Epidemics 10 (2015) 72–77

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Six challenges in measuring contact networks for use in modelling

K. Eames [a,*], S. Bansal [b,c], S. Frost [d], S. Riley [e]

[a] Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[b] Department of Biology, Georgetown University, Washington, DC 20057, USA
[c] Fogarty International Center, National Institutes of Health, Bethesda, MD 20893, USA
[d] Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK
[e] MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, Norfolk Place, London W21PG, UK

ARTICLE INFO

Article history:
Received 24 February 2014
Received in revised form 21 August 2014
Accepted 22 August 2014
Available online 30 August 2014

Keywords:
Networks
Contact patterns
Disease transmission
Mathematical modelling
Network measurement
Model-driven data collection

ABSTRACT

Contact networks are playing an increasingly important role in epidemiology. A contact network represents individuals in a host population as nodes and the interactions among them that may lead to the transmission of infection as edges. New avenues for data collection in recent years have afforded us the opportunity to collect individual- and population-scale information to empirically describe the patterns of contact within host populations. Here, we present some of the current challenges in measuring empirical contact networks. We address fundamental questions such as defining contact; measurement of non-trivial contact properties; practical issues of bounding measurement of contact networks in space, time and scope; exploiting proxy information about contacts; dealing with missing data. Finally, we consider the privacy and ethical issues surrounding the collection of contact network data.

Introduction

Early mathematical models of infectious disease dynamics treated all individuals as identical, and assumed that they all interacted with each other at the same constant rate. More recent models customarily split the population into groups distinguished by characteristics such as age and location, introducing mixing rates defining interactions between groups (Mossong et al., 2008). Such models still assume that all individuals within a group are identical, and that the interaction between two given individuals is determined solely by the groups to which they belong.

In reality, social interactions are more nuanced and structured than such assumptions allow. Each person has an individual set of contacts that determine whom she may be infected by and whom she may infect. These contacts can be described by a network: a set of links between members of a population. Each link represents a (pathogen-dependent) opportunity for transmission.

There is a long history of the use of networks in epidemiology, in particular associated with contact tracing and outbreak investigations, which seek to identify risky interactions within a population (Klovdahl et al., 1994; Riley and Ferguson, 2006; Fraser et al., 2004). Likewise, there has been a great deal of recent work in developing models of transmission through networks (Keeling and Eames, 2005); challenges associated with such modelling are discussed elsewhere in this series (Pellis et al., in this issue).

There are many different types of network in epidemiology: for example, we can consider social or sexual contacts between individuals; patient movement between hospitals; airline travel between cities. In each case, the nodes represent relevant epidemiological units (individuals, hospitals, cities) and the links describe connections between nodes that could facilitate transmission. Here our focus is predominantly on the measurement of links between individual people, but many of the challenges below apply more generally; we note that recent progress has been made in using networks for understanding animal epidemiology, prompted in part by large-scale measurement of livestock movements (Brooks-Pollock et al., 2014).

In models that consider subgroups rather than individuals, rather than requiring the strength of contact between *individuals*, we need to know about contacts between *groups*. This group-level information is often collected through studies carried out at the individual level (Mossong et al., 2008; Read et al., 2012; Eames et al., 2012); thus the data-collection challenges associated with models that contain any sort of contact structure are related to those inherent in network models.

* Corresponding author. Tel.: +44 02079272469.
E-mail addresses: ken.eames@lshtm.ac.uk (K. Eames), shweta@sbansal.com (S. Bansal), sdf22@cam.ac.uk (S. Frost), s.riley@imperial.ac.uk (S. Riley).

http://dx.doi.org/10.1016/j.epidem.2014.08.006

With perfect knowledge, any outbreak of a directly transmitted infectious disease could be described by linking each infection to its infector. This *transmission tree* would show the course of infection through a population (Gardy et al., 2011). Such networks provide a natural way of visualising and conceptualising infection processes, and contain much information about those types of interactions that result in transmission (Cauchemez et al., 2011).

