Connecting the dots: network data and models in HIV epidemiology

Wim Delva[a,b,c,d,e], Gabriel E. Leventhal[e,f] and Stéphane Helleringer[g]

Effective HIV prevention requires knowledge of the structure and dynamics of the social networks across which infections are transmitted. These networks most commonly comprise chains of sexual relationships, but in some populations, sharing of contaminated needles is also an important, or even the main mechanism that connects people in the network. Whereas network data have long been collected during survey interviews, new data sources have become increasingly common in recent years, because of advances in molecular biology and the use of partner notification services in HIV prevention and treatment programmes. We review current and emerging methods for collecting HIV-related network data, as well as modelling frameworks commonly used to infer network parameters and map potential HIV transmission pathways within the network. We discuss the relative strengths and weaknesses of existing methods and models, and we propose a research agenda for advancing network analysis in HIV epidemiology. We make the case for a combination approach that integrates multiple data sources into a coherent statistical framework.

AIDS 2016, **30**:2009–2020

Keywords: contact tracing, epidemiology, HIV, mathematical models, partner notification, phylogenetics, sexual networks, survey data

Introduction

HIV is transmitted through social networks that are formed primarily by the sexual relationships and needle-sharing practices individuals engage in. These networks are important determinants of the magnitude of HIV epidemics [1–6], and the effectiveness of HIV prevention [7–12]. Unfortunately, investigating these networks empirically is difficult, because it requires not only collecting data on sensitive risk behaviours, but also the mapping of relationships that connect people at risk of transmitting and acquiring HIV [13].

In this study, we review how the disciplines of the social sciences, molecular biology, and public health have developed various approaches to collect and analyse network data for the purpose of understanding HIV epidemiology and enhancing HIV prevention and treatment. First, we consider key aspects of network analysis relevant for HIV epidemiology. Next, we provide an overview of existing methods for collecting and analysing network data, and we discuss their relative strengths and weaknesses. Lastly, we emphasize the need for an interdisciplinary approach to network analysis in HIV epidemiology and provide examples of the

[a]Center for Statistics, Hasselt University, Diepenbeek, Belgium, [b]The South African Department of Science and Technology-National Research Foundation (DST/NRF) Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa, [c]International Centre for Reproductive Health, Ghent University, Gent, [d]Rega Institute for Medical Research, KU Leuven, Leuven, Belgium, [e]Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland, [f]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, and [g]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.

Correspondence to Wim Delva, MD, PhD, Professor of Global Reproductive Health, Ghent University, Belgium and Deputy Director Research, SACEMA, Stellenbosch University, 19 Jonkershoek Road, Stellenbosch 7600, South Africa.

Tel: +27 21 808 2589; fax: +27 21 808 2586; e-mail: wimdelva@gmail.com
Received: 23 February 2016; revised: 29 May 2016; accepted: 7 June 2016.

DOI:10.1097/QAD.0000000000001184

synergistic benefits of a combination approach to the challenge of inferring the HIV transmission pathways from which HIV epidemics emerge.

Network epidemiology and its challenges

Social networks are sets of links (also called 'connections', 'ties' or 'edges') between individuals (also called 'nodes' or 'vertices'). In common visualizations of networks (Fig. 1), individuals are represented by symbols. The shape and colour of each symbol may capture characteristics relevant to HIV transmission (e.g. gender, race, ethnicity and viral load). Lines that connect the symbols represent links, and the line width and dotting pattern can reflect link characteristics (e.g. frequency of sexual intercourse, condom use, needle sharing). A large number of mathematical tools exist to analyse the data that underlie networks [14]. These tools allow to, for example, identify who is connected to whom, or measure the (social) distance between people living with HIV (PLWH) and uninfected people.

Network analysis in HIV epidemiology revolves around identifying HIV transmission pathways, that is, the subsets of links and individuals across which HIV can spread. Some of these pathways are realized; they connect the PLWH that transmitted HIV to one another. Other pathways represent chains of potential transmission events; they link PLWH to individuals who are not (yet) infected, but are at risk of acquiring HIV in the future because of their network connections. Being able to map and anticipate these pathways helps determining which individuals are most at risk of HIV infection, and select the interventions most likely to interrupt HIV transmission.

The social networks relevant for HIV transmission have complex structures. They often connect individuals both through sexual and needle-sharing relationships [12,15,16]. In these 'multiplex' networks, the interactions between the different modes of HIV transmission can affect the impact of HIV prevention interventions [17]. Importantly, social networks evolve over time as individuals end current relationships and form new ones. Through this dynamic process, new potential HIV transmission pathways emerge, whereas others are interrupted [18–21]. For example, if two individuals 'A' and 'B' are in a relationship at time 1, but then break up, and 'B' forms a relationship with 'C' at time 2, there is a path for HIV transmission from 'A' to 'C' (via 'B'), but there is no path through which 'C' can transmit HIV to 'A'. The time ordering of relationships can be incorporated into network data [19] and can be represented visually in network animations [22]. Finally, network data can include information about who infected whom. For example, if 'A' infected 'B', who subsequently infected 'C', then the path from 'A' to 'B' to 'C' could be represented using arrows (Fig. 1d). Such 'directed' network data help better understand which individuals are central in HIV transmission pathways.

It is rarely possible to collect complete data on the social networks that propagate HIV because it is challenging to list all the individuals and links that form a network. Social scientists, molecular biologists, and public health specialists have thus developed approaches to collecting partial or indirect network data, and subsequently infer HIV transmission pathways from such incomplete information. The social science approach uses data from behaviour and relationship surveys, often but not necessarily in combination with HIV testing, to infer potential HIV transmission pathways (Fig. 1c), whereas the molecular biology approach uses HIV genetic sequences from PLWH to identify realized HIV transmission pathways (Fig. 1e). The public health approach seeks to identify high-risk networks by tracing and testing people connected to newly diagnosed HIV cases (Fig. 1f).

The social science approach

The workhorse of the social science approach is the egocentric network survey [13]. In this type of study, a random sample of individuals from the population of interest (Fig. 1c) is asked to provide information about their recent sexual and/or needle sharing partners (e.g. their age, race, and gender), and to describe the characteristics of these relationships (e.g. start and end dates, condom use). Respondents may also be invited to test for HIV infection [24].

Egocentric surveys have been conducted worldwide [25–34]. They permit measuring characteristics of the personal networks of respondents such as homophily (the propensity to engage in partnerships with others who share similar characteristics), or concurrency (the likelihood of having more than one ongoing relationship at one point in time) [35]. But they do not provide data on HIV transmission chains because the partners of sampled respondents are not typically enrolled in the study [13].

In this data context, potential HIV transmission pathways can only be inferred. Network inferences have greatly improved recently, due to the development of exponential random graph models (ERGMs). ERGMs are a family of statistical models that can accommodate the interdependencies between individuals that characterize network datasets [36–42]. They were originally developed for the analysis of complete network datasets in which all individuals and links are listed. But they can also be used with incomplete data from egocentric studies under certain simplifying assumptions [43–45].

