Title: Using social networks to understand and overcome implementation barriers in the global HIV response

Running head: Social networks for implementation science

Authors: Guy Harling 1,2,3,4, Alexander C. Tsai 3,5,6

Affiliations:

1. Institute for Global Health, University College London, London, United Kingdom
2. Africa Health Research Institute, KwaZulu-Natal, South Africa
3. Department of Epidemiology and Harvard Center for Population and Development Studies, Harvard University, Cambridge MA, United States
4. MRC/Wits Rural Public Health & Health Transitions Research Unit (Agincourt), University of the Witwatersrand, Johannesburg, South Africa
5. Chester M. Pierce, MD Division of Global Psychiatry, Massachusetts General Hospital, Boston MA United States
6. Mbarara University of Science and Technology, Mbarara, Uganda

Corresponding author: Guy Harling. Institute for Global Health, Mortimer Market Centre, London, WC1E 6JB. Email: g.harling@ucl.ac.uk; Phone: +44 20 3108 2069.

Conflicts of Interest and Sources of Funding: GH receives funding from the Wellcome Trust and Royal Society (210479/Z/18/Z) and the U.S. National Institutes of Health (NIH) (R01MH114560, R01HD084233). ACT acknowledges funding from NIH R01MH113494.
Acknowledgements: The authors would like to thank Jimi Adams and Stephane Helleringer for several helpful comments, and Justin D. Rasmussen for his assistance in producing the sociocentric social network graph using data from the HopeNet cohort study.

This is the Accepted Manuscript (not final published version) of an article published in the Journal of Acquired Immune Deficiency Syndromes on DATE, available online at: DOI, under the citation: CITATION.
Abstract

Despite the development of several efficacious HIV prevention and treatment methods in the past two decades, HIV continues to spread globally. Uptake of interventions is non-randomly distributed across populations, and such inequality is socially patterned and reinforced by homophily arising from both social selection (becoming friends with similar people) and influence (becoming similar to friends). In this article we describe how social network analysis (SNA) methods, including egocentric, sociocentric, and respondent-driven sampling designs, provide tools to measure key populations, to understand how epidemics spread, and to evaluate intervention take-up. We highlight that SNA-informed designs can improve intervention effectiveness by reaching otherwise inaccessible populations and improve efficiency by maximizing spillovers to at-risk but susceptible individuals through their social ties. We argue that SNA-informed designs thus have the potential to be both more effective and less unequal in their effects compared with SNA-naïve approaches. While SNA-informed designs are often resource-intensive, we believe that they can provide unique insights that can help reach those most in need of HIV prevention and treatment interventions. Increased collection of social network data during both research and implementation work would provide important information to improve the roll-out of existing studies in the present and to inform the design of more data-efficient, SNA-informed interventions in the future.
Introduction

The global HIV response has been transformed by new discoveries during the past two decades, including the preventive efficacy of antiretroviral therapy, pre-exposure prophylaxis (PrEP), insertable microbicides, and medical male circumcision.\(^1\)\(^-\)\(^5\) However, the field has had less success translating these efficacious interventions into effective implementation at scale. Quantifying this gap between evidence and practice, understanding why it exists, and developing strategies to overcome it is a core goal of HIV implementation science.\(^6\) One of the most challenging aspects of successful implementation of HIV prevention programs has been motivating participants to initiate, adhere to, or maintain use of these interventions.\(^7\)\(^-\)\(^9\) In the setting of sub-optimal utilization, one approach to maximizing HIV prevention, as illustrated by the DREAMS partnership,\(^10\) involves providing an intensive package of multiple evidence-based interventions to those at greatest risk.

However, insofar as uptake and adherence are correlated across interventions, such user-dependent approaches to public health intervention at scale might risk iatrogenically increasing inequalities in HIV risk within populations, leaving the most vulnerable at risk of both HIV acquisition and HIV-associated stigma.\(^11\)\(^-\)\(^13\) This phenomenon – variously described as the “inverse care law”,\(^14\) the “inverse equity hypothesis”,\(^15\) and the “inequality paradox”\(^16\) – arises when those most in need of interventions have least access to or ability to use them, while those least in need can leverage their social or economic resources to gain access to these interventions.\(^11\)\(^-\)\(^13\) As a result, the gap between less- and more-vulnerable populations grows. This pattern is often reinforced by social separation between those at greatest risk (e.g., people who use drugs, sex workers) and the rest of the population.

