Published in final edited form as:

Stat Med. 2018 January 30; 37(2): 236–248. doi:10.1002/sim.7259.

Effects of contact network structure on epidemic transmission trees: implications for data required to estimate network structure

Nicole Bohme Carnegie^{a,*}

^aJoseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, U.S.A

Abstract

Understanding the dynamics of disease spread is key to developing effective interventions to control or prevent an epidemic. The structure of the network of contacts over which the disease spreads has been shown to have a strong influence on the outcome of the epidemic, but an open question remains as to whether it is possible to estimate contact network features from data collected in an epidemic. The approach taken in this paper is to examine the distributions of epidemic outcomes arising from epidemics on networks with particular structural features to assess whether that structure could be measured from epidemic data, and what other constraints might be needed to make the problem identifiable. To this end, we vary the network size, mean degree, and transmissibility of the pathogen, as well as the network feature of interest: clustering, degree assortativity or attribute-based preferential mixing. We record several standard measures of the size and spread of the epidemic, as well as measures that describe the shape of the transmission tree in order to ascertain whether there are detectable signals in the final data from the outbreak. The results suggest that there is potential to estimate contact network features from transmission trees or pure epidemic data, particularly for diseases with high transmissibility or for which the relevant contact network is of low mean degree. Copyright

Keywords

Clustering; Assortativity; ERGM; epidemic model

1. Introduction

Understanding the dynamics of disease spread is key to developing effective interventions to control or prevent an epidemic. These dynamics are often investigated through the use of epidemic models. Historically, epidemic modeling has been dominated by mean-field models [1, 2], but there has been increasing interest in network-based models of disease spread, in which transmission can only occur over an existing edge in the network, representing contact sufficient for transmission between two individuals [3, 4, 5, 6, 7, 8, 9, 10].

*Correspondence to: Nicole Carnegie, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201-0413. carnegin@uwm.edu.

Mathematical results and simulation studies have established that different network structures lead to widely varying epidemic potential and outcomes [11, 12, 13, 14, 15, 8, 16, 17, 18], but it is not often known what the structure of an underlying contact network is when studying an epidemic. An open question remains as to whether we can learn something about contact network features from data collected in an epidemic.

Recent work in this area has attempted to make inferences about network structure and epidemic parameters strictly from dates of infection and recovery of infected individuals in the population [19, 20, 21, 22, 23]. A related area of research examines the potential for using genetic data from pathogens to estimate properties of the contact or transmission networks. Opinions on this potential for contact networks have been mixed: some investigators suggest that the signal is too weak for precise estimation [24, 25, 26], but others have shown that contact networks explain some variability in phylogenetic tree structures [27, 28, 29]. Other studies develop an approach to infer transmission network parameters directly from phylogenetic analysis of viral sequences [30, 31, 32].

The approach taken in this paper is to examine the distributions of epidemic outcomes arising from epidemics on networks with particular structural features to assess whether that structure could be measured from epidemic data, and what other constraints might be needed to make the problem identifiable. To this end, we vary the network size, mean degree, and transmissibility of the pathogen, as well as the network feature of interest: clustering, degree assortativity or attribute-based preferential mixing.

Degree assortativity is a tendency for edges to occur between nodes with a similar number of edges. Assortativity has significant effects on the size and spread of an epidemic [11, 12, 14, 16, 18], and can impede prevention efforts by making the network structure more robust [15]. Clustering occurs when there is a tendency for subgroups of nodes to have a high density of edges, with low density between subgroups. This may slow the spread of an epidemic or reduce the final size [33, 34, 18].

Attribute-based preferential mixing occurs in a heterogeneous population when nodes of different types have more or fewer edges between them than would be expected by chance. A classic example in social networks would be a tendency of people to have more friends of a similar age than would be expected in friendships were distributed randomly with respect to age. In exploring this feature, we specifically suppose that we have two communities and we wish to measure what percentage of edges are between communities rather than within community. This problem arose in planning for a community-randomized trial of an HIV prevention intervention [35, 36]. If the level of cross-community contamination is high, the effect of the intervention within community will be diluted compared to what it would be if the intervention had been rolled out in all communities. Thus, an estimate of contamination may allow for treatment effect estimates to be adjusted to reflect probable effectiveness if all communities were provided with the intervention.