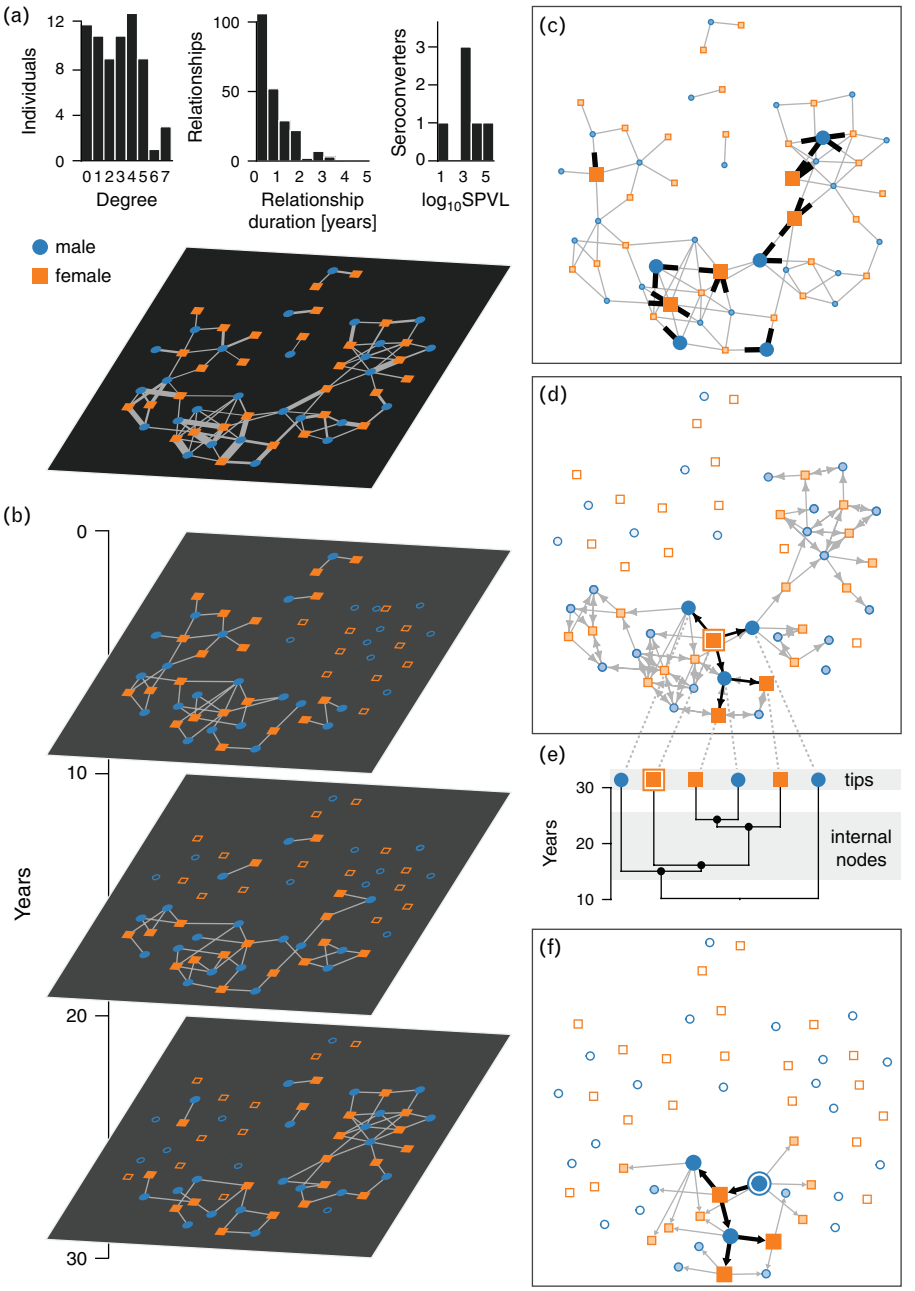

Fig. 1. Graphical representations of simulated social networks and HIV transmission pathways. All data from Fig. 1 are synthetic and were generated using Simpact, a freely available agent-based modelling tool for simulating HIV transmission in dynamic sexual networks [23]. (a) Complete (cumulative) network of all relationships that were formed over a 30-year time period between members of the simulated population. Thicker links represent longer relationships. Histograms at the top of (a) represent the degree distribution (lifetime number of partners) in the population, the distribution of relationship durations, and the distribution of $\log_{10}$ SPVL among PLWH. (b) Each graph represents sexual relationships that existed in the first, middle, and last 10-year time slice. (c) Simulated egocentric network survey conducted in the population represented in (a). Larger symbols are participants in the survey and the black edges represent their reported relationships in the last 10 years prior to the survey at the end of the 30-year simulation period. (d) In grey, the potential transmission pathways resulting from the introduction of HIV into the simulated population from the index case (framed square). The black links represent the realized HIV transmission pathway. (e) Phylogenetic tree, reconstructed from virus samples. (f) Simulated partner notification investigation starting from the framed circle, that is, the first person to present to a clinic and be diagnosed with HIV in this simulated population. Arrows indicate relationships reported during partner notification interviews. Black arrows connect individuals that were seropositive after testing. PLWH, people living with HIV; SPVL, set-point viral load.

To infer potential HIV transmission pathways from an egocentric study, information about the process of network formation (who forms relationships with whom) first needs to be extracted from the survey data. To do this, an ERGM is fitted to the reported network data to determine the extent and patterns of homophily or concurrency in the population [44]. Next, a set of complete networks is generated on the basis of this model of network formation, and the size and composition of the population [39,46]. Finally, HIV epidemics are simulated on the generated complete networks, by infecting a small number of individuals at random and recording who subsequently gets infected and when. This approach has provided key insights into HIV transmission dynamics [47–50]. Recent advances in ERGM methodology include the development of tools for modelling dynamic networks [51], and freely available software [52,53].

Apart from ERGMs, agent-based models have also been used to infer network characteristics from egocentric data [9,54–58]. In these models, the social network and the subsequent spread of HIV emerge from behavioural and biological 'rules' specified by the analyst at the level of the individual [59].

Naturally, the accuracy of inferences about HIV transmission pathways derived from egocentric data depends on the validity of the model of network formation. If salient aspects of partner choice are not included in this model (either because they are left out by the analyst, or because they were not documented during the survey), then the resulting network inferences may be affected. For example, in a population where PLWH primarily seek to form new partnerships with other PLWH [60–63], ERGMs that do not account for serosorting may misrepresent patterns of exposure to HIV in the population.

Furthermore, important groups of individuals such as mobile and marginalized key population (sex workers, injecting drug users) may be underrepresented in egocentric surveys [64]. Among included individuals, response rates in egocentric surveys are often imperfect and may be associated with HIV transmission risks. For example, PLWH who are aware of their status are significantly less likely to participate in surveys that include HIV testing [65].

Finally, egocentric data provide incomplete and often inaccurate data on the links that connect individuals in a population. To minimize respondent fatigue and recall errors, egocentric surveys often only elicit responses for a small number of relationships per respondent, for example, their three or five most recent relations. The reported personal networks are thus likely truncated. Furthermore, because of recall bias or social desirability bias, survey respondents often omit to report some of their relationships during interviews [66–70] and misreport the characteristics of some of their partners [71,72]. These various forms of missing data affect network inferences from egocentric survey data [73].

Social scientists have also occasionally attempted to map the complete network of a population [5,18,74,75]. In such sociocentric studies (also called 'network censuses') [13], each network member is interviewed and asked to provide the names of their sexual or injection partners. These nominations are then linked to identify the relationships that connect population members [76]. Potential and realized HIV transmission pathways can be observed directly if the study also includes HIV testing [5,74]. Sociocentric studies have been restricted to small, isolated population because of the challenges of delineating network boundaries, and linking survey data to identify relationships [13,77]. In addition, as during an egocentric study, some network members may be absent, decline to be interviewed, or omit to report some of their relations. Sociocentric studies are thus never fully complete.

The molecular biology approach

Whereas social scientists often start from a random sample of individuals irrespective of their HIV serostatus, the molecular biology approach focuses on PLWH. For HIV, as for many other retroviruses, the rate at which viral populations undergo genetic changes within each HIV-positive person (or 'host') is orders of magnitude faster than the rate at which they are transmitted between hosts [78]. These genetic differences between viral populations in different hosts can be used to infer the most likely evolutionary history of the pathogen. The molecular study of networks then entails obtaining viral sequences from PLWH [79], and grouping HIV sequences by genetic similarity [80]. These groupings are depicted as phylogenetic trees (or phylogenies), where tips in the tree represent PLWH, the branching pattern indicates the genetic similarity between sequences from different PLWH, and the internal nodes of the tree represent past transmission events (Fig. 1e). The phylogenetic tree is thus a proxy of a realized HIV transmission pathway.

Phylogenies have been used to identify HIV transmission clusters, that is, groups of PLWH with highly similar viral populations and who are likely connected by an HIV transmission pathway [81–88]. When cluster analysis incorporates the demographic, behavioural, and clinical characteristics of cluster members [89,90], this information may help guide the targeting of HIV prevention and treatment programmes [91]. Viral linkage analysis in HIV clinical trials with serodiscordant couples has enabled more accurate estimation of the efficacy of early antiretroviral therapy (ART) [92] and genital herpes

suppression [93] to prevent HIV transmission. It has also been used to substantiate [94,95] or reject [84,96] alleged (iatrogenic) HIV infection in forensic medicine, and to discern whether new infections originate from within or outside of the intervention communities during cluster-randomized trials [97–101].

