American Journal of Epidemiology
Copyright © 2005 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved; printed in U.S.A.
Vol. 162, No. 10
DOI: 10.1093/aje/kwi308
Advance Access publication September 21, 2005

## Practice of Epidemiology

# Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals

**R. M. Christley[1], G. L. Pinchbeck[1], R. G. Bowers[2], D. Clancy[2], N. P. French[1], R. Bennett[2], and J. Turner[1]**

[1] Epidemiology Group, Faculty of Veterinary Science, University of Liverpool, Liverpool, United Kingdom.
[2] Department of Mathematics, Faculty of Science, University of Liverpool, Liverpool, United Kingdom.

*Received for publication December 21, 2004; accepted for publication June 7, 2005.*

Simulation studies using susceptible-infectious-recovered models were conducted to estimate individuals' risk of infection and time to infection in small-world and randomly mixing networks. Infection transmitted more rapidly but ultimately resulted in fewer infected individuals in the small-world, compared with the random, network. The ability of measures of network centrality to identify high-risk individuals was also assessed. "Centrality" describes an individual's position in a population; numerous parameters are available to assess this attribute. Here, the authors use the centrality measures degree (number of contacts), random-walk betweenness (a measure of the proportion of times an individual lies on the path between other individuals), shortest-path betweenness (the proportion of times an individual lies on the shortest path between other individuals), and farness (the sum of the number of steps between an individual and all other individuals). Each was associated with time to infection and risk of infection in the simulated outbreaks. In the networks examined, degree (which is the most readily measured) was at least as good as other network parameters in predicting risk of infection. Identification of more central individuals in populations may be used to inform surveillance and infection control strategies.

disease outbreaks; disease transmission; infection; population surveillance

Abbreviations: HIV, human immunodeficiency virus; $R_0$, basic reproductive number.

Recent high-profile outbreaks of infectious diseases have demonstrated the important role of complex networks of transmission routes. Human immunodeficiency virus (HIV) transmits through networks of people linked through a range of contacts, including sexual contact and intravenous drug use (1, 2). In the 2003 outbreak of severe acute respiratory syndrome, long distance travel, particularly by air, resulted in rapid and apparently unpredictable global dissemination (3) with subsequent local transmission (4, 5). Similar patterns are evident in diseases of other species; during the foot-and-mouth disease outbreak in livestock in the United Kingdom, spread during the critical first days of the outbreak was influenced by long range animal movements (6, 7), followed by local transmission. These outbreaks clearly illustrate the role of complex heterogeneous contact networks in the transmission of agents. Complex networks can be conceptualized and modeled using social network analysis. Numerous measures have been developed to determine each individual's role within a population. "Centrality" measures the importance of individuals within a population according to one criterion or another (8). These criteria may be based, among other things, on the number of contacts made by each individual, the number of steps from each individual to all others, or the frequency that an individual appears on paths between others (refer to the Appendix). Social network analysis has already been applied to the study of infectious disease, including HIV and syphilis in human populations (9, 10) and *Mycobacterium*

Correspondence to Dr. Robert M. Christley, Epidemiology Group, Faculty of Veterinary Science, University of Liverpool, Leahurst, Neston CH64 7TE, United Kingdom (e-mail: robc@liverpool.ac.uk).



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

*bovis* in captive possums (11). In these examples, centrality measures were used to suggest which individuals would be pivotal to the spread of infection. In one study (11), predictions were confirmed by experimental introduction of *M. bovis* into the possum population. While there appears to be an association between many network parameters and risk of infection, only linear relations have previously been considered (10). Furthermore, the most appropriate parameters for predicting vulnerability to infection have not been identified. The identification of such parameters will help to identify targets for interventions, such as control procedures and surveillance, and to inform mathematical models and simulations to investigate disease transmission through population networks (10, 12–16). Identification of individuals with high vulnerability to infection could greatly enhance targeted surveillance and prevention programs. In addition, as node degree can be measured on individuals without requiring observation of entire population networks (as for the other centrality measures), estimation of the relative utility of node degree, compared with other measures, will help to evaluate the usefulness of this relatively easy-to-measure parameter.