This paper takes a simulation approach similar to those in [24, 33, 34, 8]. Because variability in the number of contacts individuals have is known to have a strong influence on epidemic spread [37, 38, 34], we maintain the degree distribution as fixed across simulations in order

to isolate effects of network features from covarying effects of degree distributions. In addition, we consider how effects may change with network size and with the transmissibility of the pathogen; both will influence the development of the epidemic and thus may interact with network features. We find more hope for modeling network features from epidemic data than previously suggested [24], especially for low-degree networks or highly transmissible pathogens.

In Section 2 we describe the simulation methodology used and the choice of epidemic and transmission tree measures used to assess signal for detection of network features. Results from the simulations are presented in Section 3 and Section 4 concludes.

2. Methods

The simulation proceeds by first generating sets of contact networks, then simulating disease spread over those networks, and finally calculating outcome measures on the resulting transmission trees. We use the *statnet* suite of packages [39] for network generation and the *epinet* package [20] for epidemic simulation over those networks (both in **R** [40]). Networks are distinguished by one of three features: clustering intensity, degree assortativity, or assortative mixing by community.

2.1. Contact network generation

In all simulations, we consider a population of N nodes with mean degree m . We first simulate an Erdős-Renyi network (also called Bernoulli or simply “random”), Y , in which each edge between a pair of nodes is present with probability $\frac{m}{N-1}$, yielding a binomial degree distribution with an expected mean degree of m . This is straightforward to generate by independently sampling edges $Y_{ij} \sim \text{Bernoulli}(\frac{m}{N-1})$ for $\{Y_{ij}; i < j\}$. When modeling community mixing, half of the nodes are assigned to each community.

For each sampled random network Y , we then generate networks with desired features that preserve the degree of nodes in Y (thus preserving the degree distribution) using a rewiring algorithm in the *ergm* package (see [41] for details). When examining clustering and assortativity, the set of networks generated from each Y include high/low clustering (Y^{HC} , Y^{LC}) and high/low degree assortativity (Y^{HA} , Y^{LA}). When studying cross-community mixing, the set of networks generated from each Y have one to 51 percent of all edges across communities (Y^{MP} , $P \in \{1, 6, 11, \dots, 51\}$).

The distribution from which these networks are sampled follows the general form

$$P_{\theta, \mathcal{Y}}(Y^* = y^*) = \frac{\exp(\theta g(y^*))}{\kappa(\theta, \mathcal{Y})}, y \in \mathcal{Y}, \quad (1)$$

where $\theta \in \Omega \subset \mathbb{R}$; is the model coefficient and $g(y)$ is the statistic of interest. This gives a probability distribution on the space of possible networks such that the expected values of the included statistics $g(y)$ are equivalent to the observed (when fitted to network

data) or target (when used for simulation), but with maximum entropy with regard to all other network features. The normalizing constant $\kappa(\theta, \mathcal{Y}) = \sum_{z \in \mathcal{Y}} \exp(\theta g(z))$. The space of possible networks $\mathcal{Y}$ in our setting is constrained to networks of size N with a set of node degrees equivalence to the base network Y .

For network simulation, we can specify either a target value for the statistic $g(y)$ or the parameter value θ . Table 1 gives the measurement statistic, model statistic and target values for simulated networks. Note that we measure clustering using a clustering coefficient, but the model statistic is a geometrically weighted edgewise shared partner (GWESP) statistic. GWESP measures the propensity to form edges between nodes that already have one or more nodes to which they are both connected. This is closely related to the clustering coefficient as defined by [42], a measure of the “cliquishness” of local networks. Degree correlation is calculated as **NEED DEFINITION FROM ERGM TERMS**.

We compare results for mean degree ranging from 2 to 28. Initial results are for networks of size 500 with mean degree 7 or 28, and networks of size 2000 with mean degree 28, which allows us to examine whether results scale with the density of the network ($= 0.014$ for the 500/7 and 2000/28 cases), or with mean degree (equal for the 500/28 and 2000/28 cases). Further results are from simulations of networks of size 500 over the full range of mean degree.

2.2. Epidemic spread

Once networks are obtained, we simulate a stochastic SEIR (Susceptible – Exposed – Infectious – Recovered) epidemic over each network. The population is fully susceptible, with one randomly selected node exposed at time 0. The waiting time in the exposed and infectious periods follow gamma distributions with parameters $(k_E, \theta_E) = (10, 5/10)$ and $(k_I, \theta_I) = (20, 8/20)$, yielding average durations of five and eight days, respectively. Gamma distributions are used to reflect empirical results showing a duration distribution more symmetrically distributed about the mean than an exponential distribution would allow [2, 43, 44].