A transmission tree is a real and – in theory – measurable entity for pathogens that do not reproduce outside their host(s). However, the contact network – the network over which transmission *might* occur – is a more challenging theoretical concept. To make the fullest use of network methods, we require not merely the transmission tree of one outbreak of a pathogen in a population, but a network that contains all contacts relevant for transmission, whether or not they have been involved in transmission during a particular outbreak. Ideally, we would measure not only the presence/absence of a contact, but additional properties such as its strength (weight), duration, and when it occurred. Although conceptually straightforward, the challenge of obtaining information about the weights and dynamics of interactions between all possible pairs of individuals at all possible times is vast. Approximations, such as assuming that all weights are constant, are often made.

While some studies have sought to measure epidemiologically relevant networks in populations of interest (Klovdahl et al., 1994; Cauchemez et al., 2011; Conlan et al., 2011; Bearman et al., 2004; Salathé et al., 2010; Isella et al., 2011), the number of such studies is small. Below, we offer a set of challenges in collecting contact network data relevant to dynamic transmission modelling. It is not intended to be a complete list, and is biased by the preferences and interests of the authors. We hope that it will contribute to seeding conversations, research projects, and healthy disagreements.

1. Defining a contact

Our ability to define a potentially infectious contact depends on our knowledge about the dominant mode of transmission. For some infections, e.g. sexually transmitted and vector-born diseases, the relevant contacts may be difficult to measure but are clearly defined, and we can propose empirical studies to refute or confirm the existence of specific network structures (Lewis et al., 2008). However, the infection event is harder to define for respiratory pathogens, where it is not always clear precisely how infection passes from person to person. The infector and infectee must be in the same physical space within a short period of time, but it is difficult to be more precise. For example, if modelling transmission on public transport, can anyone on a bus be infected by anyone else or only those "nearby"? Decisions about what types of interaction matter are crucial when setting up network studies.

Proxy measures of contacts

We are often obliged to work with plausible proxy measures of contacts. Self-reported face-to-face conversations and skin-on-skin contact are frequently used as proxy measures of potentially infectious contacts (Mossong et al., 2008; Read et al., 2008). Age-specific mixing patterns from questionnaire studies have been highly influential in parameterising models of respiratory infection (Mossong et al., 2008), despite potential problems of inaccurate reporting and recall bias. The key challenge for the use of self-reported contact data to inform network models is to validate the relationship between reported contacts and infection. Modelling work has used different measured contact patterns to fit age-structured incidence or serology data (Goeyvaerts et al., 2010; Melegaro et al., 2011), but further work is needed to understand how to interpret the results. For example, if patterns of interactions involving physical contact provide the best fit to serological sampling, does this mean that infection actually spreads via physical contact, or just that such contacts provide a good proxy in a particular population? Extending such studies to multiple populations and multiple pathogens may shed further light on this issue.

Integrating genetic data

Genetic data potentially allow the full description of the infection tree (Gardy et al., 2011). The combination of self-reported social contacts and an accurate infection tree should permit much more accurate assessments of the relative importance of different routes of transmission, resulting in better predictive models of infection events. A necessary step for making genetic approaches useful is the collection of both genetic and detailed traditional "contact" information in the same study. A key issue is the completeness of the data collected: if only a small fraction of infections and/or a small fraction of relevant contacts are sampled, then it will be difficult to reconstruct infection trees or to draw conclusions about networks (Volz and Frost, 2013). Complete sampling is made even more difficult in cases of asymptomatic infection or when our understanding of what constitutes a relevant contact is incomplete (Resik et al., 2007).

Counterfactual contact data

Many studies do not measure contacts that have actually led to transmission; rather, they measure contacts that could potentially lead to transmission. However, there is no guarantee that individuals would behave in the same way when infectious (or when interacting with infectious individuals) (Van Kerckhove et al., 2013); such counterfactual scenarios are inherently unmeasurable. However, large-scale studies that quantify links made during infectiousness would add greatly to our ability to select the right mapping between "healthy" and "ill" contact patterns.