Here, we investigate the relation of several measures of network centrality to the risk of infection after emergence of a novel infectious disease among fully susceptible individuals living in isolated populations. We utilize individual-based modeling approaches based on explicitly described contact networks in which individuals are linked by stable (rather than dynamic) connections. These networks represent both random and small-world mixing patterns, the latter characterized by a similar average path length but greater clustering compared with the random network (17). The networks were chosen to represent commonly used theoretical models (random mixing) and more realistic models of social structures (small-world models).

## MATERIALS AND METHODS

### Network generation

In social network analysis, the units of interest are sometimes referred to as ''nodes''; these may be individuals or collections of individuals (such as a family or town). Nodes are connected via edges or arcs; ''edges'' are contacts by which an infectious agent (or other entity) can travel in either direction (e.g., transmission of a respiratory virus by close social contact), and an ''arc'' is a contact by which an agent can transmit in only one direction (e.g., transmission via transfused blood). Here, we consider only contacts in which transmission can occur in either direction (edges) and that were stable over time.

In this study, two networks were generated, each consisting of 100 nodes. Each pair of nodes $(i, j)$ was either connected or disconnected, with the connections being undirected, that is, if $i \rightarrow j$, then $j \rightarrow i$. The connections were defined through the following variable:

$$\delta_{ij} = \begin{cases} 0, & \text{if } i \not\rightarrow j \\ 1, & \text{if } i \rightarrow j \end{cases}.$$

The matrix $\Delta = \delta_{ij}$ defines an undirected graph. For each network, $\Sigma\delta_{ij}$ was set at 2,000. Hence, the total number of edges was 1,000 (as each edge connects to two nodes), and the network density $(\frac{1}{2}\Sigma\delta_{ij})/(N/2(N-1))$ was 0.2.

The method of selecting nodes for connection differed for the two networks. For the small-world network, each node was randomly assigned to one of $G = 24$ groups, with each group having between one and eight members. Initially, $\Sigma\delta_{ij} = 0$, and two nodes were selected with a probability proportional to their degree + 1 ($\Sigma_j\delta_{ij} + 1$). The nodes, $i$ and $j$, were connected with probability 1 if they were in the same group, $g_i = g_j$, and with 0.05 if $g_i \neq g_j$. The process was repeated until $\Sigma\delta_{ij} = 2{,}000$. This method incorporates the preferential selection of highly connected nodes, similar to the Barabasi-Albert network-generating model (18), but also generates clustering, because of the increased probability of contacts forming between individuals within the same group. This method resulted in a network with short-path length and high clustering, characteristics of small-world networks (17).

For the random network, two nodes were randomly selected and, if not already connected, these were connected with probability 1. This process was continued until $\Sigma\delta_{ij} = 2{,}000$ and generated a random network, in which each pair of nodes had the same probability of being connected, regardless of their original group allocation (19).

### Network characteristics

All network parameters listed in the Appendix and table 1 (with the exception of the random-walk betweenness and the assortativity coefficient) were calculated using UCInet version 6.16 for Windows (Analytic Technologies, Inc., Harvard, Massachusetts). Random-walk betweenness was calculated using the method described by Newman (20), implemented in Fortran 95 (The Fortran Company, Tucson, Arizona). The assortativity of each network was calculated with regard to mixing within and between allocated groups (21). Each individual was assigned to the core or periphery of the network using algorithms described by Borgatti and Everett (22). The algorithm attempts to identify core individuals such that the density of contacts within the core is maximized and the density between peripheral individuals is minimized, where ''density'' is the proportion of possible connections between nodes that are actually present.

### Outbreak simulation

Transmission of an infectious agent through each network was simulated using a susceptible-infectious-recovered model (23).

*Infection, $S \rightarrow I$.* Before introduction of the infectious agent, the entire population was assumed to be susceptible, in keeping with the emergence of a novel infection. An outbreak was initiated by randomly selecting a single node from the entire population to become infected. At each subsequent time step, transmission to a susceptible node from each in-contact infected node was tested. If a susceptible node $n_i$ has $k_i$ infectious neighbors, during the subsequent time period, it can become infected with probability