The time to transmission across each edge between a susceptible and an infective node follows an exponential distribution with rate parameter β . We simulate epidemics with β corresponding to a basic reproductive number R_0 ranging from 0.9 to 20.4 in a random network with the specified density ($\beta = R_0/m \cdot k_I \cdot \theta_I$). For comparison, estimates of R_0 for some common diseases are given in Table 2; while most recent outbreaks have lower R_0 values, including a broad range helps to identify threshold effects. This is, of course, an approximation, as variations in network structure may affect R_0 , but exact computation of these quantities is beyond the scope of this paper. We condition on outbreaks of at least 20 infections; at low β such outbreaks are less likely to occur, implying that the distributions of epidemic features discussed here are not representative of all incursions of a pathogen into a society, but only of those that lead to an outbreak likely to be identified by public health surveillance efforts.

2.3. Choice of outcome measures

Outcome measures for analysis are taken from epidemic transmission trees. The transmission tree records both the time and source of infection for each new infection, enabling the construction of a tree as in Figure 1, which maps the path of transmission through the population. Each horizontal line represents one node (labeled with the numeric ID of the node); the black portion is the exposed period while red indicates the infectious period. The vertical dashed lines indicate transmissions. Thus, from the tree we can see who infected whom and when, as well as the transition times between stages for those who were infected. We characterize as “epidemic data” the times of infection and recovery, a subset of the information included in the tree that is observable from individual medical records. It is often the case that we observe not the time of infection, but the time of symptom onset. This should only pose a problem for inference if the incubation period is highly variable, thus making the time of symptom onset an unreliable measure of the sequencing of infections in the population.

We record several standard measures of the size and spread of the epidemic, as well as measures that describe the shape of the transmission tree in order to ascertain whether there are detectable signals from clustering in the final data from the outbreak. The measures considered are:

- Epidemic measures (require only infection and recovery times):
  - Total size of the epidemic (cumulative total number of people infected)
  - Maximum size of the infected population (at a single point in time)
  - Time of the maximum size of the infected population
  - Duration of the epidemic (time from the first infection to the last recovery)
- Transmission tree measures (require identity of infector):
  - Number of cherries (defined below)
  - Normalized Sackin balance index (defined below)

The epidemic measures are standard measures of the spread of an epidemic. These features seem likely to be determined by the epidemic parameters and degree distribution more so than the contact structure. Clustering or degree assortativity may extend the duration of the epidemic by forcing transmission through chains of low-degree nodes after the more concentrated core of the population has been exhausted. Clustering, and community clustering in particular, is likely to affect the maximum size of the infected population by placing bottlenecks between clustered groups; the infection will likely show a pattern of spreading rapidly through one cluster, then trickling through before spreading rapidly in a new cluster, forcing the maximum number of infectives to be lower than if the pathogen could spread more evenly through the whole population.

Cherries are pairs (A,B) where A infected B and neither went on to infect anyone else. When holding the degree distribution constant, it would seem at first that the number of cherries

will be primarily determined by the infectiousness of the pathogen - how likely is it that someone who is infected does not infect any others? A network that is tightly clustered or strongly assortatively mixed might have more cherries, however, as groups of nodes who share alters (nodes to whom they are connected) become saturated, leaving the later infections with no uninfected alters. Previous studies of network influence on phylogenetic tree structure using cherries have found strong differences in the number of cherries produced across network classes (random graphs, core-periphery graphs, etc.) [25, 29].

The normalized Sackin balance index is a commonly used measure of phylogenetic tree imbalance [45]. Imbalance in a transmission tree measures how evenly divided infectees are among infectors; if many nodes infect no others while a few infect many, the tree is considered imbalanced. Results from studies of the influence on the underlying network on balance have shown mixed results [27, 26, 25, 29].

3. Results

3.1. Clustering and assortativity vary independently

A summary of the clustering coefficients and degree correlations of simulated networks relative to the random base network is shown in Figure 2.