2. Bounding networks in space, time, and scope

An epidemiologically relevant network could, in theory, include practically everyone in the world. Although sophisticated mathematical models may include the population of the entire world (van den Broeck et al., 2011), we are unlikely to attempt to measure this network. Therefore in any study we must choose where to bound our network. The decision will depend on available time and resources, and on our understanding of what constitutes a relevant study community.

Permeable boundaries

Almost all network studies are constrained to be within a particular pre-defined study population, e.g. a school or hospital (Conlan et al., 2011; Salathé et al., 2010; Isella et al., 2011). However, it is only rarely – if ever – that there are no relevant contacts with individuals outside the study population. How much does it matter that we miss these "external" connections? In particular, how do we deal with the seeding of infection into our population without information about external contacts?

Time horizons

Studies may provide snapshots of contact networks, but the dynamic nature of interactions means that we expect networks to change over time (Bansal et al., 2010) (see Challenge 4, below). How can we best use networks collected over short time windows

to understand the spread of epidemics over periods of months or years?

Pre-defined contacts

One approach to carrying out a network survey involves providing participants with a pre-defined list of possible contacts. This offers straightforward completion and data management, but is impractical for large populations, and requires knowing in advance who will appear in the measured network, precluding the identification of additional contacts. Ideally, we would test different methods of measuring the same network, but such comparisons are challenging because of their non-independence if completed by the same individuals, and because of their non-comparability if completed by different individuals or at different times.

Setting the scale

Nodes need not represent individuals: they may be households, towns, or countries. Contact data can be defined at multiple scales, for example the movement of animals between farms (Brooks-Pollock et al., 2014), or the movement of people between cities (Colizza et al., 2006). To what extent is measuring contact/movement at coarse scales whilst ignoring the finer details of the network worthwhile? The answer will depend on the use to which data are to be put, and the details of the population and pathogen considered. Since detailed data about networks of contacts between individuals in a large population or a large spatial scale are generally absent, this challenge provides an opportunity for useful interplay between models and data collection, with modelling able to test a range of plausible scenarios and guide future contact studies (Pellis et al., in this issue; Riley et al., in this issue).

3. Dealing with missing data

When contact data are used to parameterise standard mathematical models of disease transmission it generally makes little difference if we fail to measure a few links. However, when using a network to predict disease spread, missing links can have a huge impact (Watts and Strogatz, 1998). Having determined the scope of our study (see Challenge 2), if it misses some links, can we meaningfully use it to parameterise a network model of disease spread? This question cuts to the heart of network studies and network models: if the answer is "no", then it may be the case that a network model, perversely, is a less wise use of network data than a simpler subgroup model.

Recall and identification

Whether presented with a list of possible contacts, or asked to name those with whom they interact, traditional network studies demand that participants know the names of their contacts. Participants may be unwilling or unable to identify all their contacts. Studies have shown that a substantial fraction of reported links within networks are not reciprocated (Read et al., 2008; Smieszek et al., 2012). The missing information is not a random sample of the links, with "stronger" connections more likely to be recalled and reported accurately. Can we afford to ignore unreported links, on the assumption that they are epidemiologically insignificant?

Using ego data

Measuring a whole network – for the entire population that might be affected by an outbreak, for example – is almost certainly impossible in practice (though, in contrast to the counterfactual scenarios of Challenge 1, possible in theory). However, we may be able to measure the personal networks of contacts reported by a set of individuals – egonets – perhaps even representatively. What is the best we can do with these egonets? Do higher-order network properties that egonets do not give information about – such as path length and centrality – significantly affect epidemic behaviour, or can everything beyond number of contacts (degree) and clustering be safely ignored?

Social network size

Perhaps the simplest social network question that one can ask is 'how many people do you know'? Greater variation between individuals in their number of contacts has been proposed as a measure of clustering in the underlying network (Zheng et al., 2006), which shows some promise as a means of obtaining social network information (Drumright and Frost, 2010). However, like all self-reported information, it is subject to numerous errors, and it remains to be seen whether additional information, for example on the psychology of estimation, could be used to improve accuracy.