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

**TABLE 1. Characteristics of the networks**

| Variable | Small-world network | Random network |
|---|---|---|
| Size (no. of nodes) | 100 | 100 |
| Total no. of edges | 1,000 | 1,000 |
| Density* | 0.20 | 0.20 |
| Degree (no. of contacts) | | |
| Mean | 20 | 20 |
| Maximum | 52 | 36 |
| Minimum | 1 | 8 |
| Degree coefficient of variation | 0.64 | 0.22 |
| Path length† | | |
| Average | 1.9 | 1.8 |
| Maximum | 4 | 3 |
| Average clustering coefficient | 0.43 | 0.21 |
| Size of core (no. of nodes in core) | 43 | 37 |
| Density of core | 0.51 | 0.34 |
| Density of periphery | 0.06 | 0.18 |
| Density of core-periphery connections | 0.18 | 0.18 |

* "Density" is the proportion of all possible edges that is actually present in the network.

† "Path length" is the number of contacts on the shortest route between a pair of nodes in the network.

$1 - (1 - \beta)^{ki}$, where $\beta$ is the probability of transmission per contact.

*Recovery, $I \rightarrow R$.* Infectious nodes were recovered after 14 time steps (i.e., the duration of infectivity), after which time they could not transmit the infectious agent and were not susceptible to repeat infection (i.e., immunity was "lifelong").

The process was repeated until $I = 0$ (i.e., until the agent became extinct), at which time the final status of each node (susceptible or recovered) and the time to infection (survival time) for each infected node were recorded. For each network, three sets of 500 simulations were undertaken. Between sets, the duration of infectivity remained constant (14 time steps), but the transmission probability of the agent, $\beta$, was varied (0.00375, 0.0075, and 0.015).

### Probability of infection

The probability of infection for node $i$ was calculated from the number of simulations in each set in which node $i$ was infected and the total number of simulations (in both cases, excluding those simulations in which node $i$ was the index case). The relation between the probability of infection and the node characteristics (degree, random-walk betweenness, shortest-path betweenness, and farness) was explored using scatter plots. Generalized additive models were used to investigate the functional form of the relation between each centrality measure and the probability of infection for each set of simulations (24). The proportion of the null deviance explained by the generalized additive models was also calculated. Analysis was performed using R, version 1.9.0, software (http://www.r-project.org/).

### Survival analysis

Times to infection in the set with $\beta = 0.0075$ for the small-world and random networks were compared using Kaplan-Meier estimates of survival. The relation between continuous covariates and the time to infection was explored using penalized Cox regression models (25). This analysis was limited to the small-world network with $\beta = 0.0075$. Cox proportional hazards models, with a frailty (random) term included to allow for clustering within each set of simulations, were used to explore multivariable relations between covariates and the probability of infection. As the measures of network centrality were highly correlated, separate models were fitted for each centrality measure. The model diagnostics performed included Schoenfeld residual plots to test for nonproportionality (25). Analysis was performed using S-plus software (Insightful Corporation, Seattle, Washington).

## RESULTS

### Network structure and characteristics

The average path length and clustering coefficients for the two networks indicated a similar average distance between all nodes in each network but considerably greater clustering in the small-world network (table 1). The small-world network had a broad degree distribution (mean: 20, standard deviation: 12.8; coefficient of variation: 0.64), whereas that for the random network was approximately binomially distributed with a relatively small range (mean: 20, standard deviation: 4.4; coefficient of variation: 0.22). The measures of centrality for each of the networks were all associated with each other (data not shown).

Both networks demonstrated proportional (nonassortative) mixing based on allocated groups, with assortativity coefficients of 0.016 and 0.013 for the small-world and random networks, respectively. Analyses suggested the existence of a 43-node core in the small-world network and a 37-node core in the random network. The density of contacts within the core of the small-world network was 0.51. That is, each node within the core contacted, on average, 51 percent of the other nodes in the core. This was more than eight times greater than the density of 0.06 within the periphery of this network. In contrast, the density of contacts within the core was lower in the random network (compared with the small-world network), 0.34, and was only double that of the periphery of the random network, 0.18. The density of contacts between the core and periphery was 0.18 for both networks. The identified core predicted well those individuals at greatest risk of infection (e.g., figure 1, the small-world network).