The shape of a phylogenetic tree (also called its 'topology') can help elucidate aspects of the broader structure of the social networks and HIV transmission pathways within them. For example, measures of tree asymmetry have been used to test whether individuals form partnerships at random, or rather according to more structured processes. Highly asymmetric trees [102,103], in which some branches led to many infections (Fig. 2c), thus suggest that the underlying social networks may be formed through preferential attachment, leading to scale-free networks [104,105] (with super-spreaders), whereas more balanced trees are associated with random or small-world networks (high interconnectedness among local cliques) [106]. However, different transmission networks

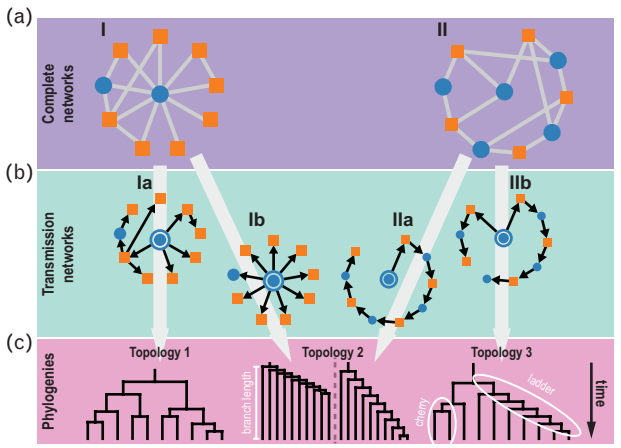

Fig. 2. The complete network influences the realized transmission pathway and subsequent phylogenetic tree, but only to a certain degree. (a) Two examples of differing complete networks. (b) Many different HIV transmission pathways may emerge following the infection of a single individual (framed circle). Individual characteristics such as sex (circles/squares) can restrict the possible transmission networks, but mixed modes of transmission (e.g. heterosexual, MSM, and needle-sharing) further extend the range of possible transmission networks. (c) The transmission networks from (b) leave a trace in the phylogenetic tree topology. The topology of a phylogenetic tree is defined as the structure of the tree with its tips (or 'leaves', i.e. the end points of the tree), without paying attention to branch lengths and left-right ordering of branches and leaves. Topologies can be compared using imbalance (topologies 2 and 3 are less balanced, i.e. more asymmetrical than topology 1), but different transmission networks (Ib and IIa) may result in the same topology (topology 2). In these cases additional information such as branch lengths and the number of cherries and ladders may help differentiate between transmission networks.

can yield phylogenetic trees with similar topologies. Additional summary tree statistics, such as branch lengths, tree width, tree depth, and the occurrence of 'cherries' and 'ladders' (Fig. 2c), are then needed to differentiate between homogeneous, chain-like and super-spreader transmission networks (Fig. 2b) [106–108]. So-called phylodynamic methods provide an alternative approach that avoids having to simulate complete network data. By combining a particular epidemiological model (e.g. a Susceptible-Infected-Removed model) with phylogenetic data, they permit directly estimating properties of transmission pathways such as the fraction of super-spreaders [109,110].

A major advantage of phylogenetic analyses over interview-based methods is that they are not subject to recall or social desirability bias. And unlike egocentric studies, they elicit indirect connections between individuals who may be at two or more degrees of separation. Molecular biology thus allows longer range investigations of the connectivity of networks, including across geographical regions [88,111,112], age and racial/ethnic groups [113], and between subpopulations [114–117].

There are several limitations associated with the molecular biology approach to network inferences, however. First, the evolutionary model for how genetic changes accumulate over time within and between PLWH typically makes several simplifying assumptions. Frequent model assumptions include neutral evolution (no effects of mutations on viral fitness); no selection of viral subpopulations during transmission events; viral evolution driven by mutation only (not by recombination); and no further transmission after HIV diagnosis. Second, a specific phylogenetic tree is only an 'estimate' of the actual true evolutionary history. Reconstruction methods that account for the uncertainty in the tree exist, but are very computationally intensive, thereby limiting the size of datasets that can be analysed. Third, the network inferences drawn from phylogenetic data depend on the sampling density, which is the proportion of PLWH for whom a viral sequence is available [118]. If it is too low, then it will be difficult to link PLWH to the source of their infection, and the phylogenetic tree will be sparser than it really is [79,88]. A recent study in Botswana suggested that a sampling density of 50–70% was required for accurate identification of transmission clusters [111,118]. This is particularly problematic for populations with high HIV prevalence and incidence, where a prohibitively large number of HIV sequences may be needed to achieve an appropriately high sampling density. Fourth, further sampling biases emerge when the individuals for whom sequences are available are not a random sample of the population of PLWH [100,119]. A large number of phylogenetic studies, see [120–122] for recent examples, rely on sequences routinely collected for antiretroviral resistance testing among PLWH who are about to initiate treatment. Phylogenetic trees

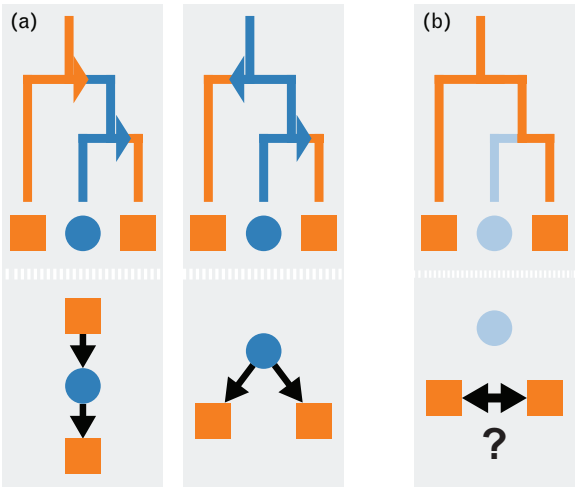

Fig. 3. Phylogenies provide incomplete transmission network data. (a) The topology of the tree alone cannot resolve who infected whom. The topology of the tree represented in the upper panels of Fig. 3a is compatible with multiple directed networks (lower panels of Fig. 3a). (b) Incomplete sampling of PLWH may also result in incorrect identification of transmission pairs in phylogenetic trees. In this example, if the node represented by the circle is not sampled, we may incorrectly infer transmission between the two squares, whereas the transmission actually occurred through a longer chain. PLWH, people living with HIV.

reconstructed from such databases exclude PLWH who have not been diagnosed or have not been linked to care. Fifth, it is not possible to determine who infected whom using only phylogenetic trees (Fig. 3). The basic network inferred by molecular biology is thus undirected (Fig. 1e), which limits its ability to identify individuals who are central in HIV transmission pathways. Determining the temporal connectivity of the transmission network requires either more detailed analyses of the phylogeny [106], or additional data about risk behaviours and the timing of seroconversions [115,123,124]. Finally, phylogenetic methods do not distinguish between transmission events that occurred through sexual intercourse or through sharing of infected injection equipment, thus limiting our ability to disentangle the contributions of different modes of transmission in bridging population that concurrently engage in multiple high-risk practices.

The public health approach

Network data can also be generated when HIV treatment and prevention programmes offer HIV partner notification services to PLWH. In partner notification, newly diagnosed PLWH are asked to provide a list of their recent sexual or injection partners, along with contact details, so that these partners can be traced. Identifiable partners are then contacted and visited in person by a disease notification specialist. They are informed about their potential exposure to HIV, and are invited to visit a health facility for HIV testing and linkage to care, if indicated. The process of contact solicitation, contact tracing and testing is repeated for notified partners who test positive for HIV, but it stops for partners who are HIV-negative. Several electronic platforms have emerged that facilitate this process [125–127]. Although face-to-face partner notification is often preferred by index cases, e-postcards, or mobile text messages are also increasingly used [128,129].

Partner notification significantly increases the number of diagnosed PLWH [130–132] and can help prevent HIV transmission. For partner notification to serve as a network data collection tool, partner notification records must be systematically compiled into datasets. Index cases and their notified partners must be uniquely identified, as one person may be notified by more than one index case in some networks, and then linked. Partner notification then yields network data on realized HIV transmission pathways that connect PLWH, and the links between PLWH and some of their uninfected partners (Fig. 1f). Partner notification thus also locates the entry points of potential HIV transmission pathways through which HIV may spread further within the population. These are crucial targets for preventive interventions such as pre-exposure prophylaxis (PrEP) or male circumcision, but they are not directly identified by other network data collection and modelling approaches.

Partner notification has several other strengths. First, similar to phylogenetic studies, it enables routine collection of network data, integrated within HIV care and treatment programmes. Second, partner notification enables tracing of networks well beyond the narrow geographical boundaries of an administrative unit. In the early days of the AIDS epidemic in the United States, for example, contacts of the so-called 'patient zero' were traced to New York, Los Angeles, and even abroad [133]. Third, partner notification may help uncover subsets of HIV transmission pathways that often remain hidden in egocentric or phylogenetic studies. These include hard-to-reach high-risk groups [134], stigmatized population that rarely attend health facilities and mobile individuals who are less likely to be included in traditional survey sampling frames. Partner notification is more effective at reaching such groups because its network sampling process is 'adaptive' in the sense that it uses information provided directly by network members to guide the selection and recruitment of individuals [43].