Contact tracing data

Contact tracing is a valuable source of network information, and has potential to be tied in with genetic studies (see above). However, there is a degree of circularity in using contact tracing data in network models: having decided a priori which contacts to trace, we can only detect transmissions resulting from those contacts. To what extent can we use contact tracing data to parameterise network models? Until we are able to test biological outcome data against reported contacts, we cannot know what fraction of infections contact tracing captures; genetic sequencing of pathogens from potential infector–infectee pairs could help interpret contact tracing data.

4. Measuring weighted and dynamic networks

We do not expect networks to remain constant over time. Most studies provide at best a snapshot of a network, and measure links as present/absent, so most models assume that links are fixed and unweighted. How do we handle situations when links within a network have their own properties and dynamics?

Weighted networks

Are measurable weights (e.g. duration or frequency of interaction; presence or absence of physical contact) well-correlated with risk? Studies indicate that some measures of link properties generate better model fits to incidence data than others (Goeyvaerts et al., 2010) (see Challenge 2, above), but can we conclude that these measures of weight are more generally suitable? If we can robustly quantify link strength using available data for a pathogen of interest, we will be able to make network studies substantially more effective.

Dynamic networks

Can we make progress on dynamic networks when we often measure little more than binary snapshots? If we ask about the dynamics of particular links (e.g. frequency or duration of a link), do unmeasured correlations matter? Electronic data collection gives scope to delve in much more detail into the fine scale dynamics of links, but is limited by boundary issues – further exacerbated by the fact that the members of the network that we would like to measure will change over time. Detailed measurement of dynamic links would allow us to understand how contact networks evolve in

response to extrinsic and disease-related processes, and how these topological changes influence the spread and evolution of infection.

Stability of superspreading

Superspreaders, individuals with high connectivity, increase the heterogeneity in a network. How can we distinguish between heterogeneous individuals and a heterogeneous population? Are the superspreaders of today also the superspreaders of tomorrow? What are the time-scales of individual-level heterogeneity? Similarly, should superspreaders be defined just by their degree or by their link weight (i.e. weighted degree). This has potential ramifications for targeted interventions (Eames et al., 2009).

Changes in response to infection

Individuals respond to infection through changes such as avoidance of infected individuals or a change in daily activities due to illness (Van Kerckhove et al., 2013). Are these changes predictable? Do the changes lead to a change in first-order network properties (e.g. degree) or does global structure also change (e.g. a reduction in long-range travel)? This has consequences for use of "non-infection" network studies in predicting infection dynamics; if we knew which links were likely to be retained during illness we would be able to make far more confident use of available network data.

Generalisation of networks

Network structure varies from place to place. Are network properties consistent between comparable places? What does a UK social network from 10 years ago tell us about infection in Poland today? What characteristics of a place should be similar to consider generalisation of network data? This has consequences for decisions about where to carry out network studies, particularly with reference to hard-to-access populations.

5. Exploiting indirect information about networks

Although it may be costly to collect detailed network information, the method of sample collection itself may contain indirect information pertinent to contacts. Even if indirect measures of the contact network itself may be unavailable, such information may inform mechanistic models of network formation.

Respondent driven sampling (RDS)

Many studies, particularly in the context of HIV surveillance, have employed RDS, in which individuals invite their friends and acquaintances to participate in the study. This generates a 'tree' of who recruited whom, embedded in the underlying social network (Poon et al., 2009; Wejnert, 2010). As individuals are often asked about their relationship with their recruiter, information on the overlap between social and sexual networks may be obtained. However, RDS may generate a biased sample of the underlying network, as 'seed' individuals are often sampled at convenience, and complex mechanisms may be in play when participants choose from amongst their contacts and when individuals decide whether to join the study. It remains a challenge to determine whether RDS can provide accurate information on the underlying network.