### Probability of infection

All four network centrality measures were associated with the probability of infection in both the small-world network and the random network (figure 2). The univariable



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

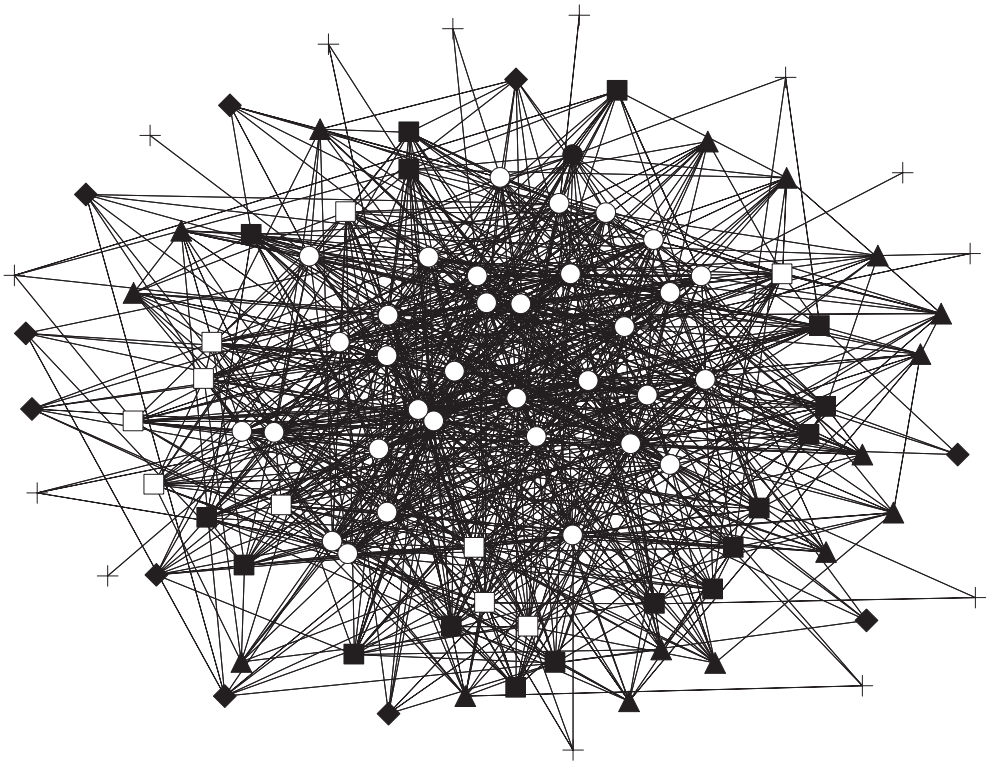

**FIGURE 1.** The small-world network. The shape of the nodes indicates the proportion of times each node was infected during the 500 simulations with transmission probability = 0.0075 (cross = 0–0.2; diamond = 0.201–0.4; triangle = 0.401–0.5; square = 0.501–0.6; circle = 0.601–0.7). White nodes belong to the core of the network, and black nodes belong to the periphery.

generalized additive models explained considerable proportions of the null deviance in most models (table 2). Increasing the basic reproductive number ($R_0$) (by increasing the transmission probability; $R_0 \propto \beta$) tended to increase the proportion of deviance explained by the generalized additive model. Increasing the basic reproductive number also increased the proportion of nodes at, or near, the highest probability of infection. For all transmission probabilities investigated, the maximum probability of infection was greatest for the random network, compared with the small-world network.

### Survival analysis

Kaplan-Meier estimates of survival indicate more rapid transmission, but smaller total outbreak size, in the small world, compared with the random, network (figure 3). For the small-world network, distance from the index case was associated with the risk of infection (table 3). Nodes one or two steps from the index case had similar probabilities of infection, whereas those three or more steps from the index case were less likely to become infected. In addition, the degree of the index case, the clustering coefficient, and each of the centrality measures were associated with risk of infection (table 3). Distance and degree of the index case both showed some evidence of nonproportionality. This was not surprising, as the characteristics of the index case would have less influence on whether a susceptible node becomes infected as time progresses and more nodes are infected, as these would act as alternate sources of infection.