Partner notification also has limitations. Whereas partner notification data indicate who notified whom about HIV infection (Fig. 1f), it is difficult to determine who infected whom from such data. In some relations, the person who acquired the infection may in fact seek care – and thus be offered partner notification services – before the transmitter. Additional data such as the date of

seroconversion, the clinical stage of the disease, or phylogenetic data, can help resolve the time sequence of HIV transmissions. The partner notification process can also be highly selective. Not every newly diagnosed PLWH will choose to notify their partners and the decision to use partner notification services may be related to recent risk behaviours. PLWH may deliberately choose not to mention some of their recent partners during partner notification interviews, or may not recall sufficient details about other partners to enable partner notification. And even among those partners who are sought out by the disease notification specialist, some may never be successfully contacted because of insufficient information, whereas others may reject partner notification. It is therefore unclear which parts of the HIV transmission chains the partner notification process reveal [135]. Finally, partner notification services are not always offered during the course of HIV care and treatment. In some regions, partner notification is required by law upon new HIV diagnosis [136], yet eliciting the names of partners of HIV cases may not be authorized in other countries. Laws that criminalize HIV exposure may further limit the willingness of PLWH to share the details of sex and injection partners [137]. In some countries with HIV prevalence, offering partner notification requires significant investments in health personnel, record keeping, and data linkages, which can be prohibitive in low-resource settings.

A combination approach for network analysis in HIV epidemiology

All three approaches described in previous sections shed light on key aspects of the social networks that propagate HIV, but from different angles. In combination, they may help uncover previously unknown features of HIV transmission pathways. We briefly illustrate how an interdisciplinary approach to network epidemiology could help improve our ability to understand processes of network formation and prevent future HIV infections.

Egocentric network data are affected by underreporting of links. This may bias the results of ERGMs or other models that are solely informed by egocentric network data; the inferred networks and HIV transmission pathways will be less dense than the true networks. If underreporting is especially prevalent among specific population groups (e.g. unmarried women in sub-Saharan population) [67,138], then these groups may appear at decreased risk of HIV acquisition, when, in fact, their level of exposure may be substantial. But if HIV sequence data are available for the population of interest, phylogenetic analyses could help adjust for such biases in network models. The topology of transmission clusters, for example, could be used to estimate the mean and variance of the distribution of the numbers of partners individuals have, possibly stratified by gender. These parameters can then be used as correction factors in the network model.

On the other hand, it may also be difficult to predict the potential impact of an HIV prevention intervention based solely on molecular data. Phylogenies only document realized HIV transmission pathways, whereas there may exist a (possibly large) number of competing but unrealized transmission pathways (Fig. 2). For instance, while almost half of HIV transmissions among MSM in the United States occur when the index case is in their first year of HIV infection [139], there may be other potential transmission pathways that involve chronically infected PLWH. Therefore, reducing the transmission potential during the early stages of infection could have a smaller impact on HIV incidence than expected if individuals are connected by more than one potential transmission pathway (Fig. 1d). Such network redundancies are unobserved in the molecular data, but can be recorded directly through sociocentric studies or during partner notification [5,140,141], and can be inferred from egocentric data using ERGMs [48].

As a final example, linking HIV drug resistance data (from HIV sequence analysis) to indicators of ART adherence as well as behavioural, relational, and sociodemographic details of PLWH could help identify subgroups in need of intensified treatment monitoring and support. Superimposing this information onto reconstructed phylogenies and inferred transmission networks may also help predict the evolution of transmitted and secondary drug resistance and identify targetable correlates of residual transmission by PLWH on ART.

Klovdahl *et al.* [142] successfully pioneered integration of multiple network data sources in an epidemiological study of tuberculosis, and an increasing number of studies in HIV epidemiology now also combine multiple sources of data on social networks [88,143–151]. The recently launched Phylogenetics and Networks for Generalized HIV Epidemics in Africa consortium (PANGEA-HIV) aims to sequence 20 000 total HIV genomes from Uganda, Botswana, South Africa, and Zambia, and link sequence data to clinical, demographic, and behavioural data [152].

There are, however, considerable methodological challenges to a combination approach to network analysis in HIV epidemiology. Consider the small population of 10 individuals in Fig. 4, in which all three sources of data described above are available (Fig. 4b–d). In this setting, combining data sources can yield a significantly more complete picture of the social networks, and the location of PLWH within these networks (Fig. 4e). But this still requires a network inference step to identify networks that are compatible with the reported egocentric, partner notification, and phylogenetic data.

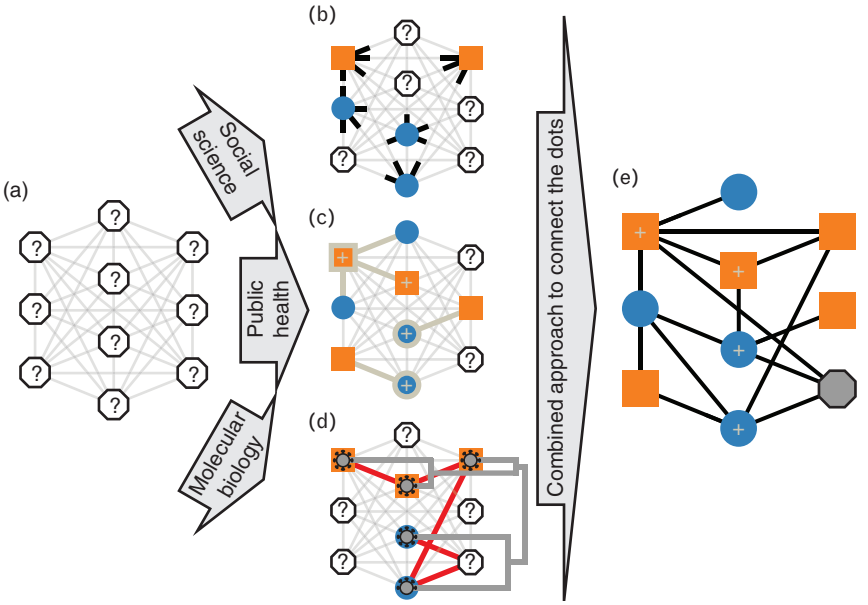

Fig. 4. Approaches from social science, public health and molecular biology can synergistically improve inferences about social networks and HIV transmission pathways. We schematize the case of a population where network data have been collected through the three data collection approaches described above. (a) The starting point is a population whose composition and network connectivity is unknown. The goal is to infer the potential transmission pathways through which HIV can spread within this population. (b) The social science approach conducts egocentric studies, which draw on a random sample of men and women (blue and orange symbols). These respondents report the extent and composition of their personal networks. (c) Public health officials conduct partner notification. Interviews are conducted with PLWH who present at health facilities ('+' in thick borders). Partner notification stops if the partner of a PLWH is HIV-negative or refuses to name their partners. (d) Phylogenetic analysis of HIV sequence data to identify longer range connections between new and previously interviewed individuals, or yet unidentified intermediate transmitters (empty octagon). (e) Ultimately, a more complete network picture emerges from the union of these three data collection approaches, through network modelling. PLWH, people living with HIV.

Unfortunately, network inference using multiple data sources is complicated by a number of factors. First, contrary to egocentric data, partner notification, and phylogenies are not typically based on random samples of the population. As a result, the probabilities of inclusion of each individual and link in the network dataset are unknown, and this may affect the results of ERGMs and other existing models of network inference. Second, in some settings, the various network datasets may not be linked at the individual and/or relationship levels. We would not know whether an individual who reported three links during an egocentric survey is the same person who tested HIV-positive and reported only one partner during partner notification. This may occur, for example, if some data sources are routinely collected during the course of HIV care and treatment (partner notification, phylogenetic data), and are subject to strict data access regulations to preserve the privacy/confidentiality of patient data. Finally, the information provided in the various network datasets may be contradictory, for example, survey respondents may report consistent condom use, but phylogenetic data suggest that they were recently infected with HIV. Unlocking the full potential of a combination approach to network analysis in HIV epidemiology will thus require a number of steps: further progress in the modelling of network data from nonrandom and adaptive samples [43], new tools for efficient linkages between multiple data sources without common identifiers [153–156], and accurate measurements of social links between individuals and the risk exposure accumulated within these relationships.