Time-location sampling (TLS)

TLS is often used when the target population congregates in some physical (e.g. bars and clubs) or virtual (e.g. an internet chatroom) location. As locations may structure social and sexual networks, data on these locations, while unlikely to offer detailed insights, may help to rule out particular network structures and generate more representative estimates of population composition (Karon and Wejnert, 2012). Data on an individual's "affiliations" to particular places have yet to be fully exploited (Frost, 2007), in part due to the lack of methods that can be applied to these sources of data. Like RDS, TLS also generates a biased sample of the network. While only a fraction of venues may be sampled, it may be possible to get indirect insights into other venues through reports of individuals who visit multiple venues.

Electronic social data

The growth of internet- and phone-derived social data offers new opportunities for the inference of social networks. Location data are captured for many people when they phone, browse, and tweet. In theory, it may be possible to link genuine social interactions and co-location. However, as with other novel ways of capturing networks of potentially infectious contacts, the key challenge remains the validation of the inferred networks. Common pathogens such as influenza (Broniatowski et al., 2013) that infect all ages offer an attractive system for the robust comparison of models based on these exciting new data with biological outcome data.

Mechanistic models of networks

Rather than try to measure the network itself, either directly or indirectly, it may be possible to build mechanistic models of interactions, through which the structure of the contact network emerges. For example, Eubank et al. (2004) used dynamic bipartite graphs to model the contact patterns that result from movements of individuals between specific locations, generated by agent based models parameterised with census, land-use and population-mobility data. It remains a challenge to verify such approaches against direct measurements of interactions, and to determine what scale of network data – whether on a group level or an individual level, for example – such methods can reliably generate.

6. Personal information and ethical concerns

Unlike most contact surveys, network studies are necessarily non-anonymous. Will ethical considerations restrict network studies? Although social media encourage us to share social information, can we continue to persuade participants that researchers can be trusted not to misuse personal information?

Sensitive information

Certain data are sensitive. Questions about sexual mixing behaviour are inherently personal, and questions about physical contacts may be considered intrusive in some communities. Researchers must engage sensitively with study participants to ask reasonable questions in reasonable ways. There is little point in asking the perfect scientific question if participants feel unable to answer it honestly. Electronic self-administered questionnaires (Prah et al., 2013) are a practical way of enabling participants to preserve their privacy.

Acceptability of automated data collection

Radio tagging has been used to collect network data in bounded settings including schools (Salathé et al., 2010) and hospitals (Isella et al., 2011), allowing contacts to be recorded anonymously. Where such studies have been successful, they have achieved a high level

of "buy in" within a specific community. It remains an open challenge to conduct and interpret such a study in a general population setting, where community engagement will be harder and uptake lower.

Second-hand information

It is generally accepted that study participants may choose to disclose information about themselves. However, in a non-anonymous study, is it appropriate for person A to disclose information about person B without B's consent? Much of the information requested in network studies would not be deemed especially sensitive, but this may not always be the case. Researchers should engage with medical ethicists and the public to help determine where should we draw the line.

Concluding remarks

Network models require a huge amount of detailed information. Ideally we would to measure everything – contacts, illness, and transmission trees – using all plausible methods; we would then seek a "minimal" set of data necessary to understand and explain epidemic behaviour.

However, this is impossible. Much that we wish to measure is unmeasurable (either in theory or in practice), and in many cases we do not know what we ought to be measuring. Even if it were possible, we would need to repeat the exercise for other populations, pathogens, and times. In practice, as with all modelling, we must make assumptions and simplifications based, to a certain extent, on intuition and hope.

Modelling has come a long way since the days of homogeneous mass action models, in part driven by better data and a better understanding of the role that contact patterns play in transmission. Continued advances will be enormously valuable in predicting the spread of infection and guiding interventions. For example, the impact of targeting "high-risk" groups depends on how strongly those groups interact with the rest of the population; the impact of social distancing and hygiene-based interventions depends on which contacts and which properties of contacts contribute to transmission. Thus, the more we know about epidemiologically relevant networks, the better use we can make of our resources.