## DISCUSSION

Despite advances in the understanding of the role of heterogeneous mixing on disease dynamics, the use of network parameters to predict infection risk has received only limited attention. Clearly, more central individuals are at greatest risk of infection during outbreaks. From these simulation studies, we have identified a number of network centrality measures that may be useful predictors of individuals' risk of infection and time to infection during outbreaks of diseases in naive populations. Importantly, in the examples investigated, the degree centrality appeared to perform at least as well as alternative measures. This may have important implications for disease surveillance and control. First, degree centrality is measured by observation of individuals and does not require assessment of entire networks (as the other measures do). Therefore, it is possible to assess the degree distribution of a population by sampling random individuals and subsequently to compare the degree centrality of individuals in a population (26). Second, in terms of disease-control strategies, it is likely to be simpler for



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

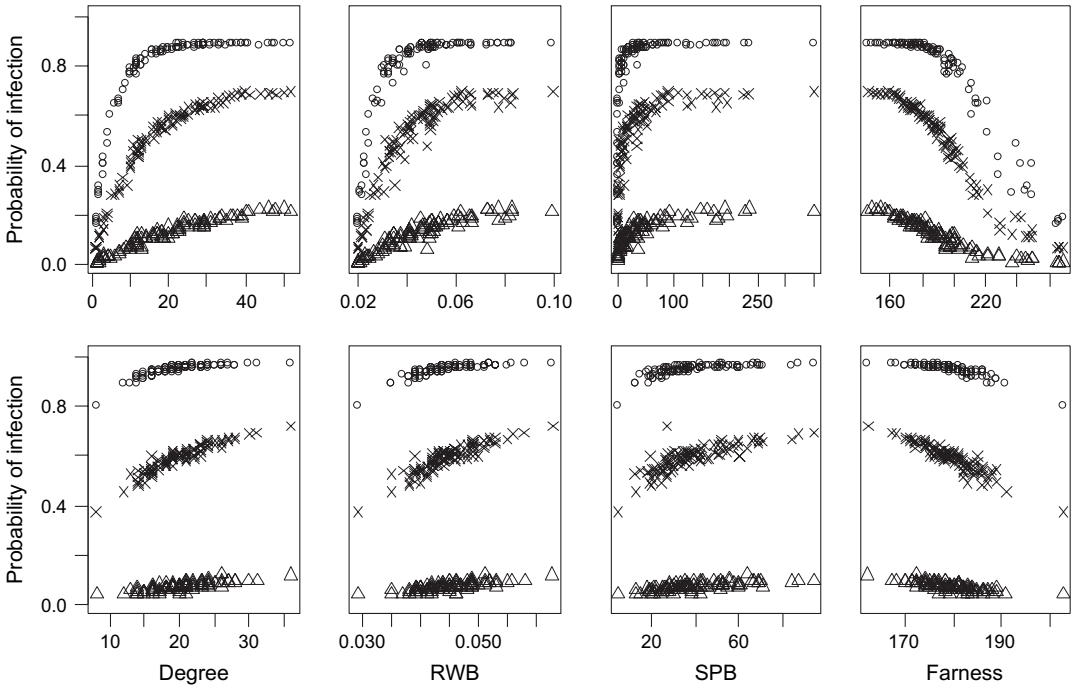


**FIGURE 2.** Relation between node degree, random-walk betweenness, shortest-path betweenness, and farness and the proportion of times infected during 500 simulations in the small-world (top) and random (bottom) networks, where $\beta = 0.00375$ (triangles), 0.0075 (crosses), and 0.015 (circles). RWB, random-walk betweenness; SPB, shortest-path betweenness.

people to respond to requests to decrease their degree centrality, as this is an egocentric parameter, rather than to alter their network-derived centrality. However, the centrality measures were associated with each other in these networks; such associations are often less evident in more complicated networks (data not shown). Hence, greater variation in the ability of different centrality measures to predict risk of infection would be expected, and measures other than degree may perform better in such networks.