Conclusion

Data on sexual networks come from an increasingly diverse array of sources, but each of these sources only document parts of the networks through which HIV may spread. Egocentric network surveys suffer from nonresponse, social desirability bias, and the inability to probe beyond the immediate network connections of individuals. Through partner notification services, realized, and potential HIV transmission pathways may be partially revealed, but in resource-poor settings with generalized HIV epidemics offering partner notification services may require prohibitively large investments. Phylogenetic tree analysis permits reconstructing parts of the HIV transmission chains by linking genetically related infections, but to be informative, HIV sequence data must be available for what may be an unfeasibly large sample of PLWH. Novel methods to combine these data sources are beginning to emerge from the collaborative efforts

of experts in computational biology, social science, statistics, public health, and epidemiological modelling. Further advances in network analysis for HIV epidemiology will require important methodological developments in network modelling, as well as a long-term, global commitment from researchers and funding agencies to ensure open access to analytical tools and multifaceted network datasets that include HIV sequences along with behavioural, demographic, clinical, and programmatic information.

Acknowledgements

We thank Jori Liesenborgs, who developed and documented the Simpact Cyan program (C++), as well as the RSimpactCyan front end (R package), and Gavin Hitchcock, Roxanne Beauclair and Niel Hens for their helpful comments on an earlier version of this study. We acknowledge support from the Eunice Kennedy Shriver National institute of Child Health and Human Development, grant # R03HD071122 (PI: Helleringer).

Conflicts of interest

There are no conflicts of interest.

References

1. Morris M, Epstein H, Wawer M. **Timing is everything: international variations in historical sexual partnership concurrency and HIV prevalence.** *PLoS One* 2010; **5**:e14092.
2. Khanna AS, Dimitrov DT, Goodreau SM. **What can mathematical models tell us about the relationship between circular migrations and HIV transmission dynamics?** *Math Biosci Eng* 2014; **11**:1065–1090.
3. Bershteyn A, Klein DJ, Eckhoff PA. **Age-dependent partnering and the HIV transmission chain: a microsimulation analysis.** *J R Soc Interface* 2013; **10**:20130613.
4. McCormick AW, Abuelezam NN, Rhode ER, Hou T, Walensky RP, Pei PP, *et al*. **Development, calibration and performance of an HIV transmission model incorporating natural history and behavioral patterns: application in South Africa.** *PLoS One* 2014; **9**:e98272.
5. Helleringer S, Kohler HP. **Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi.** *AIDS* 2007; **21**:2323–2332.
6. Helleringer S, Kohler HP, Kalilani-Phiri L. **The association of HIV serodiscordance and partnership concurrency in Likoma Island (Malawi).** *AIDS* 2009; **23**:1285–1287.
7. Hontelez JA, Lurie MN, Barnighausen T, Bakker R, Baltussen R, Tanser F, *et al*. **Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study.** *PLoS Med* 2013; **10**:e1001534.
8. Dodd PJ, Garnett GP, Hallett TB. **Examining the promise of HIV elimination by 'test and treat' in hyperendemic settings.** *AIDS* 2010; **24**:729–735.
9. Brookmeyer R, Boren D, Baral SD, Bekker LG, Phaswana-Mafuya N, Beyrer C, *et al*. **Combination HIV prevention among MSM in South Africa: results from agent-based modeling.** *PLoS One* 2014; **9**:e112668.
10. Hoffman IF, Latkin CA, Kukhareva PV, Malov SV, Batluk JV, Shaboltas AV, *et al*. **A peer-educator network HIV prevention intervention among injection drug users: results of a randomized controlled trial in St. Petersburg, Russia.** *AIDS Behav* 2013; **17**:2510–2520.
11. Hurt CB, Beagle S, Leone PA, Sugarbaker A, Pike E, Kuruc J, *et al*. **Investigating a sexual network of black men who have sex with men: implications for transmission and prevention of HIV infection in the United States.** *J Acquir Immune Defic Syndr* 2012; **61**:515–521.
12. Tieu HV, Liu TY, Hussen S, Connor M, Wang L, Buchbinder S, *et al*. **Sexual networks and HIV risk among black men who have sex with men in 6 U.S. cities.** *PLoS One* 2015; **10**:e0134085.
13. Morris M. *Network epidemiology: a handbook for survey design and data collection*. Oxford, UK: Oxford University Press; 2004.
14. Wasserman S, Faust K. *Social network analysis: methods and applications*. Cambridge, NY: Cambridge University Press; 1994.
15. Latkin CA, Mandell W, Vlahov D. **The relationship between risk networks' patterns of crack cocaine and alcohol consumption and HIV-related sexual behaviors among adult injection drug users: a prospective study.** *Drug Alcohol Depend* 1996; **42**:175–181.
16. Friedman SR, Ompad DC, Maslow C, Young R, Case P, Hudson SM, *et al*. **HIV prevalence, risk behaviors, and high-risk sexual and injection networks among young women injectors who have sex with women.** *Am J Public Health* 2003; **93**:902–906.
17. Adams J, Moody J, Morris M. **Sex, drugs, and race: how behaviors differentially contribute to the sexually transmitted infection risk network structure.** *Am J Public Health* 2013; **103**:322–329.
18. Bearman PS, Moody J, Stovel K. **Chains of affection: the structure of adolescent romantic and sexual networks.** *Am J Sociol* 2004; **110**:44–91.
19. Moody J. **The importance of relationship timing for diffusion.** *Social Forces* 2002; **81**:25–56.
20. Morris M. **Sexual networks and HIV.** *AIDS* 1997; **11 (Suppl A)**:S209–216.
21. Kretzschmar M, Morris M. **Measures of concurrency in networks and the spread of infectious disease.** *Math Biosci* 1996; **133**:165–195.
22. Moody J, McFarland D, Bender-deMoll S. **Dynamic network visualization1.** *Am J Sociol* 2005; **110**:1206–1241.
23. Liesenborgs J, Meng F, Hens N, Delva W. Simpact Cyan 0.19.4. http://www.simpact.org/how-simpact-works/ 2015. [Accessed 14 June 2016).
24. Morris M, Handcock MS, Miller WC, Ford CA, Schmitz JL, Hobbs MM, *et al*. **Prevalence of HIV infection among young adults in the United States: results from the Add Health study.** *Am J Public Health* 2006; **96**:1091–1097.
25. Schneider J, Kumar R, Dandona R, Kumar P, Kumar A, Lakshmi V, *et al*. **Social network and risk-taking behavior most associated with rapid HIV testing, circumcision, and preexposure prophylaxis acceptability among high-risk Indian men.** *AIDS Patient Care STDS* 2012; **26**:631–640.
26. Youm Y, Laumann EO. **Social network effects on the transmission of sexually transmitted diseases.** *Sex Transm Dis* 2002; **29**:689–697.
27. Laumann EO, Youm Y. **Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation.** *Sex Transm Dis* 1999; **26**:250–261.
28. Morison L, Weiss HA, Buve A, Carael M, Abega SC, Kaona F, *et al*. **Commercial sex and the spread of HIV in four cities in sub-Saharan Africa.** *AIDS* 2001; **15 (Suppl 4)**:S61–69.
29. Glynn JR, Carael M, Auvert B, Kahindo M, Chege J, Musonda R, *et al*. **Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia.** *AIDS* 2001; **15 (Suppl 4)**:S51–60.
30. Ferry B, Carael M, Buve A, Auvert B, Laourou M, Kanhonou L, *et al*. **Comparison of key parameters of sexual behaviour in four African urban populations with different levels of HIV infection.** *AIDS* 2001; **15 (Suppl 4)**:S41–S50.
31. Lagarde E, Auvert B, Carael M, Laourou M, Ferry B, Akam E, *et al*. **Concurrent sexual partnerships and HIV prevalence in five urban communities of sub-Saharan Africa.** *AIDS* 2001; **15**:877–884.
32. Morris M, Podhisita C, Wawer MJ, Handcock MS. **Bridge populations in the spread of HIV/AIDS in Thailand.** *AIDS* 1996; **10**:1265–1271.