Recently the ability to measure aspects of contact networks has increased dramatically. Large-scale contact studies have been carried out over the internet (Eames et al., 2012), or using electronic sensors (Salathé et al., 2010; Isella et al., 2011), and collection of complete cattle-movement data is becoming commonplace (Brooks-Pollock et al., 2014). The genetic sequencing revolution may help to determine more precisely which aspects of interactions are important for transmission, allowing us to refine our questions. Mathematical modelling of transmission within networks (Pellis et al., in this issue) has a key role to play in providing guidance about which details of networks and behaviour are most relevant to infection risk at the individual and population level; there is scope for beneficial collaboration between modelling and data collection, especially at the study design stage, to help network measurement studies focus on the most epidemiologically important aspects of networks. Although considerable challenges remain, there is every likelihood that network-based approaches will remain a vital method for the understanding and control of infectious diseases.

Acknowledgments

All authors are grateful to The Isaac Newton Institute. KE is supported by a Career Development Fellowship award from the National Institute for Health Research (grant NIHR-CDF-2001-04-019). SB is supported by the RAPIDD Program of the Science & Technology Directorate, Department of Homland Security, and the Fogarty International Center, National Institutes of Health, USA. SF is supported in part by the Royal Society and by the ESRC (grant no ES/J011266/1). SR is supported by Wellcome Trust Project Award 093488/Z/10/Z, R01TW008246-01 from Fogarty International Centre, The Medical Research Council (UK, Project Grant MR/J008761/1) and the RAPIDD program from Fogarty International Centre with the Science & Technology Directorate, Department of Homeland Security.

References

Bansal, S., Read, J., Pourbohloul, B., Meyers, L.A., 2010. The dynamic nature of contact networks in infectious disease epidemiology. J. Biol. Dyn. 4, 478–489.
Bearman, P.S., Moody, J., Stovel, K., 2004. Chains of affection: the structure of adolescent romantic and sexual networks. Am. J. Sociol. 110, 44–91.
Broniatowski, D.A., Paul, M.J., Dredze, M., 2013. National and local influenza surveillance through twitter: an analysis of the 2012–2013 influenza epidemic. PLOS ONE 8, e83672.
Brooks-Pollock, E., Roberts, G.O., Keeling, M.J., 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature 511, 228–231.
Cauchemez, S., Bhattarai, A., Marchbanks, T.L., Fagan, R.P., Ostroff, S., Ferguson, N.M., Swerdlow, D., Pennsylvania H1N1 working group, 2011. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proc. Natl. Acad. Sci. U. S. A. 108, 2825–2830.
Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A., 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. Proc. Natl. Acad. Sci. U. S. A. 103, 2015–2020.
Conlan, A.J.K., Eames, K.T.D., Gage, J.A., von Kirchbach, J.C., Ross, J.V., Saenz, R.A., Gog, J.R., 2011. Measuring social networks in British primary schools through scientific engagement. Proc. R. Soc. Lond. B 278, 1467–1475.
Drumright, L.N., Frost, S.D., 2010. Rapid social network assessment for predicting HIV and STI risk among men attending bars and clubs in San Diego, California. Sex. Transm. Infect. 86 (Suppl. 3), iii17–iii23.
Eames, K.T.D., Read, J.M., Edmunds, W.J., 2009. Epidemic prediction and control in weighted networks. Epidemics 1, 70–76.
Eames, K.T.D., Tilston, N.L., Brooks-Pollock, E., Edmunds, W.J., 2012. Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS Comp. Biol. 8, e1002425.
Eubank, S., Guclu, H., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184.
Fraser, C., Riley, S., Anderson, R.M., Ferguson, N.M., 2004. Factors that make an infectious disease outbreak controllable. Proc. Natl. Acad. Sci. U. S. A. 101, 6146–6151.
Frost, S.D., 2007. Using sexual affiliation networks to describe the sexual structure of a population. Sex. Transm. Infect. 83 (Suppl. 1), i37–i42.
Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M.K., Elwood, K., Jones, S.J., Brinkman, F.S., Brunham, R.C., Tang, P., 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N. Engl. J. Med. 364, 730–739.
Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., van Damme, P., Beutels, P., 2010. Estimating infectious disease parameters from data on social contacts and serological status. Appl. Stat. 59, 255–277.
Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., Gesualdo, F., Pandolfi, E., Ravà, L., Rizzo, C., Tozzi, A.E., 2011. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS ONE 6, e17144.
Karon, J.M., Wejnert, C., 2012. Statistical methods for the analysis of time-location sampling data. J. Urban Health 89, 565–586.
Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. J. R. Soc. Interface 2, 295–307.
Klovdahl, A.S., Potterat, J.J., Woodhouse, D.E., Muth, J.B., Muth, S.Q., Darrow, W.W., 1994. Social networks and infectious disease: the Colorado Springs Study. Soc. Sci. Med. 38, 79–88.
Lewis, F., Hughes, G.J., Rambaut, A., Pozniak, A., Leigh Brown, A.J., 2008. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med. 5, e50.
Melegaro, A., Jit, M., Gay, M., Zagheni, E., Edmunds, W.J., 2011. What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. Epidemics 3, 143–151.
Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Scalia-Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5, e74.
Poon, A.F., Brouwer, K.C., Strathdee, S.A., Firestone-Cruz, M., Lozada, R.M., Pond, S.L., Heckathorn, D.D., Frost, S.D., 2009. Parsing social network survey data from hidden populations using stochastic context-free grammars. PLoS ONE 4, e6777.
Prah, P., Copas, A.J., Mercer, C.H., Clifton, S., Erens, B., Phelps, A., Tanton, C., Sonnenberg, P., Macdowall, W., Wellings, K., Johnson, A.M., 2013. Consistency in reporting sensitive sexual behaviours in Britain: change