**TABLE 2. Proportion of the null deviance of the probability of infection explained by each of the centrality measures using univariable generalized additive models**

| Network and variable | Proportion of null deviance explained | | |
|---|---|---|---|
| | $\beta = 0.00375$ | $\beta = 0.0075$ | $\beta = 0.015$ |
| Small world network | | | |
| Degree | 0.96 | 0.99 | 0.99 |
| Random-walk betweenness | 0.90 | 0.96 | 0.98 |
| Shortest-path betweenness | 0.83 | 0.84 | 0.71 |
| Farness | 0.94 | 0.98 | 0.97 |
| Random network | | | |
| Degree | 0.60 | 0.90 | 0.89 |
| Random-walk betweenness | 0.55 | 0.84 | 0.88 |
| Shortest-path betweenness | 0.54 | 0.71 | 0.79 |
| Farness | 0.58 | 0.85 | 0.84 |

The results of a previous study modeling transmission of HIV in a small population of cocaine injectors by use of a susceptible-infected model (10) agreed with those presented here; degree centrality predicted infection at least as well as many network-derived measures. In that study, undirected centrality measures performed equally well to directed measures, despite transmission simulations being

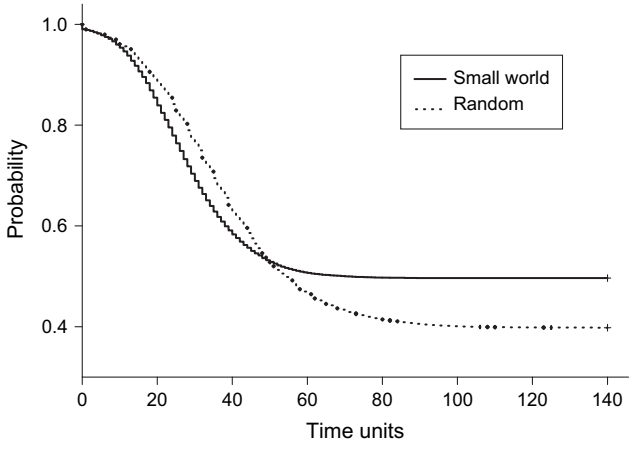


**FIGURE 3.** Kaplan-Meier estimates of survival probability (i.e., remaining uninfected) for the small-world and random networks based on the results of 500 simulations, with transmission probability = 0.0075.



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

**TABLE 3. Four Cox proportional hazards models, with a frailty (random) term included to allow for clustering within each simulation***

| Base model | Model 1 + degree | | Model 2 + SPB† | | Model 3 + farness | | Model 4 + RWB† | |
|---|---|---|---|---|---|---|---|---|
| | Hazard ratio | 95% confidence interval | Hazard ratio | 95% confidence interval | Hazard ratio | 95% confidence interval | Hazard ratio | 95% confidence interval |
| Variance of random effect | 2.84 | | 2.76 | | 2.85 | | 2.81 | |
| Base model | | | | | | | | |
| Distance to index case | | | | | | | | |
| 1 | 1.0 | | 1.0 | | 1.0 | | 1.0 | |
| 2 | 0.89 | 0.87, 0.92 | 0.76 | 0.73, 0.78 | 0.91 | 0.88, 0.93 | 0.84 | 0.81, 0.86 |
| 3 or 4 | 0.71 | 0.65, 0.77 | 0.45 | 0.42, 0.49 | 0.93 | 0.85, 1.01 | 0.63 | 0.58, 0.69 |
| CC(c)† ×10 | 1.12 | 1.10, 1.14 | 1.09 | 1.07, 1.11 | 1.05 | 1.03, 1.06 | 1.15 | 1.13, 1.17 |
| $[CC(c) \times 10]^2$ | 0.92 | 0.92, 0.93 | 0.88 | 0.87, 0.88 | 0.98 | 0.97, 0.99 | 0.92 | 0.91, 0.92 |
| Degree of the index case | 1.04 | 1.03, 1.05 | 1.03 | 1.02, 1.05 | 1.04 | 1.03, 1.06 | 1.04 | 1.02, 1.05 |
| Additional term | | | | | | | | |
| Degree | 1.05 | 1.05, 1.05 | | | | | | |
| SPB | | | 1.01 | 1.00, 1.01 | | | | |
| Farness | | | | | 0.97 | 0.97, 0.97 | | |
| RWB × 100 | | | | | | | 1.35 | 1.34, 1.36 |

* Each model consists of the base model and one additional term. Simulations were conducted using the small-world network and a transmission probability of 0.0075.