33. Morris M, Pramualratana A, Podhisita C, Wawer MJ. **The relational determinants of condom use with commercial sex partners in Thailand.** *AIDS* 1995; **9**:507–515.
34. Lin H, Ding Y, Liu X, Wu Q, Shen W, He N. **High prevalence of HIV infection and bisexual networks among a sample of men who have sex with men in eastern China.** *PLoS One* 2015; **10**:e0129300.
35. Beauclair R, Hens N, Delva W. **Concurrent partnerships in Cape Town, South Africa: race and sex differences in prevalence and duration of overlap.** *J Int AIDS Soc* 2015; **18**:19372.
36. Chatterjee S, Diaconis P. **Estimating and understanding exponential random graph models.** *Ann Stat* 2013; **41**:2428–2461.
37. Krivitsky PN. **Exponential-family random graph models for valued networks.** *Electron J Stat* 2012; **6**:1100–1128.
38. Lusher D, Koskinen J, Robins G. *Exponential random graph models for social networks: theory, methods, and applications.* New York: Cambridge University Press; 2012.
39. Goodreau SM, Kitts JA, Morris M. **Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks.** *Demography* 2009; **46**:103–125.
40. Goodreau SM. **Advances in exponential random graph (p∗) models applied to a large social network.** *Soc Networks* 2007; **29**:231–248.
41. Robins G, Pattison P, Kalish Y, Lusher D. **An introduction to exponential random graph (p∗) models for social networks.** *Soc Networks* 2007; **29**:173–191.
42. Snijders TA, Pattison PE, Robins GL, Handcock MS. **New specifications for exponential random graph models.** *Sociol Methodol* 2006; **36**:99–153.
43. Handcock MS, Gile KJ. **Modeling social networks from sampled data.** *Ann Appl Stat* 2010; **4**:5–25.
44. Krivitsky PN, Morris M. Inference for social network models from egocentrically-sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. National Institute for Applied Statistics Research Australia, 2015, Working Paper 05-15, http://niasra.uow.edu.au/content/groups/public/@web/@inf/@math/documents/mm/uow190187.pdf.
45. Koehly LM, Goodreau SM, Morris M. **Exponential family models for sampled and census network data.** *Sociol Methodol* 2004; **34**:241–270.
46. Krivitsky PN, Handcock MS, Morris M. **Adjusting for network size and composition effects in exponential-family random graph models.** *Stat Methodol* 2011; **8**:319–339.
47. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Sanchez J, Grinsztejn B, *et al*. **What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)?** *PLoS One* 2012; **7**:e50522.
48. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. **Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice.** *Am J Public Health* 2009; **99**:1023–1031.
49. Goodreau SM, Cassels S, Kasprzyk D, Montano DE, Greek A, Morris M. **Concurrent partnerships, acute infection and HIV epidemic dynamics among young adults in Zimbabwe.** *AIDS Behav* 2012; **16**:312–322.
50. Khanna AS, Goodreau SM, Gorbach PM, Daar E, Little SJ. **Modeling the impact of postdiagnosis behavior change on HIV prevalence in Southern California men who have sex with men (MSM).** *AIDS Behav* 2014; **18**:1523–1531.
51. Krivitsky PN, Handcock MS. **A separable model for dynamic networks.** *J R Stat Soc Series B Stat Methodol* 2014; **76**:29–46.
52. Hunter DR, Goodreau SM, Handcock MS. **ergm.userterms: a template package for extending statnet.** *J Stat Softw* 2013; **52**:i02.
53. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. **Statnet: software tools for the representation, visualization, analysis and simulation of network data.** *J Stat Softw* 2008; **24**:1548–7660.
54. Beck EC, Birkett M, Armbruster B, Mustanski B. **A data-driven simulation of HIV spread among young men who have sex with men: role of age and race mixing and STIs.** *J Acquir Immune Defic Syndr* 2015; **70**:186–194.
55. Huang CY. **An agent-based epidemic simulation of social behaviors affecting HIV transmission among Taiwanese homosexuals.** *Comput Math Methods Med* 2015; **2015**:867264.
56. Richardson L, Grund T. **Modeling the impact of supra-structural network nodes: the case of anonymous syringe sharing and HIV among people who inject drugs.** *Soc Sci Res* 2012; **41**:624–636.
57. Kamp C. **Untangling the interplay between epidemic spread and transmission network dynamics.** *PLoS Comput Biol* 2010; **6**:e1000984.
58. Perrin D, Ruskin HJ, Crane M. **Model refinement through high-performance computing: an agent-based HIV example.** *Immunome Res* 2010; **6 (Suppl 1)**:S3.
59. Abuelezam NN, Rough K, Seage GR 3rd. **Individual-based simulation models of HIV transmission: reporting quality and recommendations.** *PLoS One* 2013; **8**:e75624.
60. Rodriguez-Hart C, Liu H, Nowak RG, Orazulike I, Zorowitz S, Crowell TA, *et al*. **Serosorting and sexual risk for HIV infection at the ego-alter dyadic level: an egocentric sexual network study among MSM in Nigeria.** *AIDS Behav* 2016. Epub ahead of print (24 February) doi:10.1007/s10461-016-1311-3.
61. Reniers G, Helleringer S. **Serosorting and the evaluation of HIV testing and counseling for HIV prevention in generalized epidemics.** *AIDS Behav* 2011; **15**:1–8.
62. Eaton LA, Kalichman SC, O'Connell DA, Karchner WD. **A strategy for selecting sexual partners believed to pose little/no risks for HIV: serosorting and its implications for HIV transmission.** *AIDS Care* 2009; **21**:1279–1288.
63. Rowniak S. **Safe sex fatigue, treatment optimism, and serosorting: new challenges to HIV prevention among men who have sex with men.** *J Assoc Nurses AIDS Care* 2009; **20**:31–38.
64. Brewer DD, Potterat JJ, Garrett SB, Muth SQ, Roberts JM Jr, Kasprzyk D, *et al*. **Prostitution and the sex discrepancy in reported number of sexual partners.** *Proc Natl Acad Sci U S A* 2000; **97**:12385–12388.
65. Reniers G, Eaton J. **Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys.** *AIDS* 2009; **23**:621–629.
66. Helleringer S, Mkandawire J, Kohler HP. **A new approach to measuring partnership concurrency and its association with HIV risk in couples.** *AIDS Behav* 2014; **18**:2291–2301.
67. Helleringer S, Kohler HP, Kalilani-Phiri L, Mkandawire J, Armbruster B. **The reliability of sexual partnership histories: implications for the measurement of partnership concurrency during surveys.** *AIDS* 2011; **25**:503–511.
68. Cleland J, Boerma JT, Carael M, Weir SS. **Monitoring sexual behaviour in general populations: a synthesis of lessons of the past decade.** *Sex Transm Infect* 2004; **80 (Suppl 2)**:ii1–ii7.
69. Hewett PC, Mensch BS, Erulkar AS. **Consistency in the reporting of sexual behaviour by adolescent girls in Kenya: a comparison of interviewing methods.** *Sex Transm Infect* 2004; **80 (Suppl 2)**:ii43–ii48.
70. Mensch BS, Hewett PC, Erulkar AS. **The reporting of sensitive behavior by adolescents: a methodological experiment in Kenya.** *Demography* 2003; **40**:247–268.
71. Helleringer S, Kohler HP, Mkandawire J. **Women underestimate the age of their partners during survey interviews: implications for HIV risk associated with age mixing in northern Malawi.** *Sex Transm Dis* 2011; **38**:1030–1035.
72. Harling G, Tanser F, Mutevedzi T, Barnighausen T. **Assessing the validity of respondents' reports of their partners' ages in a rural South African population-based cohort.** *BMJ Open* 2015; **5**:e005638.
73. Kossinets G. **Effects of missing data in social networks.** *Soc Networks* 2006; **28**:247–268.
74. Rothenberg RB, Woodhouse DE, Potterat JJ, Muth SQ, Darrow WW, Klovdahl AS. **Social networks in disease transmission: the Colorado Springs Study.** *NIDA Res Monogr* 1995; **151**:3–19.
75. Klovdahl AS, Potterat JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW. **Social networks and infectious disease: the Colorado Springs Study.** *Soc Sci Med* 1994; **38**:79–88.
76. Jacobs JR. **Finding words that sound alike: the Soundex algorithm.** *Byte* 1982; **7**:473–474.
77. Helleringer S, Kohler HP, Chimbiri A, Chatonda P, Mkandawire J. **The Likoma Network Study: context, data collection, and initial results.** *Demogr Res* 2009; **21**:427–468.
78. Rambaut A, Posada D, Crandall KA, Holmes EC. **The causes and consequences of HIV evolution.** *Nat Rev Genet* 2004; **5**:52–61.