In all cases low clustering corresponds almost exclusively to networks with no 3-cycles (clustering coefficient = 0) and high clustering networks have substantially higher coefficients relative to the base network, while degree dis/assortativity is on target with correlations of ± 0.2 . A notable result is the near independence of clustering and degree correlation. Networks rewired to target assortativity have clustering coefficients comparable to random networks, and networks rewired to target clustering have degree correlations that may be slightly more negative than random networks, but an order of magnitude smaller than the disassortative networks.

3.2. Distribution of epidemic and tree measures by assortativity and clustering

In Figures 3 through 5, we summarize the distributions of outcome measures across experimental conditions by displaying the median and interquartile range (IQR) of simulated values across values of R_0 . As overlap of these central values declines, we would expect outcome measures to be more informative for estimation of network attributes.

We first present the simulation results for the number of cherries in the transmission tree. In Figure 3, the type of network (high clustering, etc.) is represented by color, while the network size and degree are represented by the hatching/shading. The solid line represents the median number of cherries, normalized by population size, while the shaded regions comprise the 25th through 75th percentiles of the distribution. There are two features to note: first, the distributions for high and low clustering or degree assortativity are separated from values of R_0 of 3.5 (assortativity) and 5 (clustering) upward across combinations of N and m . Second, the two networks of mean degree 28 have nearly identical profiles, but with more variability in the results for the smaller network (as would be expected due to the increase in data available in the larger network).

There is very little difference in the distribution of all other epidemic and tree measures discussed in Section 2.3 by either clustering or degree assortativity for the high-degree networks. In fact, the distributions appear nearly identical for all networks of mean degree 28, regardless of size, clustering or assortativity. We will thus focus our discussion of these measures on the networks of size 500 with mean degree 7. The total number infected shows little or no signal across network configurations and transmissibility parameters, and is not shown here. Figures 4 and 5 give the distributions of the Sackin index, duration of epidemic, the maximum size of the infected population and the time when that maximum occurs for clustering and degree assortativity, respectively. The solid line is the median of the distribution, while the dashed lines represent the 25th and 75th percentiles. In the case of clustering, we see strong and increasing separation on the maximum size of the infected population and Sackin index above moderate transmissibility levels. The time-related outcomes manifest less separation in distributions, with clear overlap of the second quartile of the high clustering distribution with the third quartile of the low clustering distribution, but the overlap does appear to be decreasing as transmissibility increases.

When comparing results by degree assortativity, there are strong parallels to the clustering case – more separation in imbalance and infected population size than in time-related outcomes, increasing separation as transmissibility increases – but with overall lower levels of separation. Indeed, the differences in distributions seen for the time of the maximum prevalence for clustering have wholly disappeared.

3.2.1. Threshold effects on mean degree—The similarity of results for networks of mean degree 28, regardless of size, suggest that the driving force behind changes in outcomes is not network density, but mean degree. This is consistent with epidemic theory in large, randomly mixing populations, wherein it is the rate of contacts that determines the spread of disease. There is also evidence that most measures examined here vary by clustering and/or degree assortativity in lower degree networks but not higher. This suggests that there may be a threshold on mean degree beyond which network structure has little impact on epidemic spread.

To examine this, we expanded the set of mean degrees for the 500 node networks to examine trends. Figures 6 and 7 show the results of simulations with 500 node networks and with m ranging from 2 to 28. Each point represents the average of the outcome measure over 200 simulations. We highlight two values of R_0 , 1.9 and 9.9. Thus we have one example below and above the transition point observed above for many measures around $R_0 = 5$.

There is evidence of a threshold effect for mean degree; the average value of many measures changes rapidly over low mean degrees and then largely stabilizes. This stabilization generally occurs earlier under high infectivity; this is to be expected as the pathogen can penetrate deeply into the population more easily.

For almost all outcomes, the differences in results between networks of different types disappear as mean degree increases. The one exception is cherries, which maintains large differences even for high mean degree when infectivity is high. In this situation, nodes are likely to infect all their alters, and high clustering would suggest that those alters are likely

to be connected to each other. Hence, they have fewer possible routes to continue to transmit, leading to the presence of a cherry in the transmission tree. There is little difference in any other outcome by degree assortativity at any mean degree when infectivity is high; we observe some difference for low mean degree and low infectivity in temporal outcomes and in balance. There are some difference by clustering for low mean degree across all measures.