in reporting bias in the second and third National Surveys of Sexual Attitudes and Lifestyles (Natsal-2 and Natsal-3). Sex. Transm. Infect., http://dx.doi.org/10.1136/sextrans-2013-051360.

Read, J.M., Eames, K.T.D., Edmunds, W.J., 2008. Dynamic social networks and the implications for the spread of infectious disease. J. R. Soc. Interface 5, 1001–1007.

Read, J.M., Edmunds, W.J., Riley, S., Lessler, J., Cummings, D.A., 2012. Close encounters of the infectious kind: methods to measure social mixing behaviour. Epidemiol. Infect. 140, 2117–2130.

Resik, S., Lemey, P., Ping, L.H., Kouri, V., Joanes, J., Perez, J., Vandamme, A.M., Swanstrom, R., 2007. Limitations to contact tracing and phylogenetic analysis in establishing HIV type 1 transmission networks in Cuba. AIDS Res. Hum. Retrovir. 23, 347–356.

Riley, S., Ferguson, N.M., 2006. Smallpox transmission and control: spatial dynamics in Great Britain. Proc. Natl. Acad. Sci. U. S. A. 103, 12637–12642.

Salathé, M., Kazandjieva, M., Lee, J.W., Levis, P., Feldman, M.W., Jones, J.H., 2010. A high-resolution human contact network for infectious disease transmission. Proc. Natl. Acad. Sci. U. S. A. 107, 22020–22025.

Smieszek, T., Burri, E.U., Scherzinger, R., Scholz, R.W., 2012. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. Epidemiol. Infect. 140, 744–752.

van den Broeck, W., Gioannini, C., Goncalves, B., Quaggiotto, M., Colizza, V., Vespignani, A., 2011. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. BMC Inf. Dis. 11, 37.

Van Kerckhove, K., Hens, N., Edmunds, W.J., Eames, K.T.D., 2013. The impact of illness on social networks: implications for transmission and control of influenza. Am. J. Epidemiol. 178, 1655–1662.

Volz, E.M., Frost, S.D.W., 2013. Inferring the source of transmission with phylogenetic data. PLoS Comp. Biol., http://dx.doi.org/10.1371/journal.pcbi.1003397.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

Wejnert, C., 2010. Social network analysis with respondent-driven sample data: a study of racial integration on campus. Soc. Netw. 32, 112–124.

Zheng, T., Salganik, M.J., Gelman, A., 2006. How many people do you know in prison? J. Am. Stat. Assoc. 101, 409–424.