Furthermore, when intervention uptake is non-randomly distributed within the population, risks for all are increased, as illustrated in the recent U.S. measles outbreak.\(^17\)\(^,\)\(^18\) While structural interventions (that are not user-dependent, and which might benefit everyone within the population)
can potentially mitigate such disparities, intervention approaches that can deliver proven methods effectively and efficiently to at-risk groups within populations will continue to remain a valuable part of the HIV prevention toolkit.

Innovative ideas, and corresponding behaviors, typically diffuse to populations via communication channels. While mass media campaigns might have broad capacity for achieving saturated awareness, changing (or maintaining) behaviors typically requires deeper engagement and interaction. Social support, the communication of social norms, and other forms of social influence that can generate such sustained engagement flow through social network ties. In this narrative review, we therefore argue that adoption of a social network approach seems natural and likely to be fruitful in helping us understand how intervention uptake and adherence are distributed across communities, how this patterning is generated and maintained, and how barriers to uptake and adherence might most efficiently be overcome. Throughout the article we use the term ‘social network’ to refer to a social structure that consists of ‘nodes’ connected by ‘ties’; nodes can represent people or organizations, and ties can represent friendship or any other type of interdependency (e.g., communication or physical proximity).

**Social network data collection**

Three social network data collection methods are commonly used. They can be roughly categorized by the extent to which the data can be linked and the extent to which the sampling frame is a priori known. ‘Egocentric’ network data are defined as those in which it is not possible to directly link the information provided by each respondent. These data are typically elicited through surveys, in which respondents (‘egos’) are asked to identify social network ties who occupy specific or general roles of social importance in their lives (‘alters’). Following the administration of these ‘name generators,’ further information is sought from the ego through ‘name interpreters,’ which
are questions that elicit information about characteristics of the named alters, and about characteristics of the ties between the ego and their alters.

One extension of egocentric data collection is to elicit from the ego perceived properties of ties between each pair of alters, sometimes called “ego-centered cognitive social structures”29,30. This design results in the ability to better typify the immediate social network of each ego, particularly the level of triadic closure, i.e., triangles linking an ego and two of their alters. However, it requires many additional questions (e.g., for five alters, there are 10 undirected alter-alter ties), often leading to interview fatigue and reduced data quality.31 Egocentric data collection is relatively straightforward, and can be conducted with a population sample using standard survey methods (Figure 1). However, egocentric data rarely allow for network-wide processes or outcomes to be directly evaluated, because typically no one reports on everyone else in the network, and because egos’ responses cannot be linked to one another.

A classic form of egocentric network study design used to study HIV is the sexual or drug-use partner roster. By collecting information on alter characteristics (e.g., age, geographic location, substance use profile) and tie characteristics (e.g., relationship duration, condom use pattern) researchers can evaluate someone’s HIV acquisition and transmission risk profile more comprehensively.32 In the context of HIV research, egocentric alter-alter ties are rarely analyzed, despite their potential importance for single-mode behaviors where triadic closure is possible (e.g. condomless anal intercourse,33 needle/syringe sharing34), and for identifying potential change agents within populations.35

In contrast to egocentric data, ‘sociocentric’ data arise when all egos in a population are sampled (Figure 2), and each alter is identified, i.e., entity resolution. Entity resolution can be done either by having egos select their alters from a closed list (e.g., school register, employee list, population census) or by eliciting sufficient information that each alter can be later identified by the
The latter approach typically leaves a substantial proportion of alters unidentified; on the other hand, having an *a priori* census may be economically or practically infeasible, especially when the population is not well described. Sociocentric data can be extremely powerful and, when captured longitudinally, can identify how infections and interventions spread through communities, something that has been done from time-to-time since the early days of the epidemic. However, the requirement for well-enumerated, quasi-closed populations makes such studies rare, particularly in the resource-limited settings where HIV is most prevalent.