79. Grabowski MK, Redd AD. **Molecular tools for studying HIV transmission in sexual networks.** *Curr Opin HIV AIDS* 2014; **9**:126–133.
80. Lemey P, Rambaut A, Pybus OG. **HIV evolutionary dynamics within and among hosts.** *AIDS Rev* 2006; **8**:125–140.
81. Holmes EC, Zhang LQ, Robertson P, Cleland A, Harvey E, Simmonds P, *et al*. **The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh.** *J Infect Dis* 1995; **171**:45–53.
82. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J. **Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis.** *Proc Natl Acad Sci U S A* 1996; **93**:10864–10869.
83. Hue S, Pillay D, Clewley JP, Pybus OG. **Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups.** *Proc Natl Acad Sci U S A* 2005; **102**:4425–4429.
84. de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, *et al*. **Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak.** *Nature* 2006; **444**:836–837.
85. Gifford RJ, de Oliveira T, Rambaut A, Pybus OG, Dunn D, Vandamme AM, *et al*. **Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1.** *J Virol* 2007; **81**:13050–13056.
86. Kouyos RD, von Wyl V, Yerly S, Boni J, Taffe P, Shah C, *et al*. **Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland.** *J Infect Dis* 2010; **201**:1488–1497.
87. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. **Episodic sexual transmission of HIV revealed by molecular phylodynamics.** *PLoS Med* 2008; **5**:e50.
88. Grabowski MK, Lessler J, Redd AD, Kagaayi J, Laeyendecker O, Ndyanabo A, *et al*. **The role of viral introductions in sustaining community-based HIV epidemics in rural Uganda: evidence from spatial clustering, phylogenetics, and egocentric transmission models.** *PLoS Med* 2014; **11**:e1001610.
89. Dennis AM, Hue S, Pasquale D, Napravnik S, Sebastian J, Miller WC, *et al*. **HIV transmission patterns among immigrant latinos illuminated by the integration of phylogenetic and migration data.** *AIDS Res Hum Retroviruses* 2015; **31**:973–980.
90. Dennis AM, Hue S, Hurt CB, Napravnik S, Sebastian J, Pillay D, *et al*. **Phylogenetic insights into regional HIV transmission.** *AIDS* 2012; **26**:1813–1822.
91. Little SJ, Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, *et al*. **Using HIV networks to inform real time prevention interventions.** *Plos One* 2014; **9**:e98443.
92. Eshleman SH, Hudelson SE, Redd AD, Wang L, Debes R, Chen YQ, *et al*. **Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial.** *J Infect Dis* 2011; **204**:1918–1926.
93. Campbell MS, Mullins JI, Hughes JP, Celum C, Wong KG, Raugi DN, *et al*. **Viral linkage in HIV-1 seroconverters and their partners in an HIV-1 prevention clinical trial.** *PLoS One* 2011; **6**:e16986.
94. Metzker ML, Mindell DP, Liu XM, Ptak RG, Gibbs RA, Hillis DM. **Molecular evidence of HIV-1 transmission in a criminal case.** *Proc Natl Acad Sci U S A* 2002; **99**:14292–14297.
95. Blanchard A, Ferris S, Chamaret S, Guetard D, Montagnier L. **Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients.** *J Virol* 1998; **72**:4537–4540.
96. Jaffe HW, McCurdy JM, Kalish ML, Liberti T, Metellus G, Bowman BH, *et al*. **Lack of HIV transmission in the practice of a dentist with AIDS.** *Ann Intern Med* 1994; **121**:855–859.
97. Hayes R, Ayles H, Beyers N, Sabapathy K, Floyd S, Shanaube K, *et al*. **HPTN 071 (PopART): rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment: a study protocol for a cluster randomised trial.** *Trials* 2014; **15**:57.
98. Cori A, Ayles H, Beyers N, Schaap A, Floyd S, Sabapathy K, *et al*. **HPTN 071 (PopART): a cluster-randomized trial of the population impact of an HIV combination prevention intervention including universal testing and treatment: mathematical model.** *PLoS One* 2014; **9**:e84511.
99. Novitsky V, Bussmann H, Okui L, Logan A, Moyo S, van Widenfelt E, *et al*. **Estimated age and gender profile of individuals missed by a home-based HIV testing and counselling campaign in a Botswana community.** *J Int AIDS Soc* 2015; **18**:19918.
100. Carnegie NB, Wang R, Novitsky V, De Gruttola V. **Linkage of viral sequences among HIV-infected village residents in Botswana: estimation of linkage rates in the presence of missing data.** *PLoS Comput Biol* 2014; **10**:e1003430.
101. Wertheim JO, Pond SL, Little SJ, De Gruttola V. **Using HIV transmission networks to investigate community effects in HIV prevention trials.** *Plos One* 2011; **6**:e27775.
102. Dearlove BL, Frost SD. **Measuring asymmetry in time-stamped phylogenies.** *PLoS Comput Biol* 2015; **11**:e1004312.
103. Frost SD, Volz EM. **Modelling tree shape and structure in viral phylodynamics.** *Philos Trans R Soc Lond B Biol Sci* 2013; **368**:20120208.
104. Barabasi AL, Bonabeau E. **Scale-free networks.** *Sci Am* 2003; **288**:60–69.
105. Dezso Z, Barabasi AL. **Halting viruses in scale-free networks.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2002; **65**:055103.
106. Leventhal GE, Kouyos R, Stadler T, Wyl V, Yerly S, Boni J, *et al*. **Inferring epidemic contact structure from phylogenetic trees.** *PLoS Comput Biol* 2012; **8**:e1002413.
107. Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. **How the dynamics and structure of sexual contact networks shape pathogen phylogenies.** *PLoS Comput Biol* 2013; **9**:e1003105.
108. Colijn C, Gardy J. **Phylogenetic tree shapes resolve disease transmission patterns.** *Evol Med Public Health* 2014; **2014**:96–108.
109. Frost SD, Volz EM. **Modelling tree shape and structure in viral phylodynamics.** *Philos Trans R Soc Lond B Biol Sci* 2013; **368**:20120208.
110. Stadler T, Bonhoeffer S. **Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods.** *Philos Trans R Soc Lond B Biol Sci* 2013; **368**:20120198.
111. Novitsky V, Bussmann H, Logan A, Moyo S, van Widenfelt E, Okui L, *et al*. **Phylogenetic relatedness of circulating HIV-1C variants in Mochudi, Botswana.** *PLoS One* 2013; **8**:e80589.
112. Lai A, Bozzi G, Franzetti M, Binda F, Simonetti FR, Micheli V, *et al*. **Phylogenetic analysis provides evidence of interactions between Italian heterosexual and South American homosexual males as the main source of national HIV-1 subtype C epidemics.** *J Med Virol* 2014; **86**:729–736.
113. Whiteside YO, Song R, Wertheim JO, Oster AM. **Molecular analysis allows inference into HIV transmission among young men who have sex with men in the United States.** *AIDS* 2015; **29**:2517–2522.
114. Graw F, Leitner T, Ribeiro RM. **Agent-based and phylogenetic analyses reveal how HIV-1 moves between risk groups: injecting drug users sustain the heterosexual epidemic in Latvia.** *Epidemics* 2012; **4**:104–116.
115. Kouyos RD, von Wyl V, Yerly S, Boni J, Taffe P, Shah C, *et al*. **Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland.** *J Infect Dis* 2010; **201**:1488–1497.
116. Bezemer D, Faria NR, Hassan A, Hamers RL, Mutua G, Anzala O, *et al*. **HIV Type 1 transmission networks among men having sex with men and heterosexuals in Kenya.** *AIDS Res Hum Retroviruses* 2014; **30**:118–126.
117. Oster AM, Wertheim JO, Hernandez AL, Ocfemia MC, Saduvala N, Hall HI. **Using molecular HIV surveillance data to understand transmission between subpopulations in the United States.** *J Acquir Immune Defic Syndr* 2015; **70**:444–451.
118. Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. **Impact of sampling density on the extent of HIV clustering.** *AIDS Res Hum Retroviruses* 2014; **30**:1226–1235.
119. O'Dea EB, Wilke CO. **Contact heterogeneity and phylodynamics: how contact networks shape parasite evolutionary trees.** *Interdiscip Perspect Infect Dis* 2011; **2011**:238743.
120. Bezemer D, Cori A, Ratmann O, van Sighem A, Hermanides HS, Dutilh BE, *et al*. **Dispersion of the HIV-1 epidemic in men who have sex with men in the Netherlands: a combined mathematical model and phylogenetic analysis.** *PLoS Med* 2015; **12**:e1001898.
121. Hue S, Brown AE, Ragonnet-Cronin M, Lycett SJ, Dunn DT, Fearnhill E, *et al*. **Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions.** *AIDS* 2014; **28**:1967–1975.