3.3. Distribution of epidemic and tree measures by community mixing

We turn now to considering a particular type of clustering: cross-community contamination. Figures 8 and 9 explore the effect of varying the percent of edges between communities on the resulting epidemic and transmission tree measures. In Figure 8 we see plots similar to those for generic clustering and degree mixing above for two extreme cases: near-complete separation (one percent of edges between communities) and effective blending into a single community (fifty-one percent of edges between communities). In this comparison, we see a drastic effect on the maximum size of the infected population, and some effect on the time at which that maximum occurred. There is also a distinct shift in the duration of the epidemic, with epidemics on average being shorter when communities are integrated. In the transmission trees, we do not see the effect on cherries that we observed with generic clustering and degree mixing, but the imbalance of the tree is strongly affected.

Of course, we expect that the true level of contamination is somewhere between these two extremes, so it is helpful to visualize how the measures change over a range of cross-community mixing rates. Figure 9 includes contour plots of the medians of epidemic and transmission tree measures across mixing rates and transmission intensities. The number of cherries and the total number infected are excluded, as these measures do not vary by mixing level. To detect the level of cross-community mixing in a given epidemic, we would like to see non-vertical contours, since community structure varies along the vertical axis. This type of behavior is most evident for the maximum size of the infected population and the Sackin index, and less so for the duration of the epidemic and the number of cherries.

4. Discussion

The results presented above suggest that contact network structure affects epidemic behavior under some conditions. Clustering, assortativity, and community structure have their strongest effects when the mean degree of the network is low or the transmissibility of the pathogen is high. We would expect as mean degree increases in a random network that epidemic behavior on that network approaches the behavior of a mean-field model. It appears that a sufficiently high mean degree outweighs the influence of higher-level structure as well. When transmissibility is low, we are unlikely to observe much variation in the onward infections caused by infected individuals (once the degree distribution is controlled for), which would dampen the influence of the network structure.

These findings are broadly consistent with several studies of network influence on phylogenetic tree shape [27, 26, 29], which find that network types (random, small world, etc.) and, in some cases, network structure are related to structural features of phylogenetic trees. Most literature on the effect of network clustering or assortativity on epidemic

outcomes focuses on outbreak probability and final size. We do not evaluate outbreak size, and find modest association with final size, in contradiction to some work [8, 17, 18]. Our finding of a significant association between the number of cherries in the transmission tree and network clustering is contradictory to the broad conclusions made by Welch (2011), though at the particular infectivity used there, we also observe little difference due to network clustering.

The observed contradictions with existing literature are likely due to two features of this simulation study: fixed degree distributions and varying transmissibility. When the mean degree in the network is fixed but the degree distribution is not, varying clustering or assortativity will tend to alter the variance of the degree distribution, which has been shown to be related to epidemic behavior. Thus, our somewhat more conservative results with regard to clustering are likely because we have controlled both the mean and variance of the degree distribution. Most existing work either considers a single value of transmissibility [13, 14, 12, 24], or a narrow range of possibilities [18, 29], and the values considered are generally in the low end of the range considered in this paper.

The effects of contact network structure are most evident in structural features of the transmission tree. Cherries appear to capture local clustering, whereas the Sackin index captures global clustering, as reflected here in community structure. This reflects the fact that cherries are essentially a local measure (it is only necessary to know the timing of onward infections of an infector-infectee pair to identify them as a cherry), while the imbalance measured by the Sackin index is a global feature.

While we show that clustering and assortativity are not necessarily entangled as structural features of a network (Figure 2), their effects on the transmission tree are similar. Thus, it may be difficult to disentangle the two features when making inferences about network structure. This is particularly true given that observed data are likely to be noisier than those used in our simulations. In particular, the precise times and sources of infection encoded in the transmission tree are unlikely to be available. Phylogenetic data on the pathogen from infected hosts could be an appropriate approximation, and the Sackin index used here and number of cherries are also valid measures on phylogenetic trees.

Statistically, these results indicate that aspects of contact network structure in a low-degree network could be inferred from full data on an epidemic outbreak, particularly if there were also data available on the degree distribution of the network. Since sensitivity to network structure varies by transmissibility, estimation would also be aided by inclusion of data on pathogen transmissibility when available. A Bayesian approach integrating multiple data sources would be a logical tool for this problem. Some models using only epidemic data have been developed [19, 20, 21, 22, 23], but a more inclusive approach may permit the measurement of complex structural features like clustering.