The collection of sociocentric data is resource-intensive within each community, compared with egocentric approaches. The need to obtain information from a whole-population sample of egos therefore excludes the possibility of sampling. However, by typically focusing on a smaller number of communities, sociocentric designs (compared with multi-site egocentric designs) can reduce the costs of travel between communities and the need for repeated community engagement. The relative benefits of the two approaches depend largely on community-level heterogeneity: if all communities are homogeneous, a sociocentric examination of a few populations will provide greater insight; if communities are heterogeneous, a sociocentric network study focused fewer communities is more likely to miss the full range of social structures and risk environments. One way to circumvent the resource requirements of sociocentric network data collection is by using existing network datasets, such as online social networks or phone records. While such methods can be extremely powerful, they rarely provide the specific kinds of detail required to evaluate HIV programs. In addition, the substantial quantity of identifiable data that must be collected and processed for sociocentric work, especially relating to HIV, requires careful ethical consideration.

Both egocentric and sociocentric methods are based on interviewing some or all of a pre-defined population. Some populations, however, are hidden insofar as their membership is not *a priori* known. In such situations, partial network sampling methods – of which ‘respondent driven
"sampling’ (RDS) is a common example – play an important role. RDS begins by interviewing a few, often visible, members of the population. After completing the initial round of interviews, first-round egos (‘seeds’) are provided with a mechanism to invite alters to participate in the survey (e.g., a voucher for food or airtime that can be redeemed with the study team). In this way, subsequent egos ‘self-refer’ to the study. These referral chains help the study team to survey individuals who might otherwise not have been identified using conventional approaches (Figure 3); however, since the overall population of interest is usually not well characterized, deriving population-representative estimates from RDS designs is difficult. Several methods have been proposed for assessing and adjusting for such potential biases. However, there is little consensus about the criterion standard method for doing so, and concerns remain about the tenability of the strong assumptions required to move from partial network samples to population-level inference.

In HIV-related research, RDS is commonly used when working with populations defined by criminalized or stigmatized activities, including injection drug use, same-sex intercourse, and sex work.

Each of these social network data collection methods has advantages and disadvantages related to cost, feasibility, and validity. None are considered to have advantages over the others as a general matter. However, depending on the specific research aims, some advantages (or disadvantages) may be more heavily weighted than others, leading investigators to choose one study design over another.

**Using social networks to understand HIV treatment and prevention**

In this article, we focus on three comparative advantages of the social network approach, namely the kinds of information it can provide where other methods cannot, and its relevance for three important aspects of the HIV epidemic: risk population definition, infection spread, and behavior patterns.
First, social network analysis (SNA), particularly using RDS methods, can help us understand the characteristics and risk behaviors of hard-to-reach, key populations. Such knowledge allows us to see who is unable or unwilling to take up interventions and, potentially, why. The power of SNA in this context arises from the principle of homophily: an ego’s alters tend to be more similar to the ego than to other egos in the population.\textsuperscript{52} As a result, studies that elicit information from egos about their alters yield additional information about the risk groups to which the egos belong, including information about seroprevalence. This exercise is particularly important in the context of HIV since many of those most at risk are members of stigmatized or criminalized populations. In the case of RDS, egos are asked to refer alters who fit a particular profile, allowing detection and identification of specific key populations,\textsuperscript{53} including people who use injection drugs,\textsuperscript{54-56} MSM,\textsuperscript{49,57-60} and people who engage in sex work.\textsuperscript{61-64} Similarly, qualitative descriptions of the social networks of key populations can provide vital insights into the activities of people at risk of HIV transmission and acquisition, notably of the interactions between members, improving understanding about barriers and facilitators of intervention delivery.\textsuperscript{65,66} One form of RDS that can be particularly important is partner notification or contact tracing. Partner notification/contact tracing avoids some of the concerns of typical RDS designs regarding selective recruitment, since the investigator is seeking to capture all contacts fitting a particular profile, and may thus play an important (if partial) role in understanding HIV-related networks.\textsuperscript{67}

