Demography* 39(4): 713–736. doi:10.1353/dem.2002.0033.
- Behrman, J. R., Kohler, H.-P. and Watkins, S. C. (2009). Lessons from empirical network analyses on matters of life and death in East Africa. In: Kleindorfer, P. R. and Wind, Y. (eds.). *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*. Upper Saddle River, NJ: Wharton School Publishing, chap. 28. URL http://books.google.com/books?id=gjq4mC3hW1QC.
- Blasnick, M. (2001). Gsoundex: Stata module to implement soundex algorithm. Unpublished manuscript, Boston College Department of Economics, URL http://ideas.repec.org/c/boc/bocode/s420901.html.
- Bloom, D. E. (1998). Technology, experimentation, and the quality of survey data. *Science* 280(5365): 847–848.
- Carrington, P. J., Scott, J. and Wasserman, S. (eds.) (2005). *Models and methods in social network analysis*. Cambridge University Press.
- Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 357: 370–379.
- Christakis, N. A. and Fowler, J. H. (2009). *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little Brown & Company.
- Clark, S., Helleringer, S. and Kohler, H.-P. (2008). Do men and women perceive sexual relationships differently? data from matched couples on Likoma Island, Malawi. Paper presented at the annual meeting of the Population Association of America, New Orleans, LA, April 17–19, 2008.
- Cliff, A. D., Haggett, P. and Smallman-Raynor, M. R. (2000). *Island Epidemics*. Oxford: Oxford University Press.
- Doherty, I. A., Padian, N. S., Marlow, C. and Aral, S. O. (2005). Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *Journal of Infectious Diseases* 191: S42–S54.
- Fiori, K. L., Antonucci, T. and Cortina, K. S. (2006). Social network typologies and mental health

among older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 61(1): P25.

- Fowler, J. H. and Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *Britisch Medical Journal* 337: a2338–. doi:10.1136/bmj.a2338.
- Ghani, A. C., Donnelly, C. A. and Garnett, G. P. (1998). Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. *Statistics in Medicine* 17: 2079–2097.
- Helleringer, S. and Kohler, H.-P. (2007). Sexual network structure and the spread of HIV in Africa: Evidence from Likoma Island, Malawi. *AIDS* 21(17): 2323–2332. doi: 10.1097/QAD.0b013e328285df98.
- Helleringer, S. and Kohler, H.-P. (2008). Cross-sectional research design and relatively low HIV incidence, rather than blood exposures, explain the peripheral location of HIV cases within the sexual networks observed on Likoma. *AIDS* 22(11): 1378–1379. doi: 10.1097/QAD.0b013e3282fc733b.
- Helleringer, S., Kohler, H.-P. and Chimbiri, A. (2007). Characteristics of external/bridge relationships by partner type and location where sexual relationship took place. *AIDS* 21(18): 2560– 2561. doi:10.1097/QAD.0b013e3282f112bd.
- Helleringer, S., Kohler, H.-P., Chimbiri, A., Chatonda, P. and Mkandawire, J. (2009a). The Likoma Network Study: Context, data collection and initial results. *Demographic Research* 21(15): 427– 468. doi:10.4054/DemRes.2009.21.15.
- Helleringer, S., Kohler, H.-P. and Kalilani-Phiri, L. (2009b). The association of HIV serodiscordance and partnership concurrency in Likoma Island (Malawi). *AIDS* 23(10): 1285–1287. doi: 10.1097/QAD.0b013e32832aa85c.
- Helleringer, S., Kohler, H.-P., Kalilani-Phiri, L., Frimpong, J. A. and Mkandawire, J. (2009c). Multiple concurrent partnerships in a sub-saharan setting: Biases in self-reported measures and their implications for HIV/AIDS prevention. Paper presented at the Annual Meeting of the Population Association of America, Detroit, MI, April 30–May 2, 2009.
- Helleringer, S., Kohler, H.-P. and Mkandawire, J. (2009d). Increasing uptake of HIV testing and counseling among the poorest in sub-Saharan countries through home-based service provision. *Journal of Acquired Immune Deficiency Syndromes* 51(2): 185–193. doi: 10.1097/QAI.0b013e31819c1726.
- Heuveline, P. (2003). HIV and population dynamics: A general model and maximum-likelihood standards for East Africa. *Demography* 40(2): 217–245.
- Kaplan, C. P., Nápoles-Springer, A., Stewart, S. L. and Pérez-Stable, E. J. (2001). Smoking acquisition among adolescents and young Latinas The role of socioenvironmental and personal factors. *Addictive Behaviors* 26(4): 531–550.
- Klovdahl, A., Potterat, J., Woodhouse, D., Muth, J., Muth, S. and Darrow, W. (1994). Social networks and infectious disease: The Colorado Springs Study. *Social Science and Medicine* 38: 79–88.
- Kohler, H.-P. (1997). Learning in social networks and contraceptive choice. *Demography* 34(3): 369–383. URL http://www.jstor.org/stable/3038290.

- Kohler, H.-P. (2001). *Fertility and Social Interactions: An Economic Perspective*. Oxford: Oxford University Press. URL http://books.google.com/books?id=03WVNSjf1KkC.
- Kohler, H.-P., Behrman, J. R. and Watkins, S. C. (2007). Social networks and HIV/AIDS risk perceptions. *Demography* 44(1): 1–33. doi:10.1353/dem.2007.0006.
- Manski, C. F. (2000). Economic analysis of social interaction. *Journal of Economic Perspectives* 14(3): 115–136.
- Mensch, B. S., Hewett, P. C. and Erulkar, A. (2003). The reporting of sensitive behavior among adolescents: A methodological experiment in Kenya. *Demography* 40(2): 247–268.
- Morris, M. (ed.) (2004). Network Epidemiology. Oxford: Oxford University Press.
- Smith, K. P. and Christakis, N. A. (2008). Social networks and health. *Annual Review of Sociology* 34: 405–429. doi:10.1146/annurev.soc.34.040507.134601.
- Soetevent, A. R. (2006). Empirics of the identification of social interactions; an evaluation of the approaches and their results. *Journal of Economic Surveys* 20: 193–228.
- Strauss, R. S. and Pollack, H. A. (2003). Social marginalization of overweight children. *Archives of Pediatrics and Adolescent Medicine* 157(8): 746.
- Whittaker, R. (1999). Island Biogeography. Oxford: Oxford University Press.