122. Villandre L, Stephens DA, Labbe A, Gunthard HF, Kouyos R, Stadler T, *et al*. **Assessment of overlap of phylogenetic transmission clusters and communities in simple sexual contact networks: applications to HIV-1.** *PLoS One* 2016; **11**:e0148459.
123. Zarrabi N, Prosperi M, Belleman RG, Colafigli M, De Luca A, Sloot PM. **Combining epidemiological and genetic networks signifies the importance of early treatment in HIV-1 transmission.** *PLoS One* 2012; **7**:e46156.
124. Drescher SM, von Wyl V, Yang WL, Boni J, Yerly S, Shah C, *et al*. **Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the Swiss HIV Cohort Study.** *Clin Infect Dis* 2014; **58**:285–294.
125. Plant A, Rotblatt H, Montoya JA, Rudy ET, Kerndt PR. **Evaluation of inSPOTLA.org: an Internet partner notification service.** *Sex Transm Dis* 2012; **39**:341–345.
126. Bourne C, Zablotska I, Williamson A, Calmette Y, Guy R. **Promotion and uptake of a new online partner notification and retesting reminder service for gay men.** *Sex Health* 2012; **9**:360–367.
127. Gotz HM, van Rooijen MS, Vriens P, Op de Coul E, Hamers M, Heijman T, *et al*. **Initial evaluation of use of an online partner notification tool for STI, called 'suggest a test': a cross sectional pilot study.** *Sex Transm Infect* 2014; **90**:195–200.
128. Carnicer-Pont D, Barbera-Gracia MJ, Fernandez-Davila P, Garcia de Olalla P, Munoz R, Jacques-Avino C, *et al*. **Use of new technologies to notify possible contagion of sexually-transmitted infections among men.** *Gac Sanit* 2015; **29**:190–197.
129. Pennise M, Inscho R, Herpin K, Owens J Jr, Bedard BA, Weimer AC, *et al*. **Using smartphone apps in STD interviews to find sexual partners.** *Public Health Rep* 2015; **130**:245–252.
130. Hogben M, McNally T, McPheeters M, Hutchinson AB. **The effectiveness of HIV partner counseling and referral services in increasing identification of HIV-positive individuals a systematic review.** *Am J Prev Med* 2007; **33**:S89–S100.
131. Brown LB, Miller WC, Kamanga G, Nyirenda N, Mmodzi P, Pettifor A, *et al*. **HIV partner notification is effective and feasible in sub-Saharan Africa: opportunities for HIV treatment and prevention.** *J Acquir Immune Defic Syndr* 2011; **56**:437–442.
132. Henley C, Forgwei G, Welty T, Golden M, Adimora A, Shields R, *et al*. **Scale-up and case-finding effectiveness of an HIV partner services program in Cameroon: an innovative HIV prevention intervention for developing countries.** *Sex Transm Dis* 2013; **40**:909–914.
133. Auerbach DM, Darrow WW, Jaffe HW, Curran JW. **Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact.** *Am J Med* 1984; **76**:487–492.
134. Adams J, Moody J, Muth SQ, Morris M. **Quantifying the benefits of link-tracing designs for partnership network studies.** *Field methods* 2012; **24**:175–193.
135. Begley EB, Oster AM, Song B, Lesondak L, Voorhees K, Esquivel M, *et al*. **Incorporating rapid HIV testing into partner counseling and referral services.** *Public Health Rep* 2008; **123 (Suppl 3)**:126–135.
136. Lunny C, Shearer BD. **A systematic review and comparison of HIV contact tracing laws in Canada.** *Health Policy* 2011; **103**:111–123.
137. Lehman JS, Carr MH, Nichol AJ, Ruisanchez A, Knight DW, Langford AE, *et al*. **Prevalence and public health implications of state laws that criminalize potential HIV exposure in the United States.** *AIDS Behav* 2014; **18**:997–1006.
138. Nnko S, Boerma JT, Urassa M, Mwaluko G, Zaba B. **Secretive females or swaggering males? An assessment of the quality of sexual partnership reporting in rural Tanzania.** *Soc Sci Med* 2004; **59**:299–310.
139. Volz EM, Ionides E, Romero-Severson EO, Brandt MG, Mokotoff E, Koopman JS. **HIV-1 transmission during early infection in men who have sex with men: a phylodynamic analysis.** *PLoS Med* 2013; **10**:e1001568.
140. Potterat JJ, Phillips-Plummer L, Muth SQ, Rothenberg RB, Woodhouse DE, Maldonado-Long TS, *et al*. **Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs.** *Sex Transm Infect* 2002; **78 (Suppl 1)**:i159–i163.
141. Potterat JJ, Muth SQ, Rothenberg RB, Zimmerman-Rogers H, Green DL, Taylor JE, *et al*. **Sexual network structure as an indicator of epidemic phase.** *Sex Transm Infect* 2002; **78 (Suppl 1)**:i152–i158.
142. Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, *et al*. **Networks and tuberculosis: an undetected community outbreak involving public places.** *Soc Sci Med* 2001; **52**:681–694.
143. Chan PA, Hogan JW, Huang A, DeLong A, Salemi M, Mayer KH, *et al*. **Phylogenetic investigation of a statewide HIV-1 epidemic reveals ongoing and active transmission networks among men who have sex with men.** *J Acquir Immune Defic Syndr* 2015; **70**:428–435.
144. Poon AF, Joy JB, Woods CK, Shurgold S, Colley G, Brumme CJ, *et al*. **The impact of clinical, demographic and risk factors on rates of HIV transmission: a population-based phylogenetic analysis in British Columbia, Canada.** *J Infect Dis* 2015; **211**:926–935.
145. Kharsany AB, Buthelezi TJ, Frohlich JA, Yende-Zuma N, Samsunder N, Mahlase G, *et al*. **HIV infection in high school students in rural South Africa: role of transmissions among students.** *AIDS Res Hum Retroviruses* 2014; **30**:956–965.
146. Avila D, Keiser O, Egger M, Kouyos R, Boni J, Yerly S, *et al*. **Social meets molecular: combining phylogenetic and latent class analyses to understand HIV-1 transmission in Switzerland.** *Am J Epidemiol* 2014; **179**:1514–1525.
147. Middelkoop K, Rademeyer C, Brown BB, Cashmore TJ, Marais JC, Scheibe AP, *et al*. **Epidemiology of HIV-1 subtypes among men who have sex with men in Cape Town, South Africa.** *J Acquir Immune Defic Syndr* 2014; **65**:473–480.
148. Robertson AM, Garfein RS, Wagner KD, Mehta SR, Magis-Rodriguez C, Cuevas-Mota J, *et al*. **Evaluating the impact of Mexico's drug policy reforms on people who inject drugs in Tijuana, B.C., Mexico, and San Diego, CA, United States: a binational mixed methods research agenda.** *Harm Reduct J* 2014; **11**:4.
149. Volz EM, Frost SD. **Inferring the source of transmission with phylogenetic data.** *PLoS Comput Biol* 2013; **9**:e1003397.
150. Lin H, He N, Zhou S, Ding Y, Qiu D, Zhang T, *et al*. **Behavioral and molecular tracing of risky sexual contacts in a sample of Chinese HIV-infected men who have sex with men.** *Am J Epidemiol* 2013; **177**:343–350.
151. Levy I, Mor Z, Anis E, Maayan S, Leshem E, Pollack S, *et al*. **Men who have sex with men, risk behavior, and HIV infection: integrative analysis of clinical, epidemiological, and laboratory databases.** *Clin Infect Dis* 2011; **52**:1363–1370.
152. Pillay D, Herbeck J, Cohen MS, de Oliveira T, Fraser C, Ratmann O, *et al*. **PANGEA-HIV: phylogenetics for generalised epidemics in Africa.** *Lancet Infect Dis* 2015; **15**:259–261.
153. Liepe J, Kirk P, Filippi S, Toni T, Barnes CP, Stumpf MP. **A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation.** *Nat Protoc* 2014; **9**:439–456.
154. Andrianakis I, Vernon IR, McCreesh N, McKinley TJ, Oakley JE, Nsubuga RN, *et al*. **Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on HIV in Uganda.** *PLoS Comput Biol* 2015; **11**:e1003968.
155. Willem L, Stijven S, Vladislavleva E, Broeckhove J, Beutels P, Hens N. **Active learning to understand infectious disease models and improve policy making.** *PLoS Comput Biol* 2014; **10**:e1003563.
156. Ratmann O, Donker G, Meijer A, Fraser C, Koelle K. **Phylodynamic inference and model assessment with approximate bayesian computation: influenza as a case study.** *PLoS Comput Biol* 2012; **8**:e1002835.