Second, SNA that captures sexual and other risk interactions provides novel understanding of how HIV spreads through populations. Such complexity is highlighted in modelling work showing how the impacts of homophily in use of antiretroviral therapy or partner selection (i.e., HIV serosorting) differ depending on the population prevalence of HIV.\textsuperscript{68} Since the 1990s, researchers have used social network data (e.g., partner notification and contact tracing) to understand sexual behavior\textsuperscript{69} and epidemic change for both HIV\textsuperscript{32,70} and other sexually transmitted infections.\textsuperscript{71} More recently, two approaches have been used to capture sexual networks \textit{de novo}: (i) sociocentric
methods focusing on quasi-closed communities, such as those confined to islands \(^{36}\) or those with well-defined social networks where boundary specification is relatively straightforward;\(^{72,73}\) and (ii) RDS-based approaches attempting to gain saturated coverage of behaviorally or geographically defined groups,\(^{74-77}\), although there remain important ethical issues to consider in such detailed inference.\(^{78}\) In both approaches, longitudinal data allow the identification of potential sources of transmission. One powerful emerging use of social networks for understanding HIV transmission is its combination with phylogenetic data, thereby attempting to minimize the limitations of both data types,\(^{79-81}\) although there remain important ethical issues to consider in such detailed inference.\(^{78}\)

Third, SNA can allow us to observe how behavior is distributed across populations and how novel ideas and behaviors spread through populations. Egocentric and sociocentric studies of MSM have demonstrated that there is homophily in behavior between egos and alters,\(^{82,83}\) and that social network characteristics are correlated with use of PrEP.\(^{33}\) SNA studies of the spread of interventions through networks are rare and are often limited to egocentric designs due to the larger budgets typically required for longitudinal sociocentric data collection,\(^{84}\) although RDS methods can also be appropriate.\(^{85}\) In the context of vaccine uptake, there is evidence that social norms are central to the generation and maintenance of vaccine hesitancy within online social networks.\(^{86}\)

SNA also provides an opportunity to understand how increasingly popular HIV self-test kits are distributed and used by sexual partners\(^ {87}\) and by friends and family.\(^{88,89}\) A key aspect of the diffusion of new technologies, such as HIV self-test kits, is the dynamic interplay between multiple overlaid networks, in terms of both structure (e.g., friends, drug use and sexual partners) and content (e.g., knowledge, behavior and infection).\(^ {90}\) Understanding how using self-test kits and learning one’s status affect social and sexual networks, and thus onward diffusion of use, is vital to evaluating the likely effectiveness of peer-based distribution methods.
Any attempt to infer causal processes in uptake must account for the strong dependencies between connected individuals, while also disentangling the effects of homophily (being friends with similar people), social selection (becoming friends with similar people) and social influence (becoming similar to friends). Several approaches that utilize all three types of social network data have been proposed for conducting causal inference around peer effects, including stochastic actor-oriented models, peer effects models, instrumental variables estimation, and dynamic matched sample estimation, but there remain ongoing debates about the extent to which causal inference is possible in these settings.

One important subset of behavior change is that which is initiated intentionally, i.e., through intervention deployment. SNA can therefore also provide important information on where and why non-network informed interventions fail. Qualitative investigation of several past implementation trials has shown non-uptake or non-adherence to protocols; SNA in such situations could provide information about how non-adherence is patterned or even spread through the trial communities – information that could help with subsequent roll-out of the same or similar interventions.

Social network-informed interventions for HIV treatment and prevention

The above three SNA uses can help improve understanding of the HIV epidemic and barriers to its implementation. Beyond these functions, several social network-informed intervention (SNI) methods can be deployed to help overcome implementation barriers encountered with other approaches. Valente has proposed four types of network interventions: individual; segmentation, or group change; inducted interaction; and network tie alteration. Randomized trials of such interventions to improve a range of health outcomes have often been successful.
Individual interventions are by far the most common SNI in HIV prevention. Examples include seeding behavior change messages (with the expectation that these messages will be passed on to social ties) and using network information to inform intervention targeting. Past experience has shown that carefully selected individuals can shift HIV risk behavior within populations (e.g. community popular opinion leaders [CPOLs]\textsuperscript{107}), but only when the CPOLs selected were truly influential.\textsuperscript{108,109} Diffusion approaches that rely on random seeding are unlikely to reach everyone in the population, much less those at highest risk who are often hardest to reach,\textsuperscript{110} since by definition previous messaging has already failed to reach them. Careful planning and evaluation of individual SNIs is thus vital.