Acknowledgments

Contract/grant sponsor: This work is supported by NIH grant R37 AI51164

References

1. Anderson RM, May RM. Infectious diseases of humans. Oxford University Press; 1991.
2. Keeling MJ, Rohani P. Modeling infectious disease in humans and animals. Princeton University Press; 2008.
3. Potterat JJ, Rothenberg RB, Muth SQ. Network structural dynamics and infectious disease propagation. *International journal of STD & AIDS*. 1999; 10(3):182–185. [PubMed: 10340199]
4. Morris M, Kretzschmar M. A microsimulation study of the effect of concurrent partnerships on the spread of HIV in Uganda. *Mathematical Population Studies*. 2000; 8(2):109–133.
5. Klov Dahl AS. Networks and pathogens. *Sexually Transmitted Diseases*. 2001; 28(1):25–28. [PubMed: 11196041]
6. Halloran ME, Longini J, Ira M, Nizam A, Yang Y. Containing bioterrorist smallpox. *Science*. 2002; 298(5597):1428–1432. DOI: 10.1126/science.1074674 [PubMed: 12434061]
7. Keeling MJ, Eames KT. Networks and epidemic models. *Journal of the Royal Statistical Society Interface*. 2005; 2:295–307. DOI: 10.1098/rsif.2005.0051
8. Keeling MJ. The implications of network structure for epidemic dynamics. *Theoretical Population Biology*. 2005; 67:1–8. [PubMed: 15649519]
9. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Sanchez J, Grinsztejn B, Koblin BA, Mayer KH, Buchbinder SP. What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? *PLoS ONE*. 2012; 7(11):e50 522.doi: 10.1371/journal.pone.0050522
10. Carnegie NB, Goodreau SM, Liu A, Vittinghoff E, Sanchez J, Lama JR, Buchbinder S. Targeting pre-exposure prophylaxis among men who have sex with men in the united states and peru: Partnership types, contact rates, and sexual role. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2015; 69(1):119–125. [PubMed: 25942463]
11. Gupta S, Anderson RM, May RM. Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS*. 1989; 3:807–817. [PubMed: 2517202]
12. Anderson R, Gupta S, Ng W. The significance of sexual partner contact networks for the transmission dynamics of HIV. *Journal of Acquired Immune Deficiency Syndromes*. 1990; 3:417–429. [PubMed: 2179528]
13. Morris M, Kretzschmar M. Concurrent partnerships and transmission dynamics in networks. *Social Networks*. 1995; 17:299–318.
14. Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of infectious disease. *Mathematical Biosciences*. 1996; 133:165–195. [PubMed: 8718707]
15. Newman MEJ. Mixing patterns in networks. *Physical Review E*. 2002; 67:026–126. DOI: 10.1103/PhysRevE.67.026126
16. Noh JD. Percolation transition in networks with degree-degree correlation. *Physical Review E*. 2007; 76:026–116. DOI: 10.1103/PhysRevE.76.026116
17. Miller JC. Spread of infectious disease through clustered populations. *Journal of the Royal Society Interface*. 2009; 6:1121–1134. DOI: 10.1098/rsif.2008.0524
18. Badham J, Stocker R. The impact of network clustering and assortativity on epidemic behavior. *Theoretical Population Biology*. 2010; 77:71–75. DOI: 10.1016/j.rpb.2009.11.003 [PubMed: 19948179]
19. Britton T, O'Neill PD. Bayesian inference for stochastic epidemics in populations with random social structure. *Scandinavian Journal of Statistics*. 2002; 29:375–390.
20. Groendyke C, Welch D, Hunter DR. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics*. 2011; 38:600–616. DOI: 10.1111/j.1467-9469.2010.00721.x
21. Groendyke C, Welch D, Hunter DR. A network-based analysis of the 1861 Hagelloch measles data. *Biometrics*. 2012; 68:755–765. DOI: 10.1111/j.1541-0420.2012.01748.x [PubMed: 22364540]
22. Ray J, Marzouk Y. A Bayesian method for inferring transmission chains in a partially observed epidemic. *Proceedings of the Joint Statistical Meetings: Conference; Denver, Colorado*. 2008; 3–7.
23. Ball F, Mollison D, Scalia-Tomba G. Epidemics with two levels of mixing. *The Annals of Applied Probability*. 1997; 7:46–89.