Induced interventions can be seen in the example of treatment supporters for HIV treatment and care. Such supporters may be either identified from existing contacts who are asked to assume new roles,\textsuperscript{111,112} or created \textit{de novo},\textsuperscript{113,114} both approaches appear to increase treatment adherence.\textsuperscript{115-117} Examples of network tie alteration strategies are rarer, not least because SNIs that eliminate ties or otherwise attempt to fragment/damage social networks pose obvious ethical problems. SNIs that use SNA to guide tie formation may hold more promise, but compelling studies in the field of HIV are yet to be done.\textsuperscript{118}

SNA highlights the importance of initial conditions for intervention targeting in two ways. First, when identifying initial seeds for behavior change, it is vital to choose those who are influential within their communities.\textsuperscript{119} Randomly selected seeds will be less efficient at inducing behavior change, since they will exert weaker spillover effects on others,\textsuperscript{120} while the earliest adopters of innovations are rarely the most influential since they are usually by definition not deeply embedded in local networks.\textsuperscript{121} Second, SNIs that select people who are already interested in behavior change will likely miss key populations.\textsuperscript{122} It is therefore important for an SNI to ensure broad coverage of
the social space of interest (e.g., a school, a village, young MSM living in a city) and, potentially, to selectively enroll individuals at greatest risk.

Another lesson from SNA is the importance of complex contagion,123 where uptake or infection requires exposure to multiple sources. The threshold level of exposure required for complex contagion may be an absolute number or a proportion (e.g., of social ties who exert influence on one’s behavior). In the context of HIV prevention, there is a close analogy to the importance of social norms, both descriptive and injunctive: behavior change becomes much more likely once the behavior is adopted or endorsed by a sufficiently large part of one’s social referent group.124,125 In practical terms, ideas of complex contagion reinforce the importance of sufficiently dense seeding of the most at-risk parts of any social network, suggesting that random targeting will do even worse than one might expect.

**Some ways forward**

SNA provides a range of insights into how HIV interventions can be changed and improved. However, implementing many of these insights will require better data than are currently available. While social network descriptions exist, notably for school-aged children and some specific well-defined populations (e.g., people who use injection drugs or MSM in higher-income settings), we know still little about the social networks of those at greatest risk. In particular, the continued high incidence of HIV among adolescents and young adults in sub-Saharan Africa126 is not reflected in an understanding of their social networks, with a few notable exceptions.72,127 A fuller understanding of the patterning of attitudes and behaviors across social space, how such factors change over time, and indeed how social networks evolve, in the context of youth and other key populations throughout sub-Saharan Africa would greatly aid ongoing efforts to reduce HIV incidence in this setting. Much of these data will need to be longitudinal, which adds standard cohort design issues (e.g., entry/exit conditions, loss to follow-up) to social network design issues (e.g., defining boundaries, missing tie
information). As we implement SNIs in these populations (often without considering them to be SNIs), it will be vital to collect sufficient social network data to enable evaluation of how social networks impact intervention effects as well as how interventions rewire social networks – we should expect social networks to evolve under the pressure of new ideas spreading across them.\textsuperscript{128}

Despite this need for better data in order to understand social processes, it is not often feasible to collect sociocentric or even egocentric data routinely. Developing scalable HIV interventions that draw on social network insights will therefore require ways to quickly identify key individuals for SNIs. While occasional examples of low-cost social network data collection processes exist, such as downloading contacts from cellphones,\textsuperscript{75} there may be other settings in which they cannot be implemented without ethical problems or substantial refusals. Sociocentric datasets that measure social influence can be used to test whether algorithms based on non-network properties can identify influential individuals within populations. Such algorithms can then be tested in randomized studies to determine whether such non-network-identified seeds are more effective than randomly- or network-selected seeds at reducing HIV risk within the network. More evaluations of SNIs are also required, although such studies can be prohibitively expensive as they require the collection of sociocentric social network data from many independent networks that can then be randomized to different SNIs.\textsuperscript{129,130}

Even without social network data, however, social network theory highlights that there are low-cost ways to improve on random selection for intervention targeting. For example, the “friendship paradox” highlights that, on average, the friends of individuals have more friends than the individuals who nominated them.\textsuperscript{131} Other evidence suggests that people tend to identify as friends people who are of higher social status than themselves.\textsuperscript{132} A theoretical prediction thusly emerges: asking a random sample of individuals to identify their friends and then enrolling a random friend should improve the reach and effectiveness of interventions.\textsuperscript{133} This prediction was borne out
in a recent SNI trial of multivitamins in Honduras;\textsuperscript{134} the same research group also showed that this theoretical prediction could be leveraged to predict an outbreak of H1N1 influenza six weeks in advance even in the absence of having full network data.\textsuperscript{135} A SNI that randomly selects the most-connected alter (rather than a random alter) may be even more effective.\textsuperscript{136}

Another insight that has been obtained through SNA is the importance of social affiliation: the connection of individuals through shared group membership or use of physical\textsuperscript{137,138} or online\textsuperscript{139,140} locations. Social affiliation is particularly important when affiliation is associated with risk behavior, and thus either acquired infection or attitudes. Social spaces that are characterized by greater prevalence of HIV transmission risk behavior are likely to be vital sites for intervention, especially when the interventions in question are network-informed – the original CPOL work engaging opinion leaders in gay bars in small midwestern US towns being a canonical example.\textsuperscript{107}

Our review has focused on individual-level intervention uptake in the context of community networks. However, social networks are also likely to play key roles in determining which evidence-based interventions are embraced by policymakers, providers, organizations, and communities. In particular, the choice of which interventions to roll out will depend on who is championing them and how they can influence those with the power to make funding decisions. Delayed roll-out of antiretroviral therapy in South Africa after the turn of the millennium,\textsuperscript{141,142} or of PrEP in the National Health Service in England in recent years,\textsuperscript{143} might well be understood through a social network lens where those with decision-making power were influenced by individuals not convinced of the interventions’ effectiveness. SNA can also be applied to organizational networks to identify isolated organizations (that may not be in a position to take up an intervention), central organizations (that might serve as either positive or negative influences on other organizations within the network), or disconnected subgroups of organizations (that could potentially be brought together by organizational brokers for the purposes of consolidating intervention efforts).\textsuperscript{144-146} A meaningful
understanding of the existing network structure of potential implementing organizations can potentially facilitate the deployment and maintenance of the implementation process. Consistent with this idea, Gesell and colleagues have proposed a method by which social networks can be monitored during the process of intervention implementation, with the aim of using “social network diagnostics” to guide program activities.147

Conclusion

The history of HIV contains many examples of proven interventions being deployed to the benefit of communities at risk. It also contains several examples where these interventions have left behind some groups – often those who are most vulnerable and least visible. Social network analysis can help us to identify who is being missed and why, and what barriers limit their capacity to use the interventions being offered. Social network-informed interventions can work with the grain of society to improve the reach of interventions, especially to key populations, and to maximize intervention impact once delivered. To achieve these objectives, it is vital that we collect more social network data now, and strategically study how to minimize our need of such resource-intensive data in the future. Achieving these goals will help curtail the HIV epidemic while not iatrogenically worsening health disparities.
**Figure 1.** A sample of egocentric social networks derived from a cross-sectional study of sources of sexual advice for young adults living in uMkhanyakude district, KwaZulu-Natal, South Africa. This figure presents social support ties between 31 respondents (women aged 18-24, dark blue squares) and their 123 social contacts. Light blue squares are female contacts; brown circles are males; dark brown circles are males with whom the respondent has now or previously had a sexual relationship. Line thicknesses are proportional to reported frequency of communication between individuals. Absent lines between respondents and sexual partners indicate no report of social support from this individual.
Figure 2. A sociocentric social network graph derived from the HopeNet cohort study, a whole-population cohort of all adult men and women who report stable residence in Mbarara, Uganda. This undirected network graph represents the emotional support ties of 140 women in one of the study villages, Buhingo, with HIV-negative women or women whose serostatus is unknown depicted with solid circles and HIV-positive women depicted with hollow circles.
Figure 3. A set of recruitment networks derived from a respondent-driven sampling survey of 405 bisexuals in Ontario, Canada. Blue squares represent initial seeds and purple circles subsequently referred recruits. Arrows represent referral links, where eligibility criteria for participation were: identification as attracted to individuals of more than one sex/gender, age 16 years or older, and residence in Ontario, Canada.
References


93. Ogburn EL. Challenges to estimating contagion effects from observational data. *Complex Spreading Phenomena in Social Systems:* Springer; 2018:47-64.


140. Young LE, Fujimoto K, Schneider JA. HIV prevention and sex behaviors as organizing mechanisms in a Facebook group affiliation network among young black men who have sex with men. *AIDS Behav.* 2018;22(10):3324-3334.