24. Welch D. Is network clustering detectable in transmission trees? *Viruses*. 2011; 3:659–676. DOI: 10.3390/v3060659 [PubMed: 21731813]
25. Frost SD, Volz EM. Modelling tree shape and structure in viral phylodynamics. *Philosophical Transactions of the Royal Society B*. 2013; 368:20120, 208.
26. Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the dynamics and structure of sexual contact networks shape pathogen phylogenies. *PLoS Computational Biology*. 2013; 9:e1003105.doi: 10.1371/journal.pcbi.1003105
27. Leventhal GE, Kouyos R, Stadler T, von Wyl V, Yerly S, Böni J, Cellerai C, Klimkait T, Günthard HF, Bonhoeffer S. Inferring epidemic contact structure from phylogenetic trees. *PLoS Computational Biology*. 2012; 8:e1002413.doi: 10.1371/journal.pcbi.1002413
28. Pluciński MM, Starfield R, Almeida RPP. Inferring social network structure from bacterial sequence data. *PLoS ONE*. 2011; 6:e22685.doi: 10.1371/journal.pone.0022685 [PubMed: 21829645]
29. Colijn C, Gardy J. Phylogenetic tree shapes resolve disease transmission patterns. *Evolution, Medicine, and Public Health*. 2014; 2014(1):96–108. DOI: 10.1093/emph/eou018
30. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Medicine*. 2008; 5:e50.doi: 10.1371/journal.pmed.0050050 [PubMed: 18351795]
31. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT. on behalf of the UK HIV Drug Resistance Collaboration. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *Journal of Infectious Diseases*. 2011; 204:1463–1469. [PubMed: 21921202]
32. Yerly S, Junier T, Gayet-Ageron A, El Amari EB, von Wyl V, Günthard HF, Hirschel B, Zdobnov E, Kaiser L. the Swiss HIV Cohort Study. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. *AIDS*. 2009; 23:1415–1423. [PubMed: 19487906]
33. Eames K. Modelling disease spread through random and regular contacts in clustered populations. *Theoretical Population Biology*. 2008; 73(1):104–111. [PubMed: 18006032]
34. Kiss IZ, Green DM. Comment on “properties of highly clustered networks”. *Physical Review E*. 2008; 78:048101.
35. Wang R, Goyal R, Lei Q, Essex M, De Gruttola V. Sample size considerations in the design of cluster randomized trials of combination HIV prevention. *Clinical Trials*. 2014; 11:309–318. [PubMed: 24651566]
36. Datz T. [Last accessed: 2 May 2013] Harvard School of Public Health awarded \$20 million CDC grant to study HIV prevention in Botswana. *HSPH News*. Sep 20, 2011. <http://www.hsph.harvard.edu/news/press-releases/20-million-grant-hiv-botswana>
37. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Physical Review Letters*. 2001; 86(14):3200–3203. DOI: 10.1103/PhysRevLett.86.3200 [PubMed: 11290142]
38. May RM, Lloyd AL. Infection dynamics on scale-free networks. *Physical Review E*. 2001; 64:066112.doi: 10.1103/PhysRevE.64.066112
39. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. *statnet: Software tools for the Statistical Modeling of Network Data*. Seattle, WA: 2003. <http://statnetproject.org>
40. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2014. <http://www.R-project.org>
41. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. *ergm: A package to fit, simulate and diagnose exponential-family models for networks*. *Journal of Statistical Software*. 2008; 24(3): 1–29. URL <http://www.jstatsoft.org/v24/i03>. [PubMed: 18612375]
42. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature*. 1998; 393:440–442. [PubMed: 9623998]
43. Sartwell PE. The distribution of incubation periods of infectious disease. *American Journal of Hygiene*. 1950; 51(3):310–318. [PubMed: 15413610]
44. Simpson REH. Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *The Lancet*. 1952; 260(6734):549–554.
45. Blum MG, François O. On statistical tests of phylogenetic tree imbalance: The Sackin and other indices revisited. *Mathematical Biosciences*. 2005; 195(2):141–153. URL <http://>

www.sciencedirect.com/science/article/pii/S0025556405000544. DOI: 10.1016/j.mbs.2005.03.003 [PubMed: 15893336]

46. Mossong J, Muller C. Estimation of the basic reproduction number of measles during an outbreak in a partially vaccinated population. *Epidemiology and Infection*. 2000; 124:273–278. [PubMed: 10813153]
47. Van Effelterre T, Shkedy Z, Aerts M, Molenberghs G, Van Damme P, Beutels P. Contact patterns and their implied basic reproductive numbers: and illustration for varicella-zoster virus. *Epidemiology and Infection*. 2009; 137(1):48–57. [PubMed: 18466660]
48. Kretzschmar M, Teunis PF, Pebody RG. Incidence and reproduction numbers of pertussis: estimates from serological and social contact data in five European countries. *PLoS Medicine*. 2010; 7(6):e1000291.doi: 10.1371/journal.pmed.1000291
49. Coburn BJ, Wagner BG, Blower S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Medicine*. 2009; 7:30.doi: 10.1186/1741-7015-7-30 [PubMed: 19545404]
50. Blower SM, McLean A. Prophylactic vaccines, risk behavior change, and the probability of eradicating hiv in san francisco. *Science*. 1994; 265:1451–1451. [PubMed: 8073289]
51. Chowell G, Hengartner N, Castillo-Chavez C, Fenimore P, Hyman J. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*. 2004; 229(1):119–126. DOI: 10.1016/j.jtbi.2004.03.006 [PubMed: 15178190]
52. WHO Ebola Response Team. Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections. *New England Journal of Medicine*. 2014; 371:1481–1495. DOI: 10.1056/NEHMoA1411100 [PubMed: 25244186]
53. Nishiura H, Kinoshita R, Mizumoto K, Yasuda Y, Nah K. Transmission potential of zika virus infection in the south pacific. *International Journal of Infectious Diseases*. 2016; 45:95–97. DOI: 10.1016/j.ijid.2016.02.017 [PubMed: 26923081]

Figure 1.

A sample transmission tree from the first 50 infections in one epidemic simulation. Each horizontal line represents one node (labeled with the numeric ID of the node); the black portion is the exposed period while red indicates the infectious period. The vertical dashed lines indicate transmissions.

Figure 2.

Median and interquartile range of clustering coefficients (L) and degree correlations (R) of generated networks. Note that the IQR is often so narrow as to be indistinguishable from the plotted median.

Figure 3.

Median and interquartile range of the number of cherries (normalized by population size) over 250 simulations in networks of varying size and mean degree with high (green) and low (purple) clustering and high (orange) and low (blue) assortativity. Results are plotted against the R_0 expected in a random-mixing model.

Figure 4.

Median and interquartile range of selected epidemic and tree measures over 250 simulations in networks of size 500 with mean degree 7 with high (green) and low (purple) clustering. Results are plotted against the R_0 expected in a random-mixing model.

Figure 5.

Median and interquartile range of selected epidemic and tree measures over 250 simulations in networks of size 500 with mean degree 7 with high (orange) and low (blue) degree assortativity. Results are plotted against the R_0 expected in a random-mixing model.

Figure 6. Median of epidemic and tree measures over 200 simulations in networks of size 500 with varying mean degree and high (green) and low (purple) clustering. Results are plotted for two values (1.9, 9.9) of R_0 as expected in a random-mixing model.

Figure 7.

Median of epidemic and tree measures over 200 simulations in networks of size 500 with varying mean degree and high (orange) and low (blue) degree assortativity. Results are plotted for two values (1.9, 9.9) of R_0 as expected in a random-mixing model.

Figure 8.

Median and interquartile range of epidemic and transmission tree measures for communities with one percent cross-community ties (low) and 51 percent (high).

Figure 9.

Median of epidemic and transmission tree measures by percent cross-community ties and the R_0 expected in a random-mixing model

Table 1

Specifications for network simulation.

Feature	Measurement Statistic	Model Statistic	Target
Assortativity	Degree correlation	Degree correlation	$g(y) = \pm 0.2$
Clustering	Clustering coefficient	GWESP	$\theta = \pm 5$
Mixing	Percent edges across communities	Number of edges (within 1, across, within 2)	$g(y) = (\frac{1-\pi}{2}, \pi, \frac{1-\pi}{2}) \cdot Y \mid \pi \in \{0.01, 0.06, \dots, 0.51\}$

Table 2

Estimated values of the basic reproductive number R_0 for selected infectious diseases.

Disease	R_0
Measles	5–18 [46, 1]
Chicken pox	3–26 [47, 1]
Pertussis	5.2–18 [48, 1]
Poliomyelitis	5–7 [1]
Flu	0.9–2.8 [49]
HIV	2–12 [50, 1]
Ebola	1.2–2.0 [51, 52]
Zika virus	1.8–5.8 [53]