† SPB, shortest-path betweenness; RWB, random-walk betweenness; CC(c), clustering coefficient (centered).

based on a directed network (in order to account for the directed transmission potential of being the second user of a shared intravenous syringe). This study, however, considered only linear relations between the network measures and risk of infection (10). However, as the probability of infection is bounded by 0 and 1, the relation cannot be linear, except over a limited range, as illustrated by our results. Hence, future studies should not simply consider the correlation between networks' measures and risk of infection, but they should examine the form of these relations. This may, in certain circumstances, usefully inform control strategies by identifying the range of a certain characteristic for which individuals may need to be targeted for control strategies.

The relation between each of the centrality measures and the probability of infection was similar in both the small-world and random networks, suggesting that these may be used somewhat interchangeably across a range of network topologies. It has been suggested that the best representation of the path of an agent through a network is likely to be a random walk (20), as agents will not necessarily follow the shortest path to a ''target'' individual. Hence, it was not surprising that the measure of random-walk betweenness appeared to predict the probability of infection better than the shortest-path betweenness. Importantly, the proportion of the null deviance explained by degree was consistently greater than that for the other centrality measures. Despite the simplifications inherent in the models used here, the results suggest that measurement of network centrality may have real utility in the design of surveillance programs and control strategies. That degree centrality performed well gives hope that this relatively simple-to-measure parameter may be at least as appropriate for predicting risk of infection as measures based on entire networks. Measuring disease contact networks is problematic for many diseases because of the wide variety of possible exposures and the varying transmission potentials associated with these. However, there is evidence that even simplified measures of centrality (e.g., ignoring edge weight and direction) still have predictive ability (10, 11). Centrality measures were associated with not only a particular individual's risk of becoming infected but also the time to infection. Increasing centrality (defined by any of the measures investigated) was associated with earlier infection in the simulated epidemics. Hence, targeting central individuals for surveillance or control interventions should limit early spread of an agent throughout the population.

The rate of spread of the agent in the population and the ultimate size of the outbreak were influenced by network structure (27). The $R_0$, which influences the initial rate of spread of infection, is the average number of secondary cases arising from the entry of a single infected individual into a wholly susceptible population (23). In the absence of clustering, it can be estimated as $R_0 = \beta\kappa D(1 + (\mathrm{CV})^2)$, where $\beta$ is the probability of transmission, $\kappa$ is the average contact rate per unit time, $D$ is the duration of infectivity, and ''CV'' is the coefficient of variation of degree (14). As $\beta$, $\kappa$, and $D$ were the same for the small-world and random networks (within any set of simulations), it was expected that infection would initially spread more rapidly in the small-world network, as the coefficient of variation of degree was almost three times greater for this network. Hence, the $R_0$ in the small-world network was approximately 34 percent greater than that in the random network (for a given transmission probability and duration of infection). However, as the small-world network had a densely connected core and sparsely connected periphery, outbreaks in



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

this network were frequently confined to the core, with rapidly developing correlations in the infection status of connected nodes within the core, leading to local depletion of susceptible nodes (28) and increased probability of extinction of the pathogen. In contrast, the greater density of connections in the periphery of the random network enabled greater transmission within this part of the network. In this way, a greater proportion of the random network was likely to become exposed to the agent, and the final size of the outbreaks tended to be smaller in the small-world networks, compared with the random network. In light of this, models that assume random mixing may tend to overestimate the size of an outbreak while underestimating the initial rate of transmission.

Similar effects on the rate of spread and final outbreak size have been observed in models of HIV transmission, where assortative (like-with-like) mixing results in rapid transmission within a particular group, compared with proportional (random) or disassortative (like-with-unlike) mixing (29). However, both of our models demonstrated proportional mixing, indicating that such effects can occur in the absence of assortative mixing. The lack of assortative mixing in our models arose because of the large number of groups to which individuals could be allocated, each group having only 1–8 individuals. This resulted in considerable mixing between unlike individuals.

Analysis of dynamic networks (such as for sexual partnerships) has identified a strong association between concurrency and risk of transmission (30). Concurrency is a measure of the proportion of relationships between individuals that are coexistent. This measure is highly appropriate for dynamic networks, such as sexual relationships, which may be monogamous, serially monogamous, or concurrent. However, for static networks, as used here, all relationships can be considered concurrent. Further studies of the ability of measures of concurrency, as well as other centrality measures, to predict the risk of infection in dynamic networks are required.

The aim of this study was to investigate agent transmission in simple networks in order to draw broad conclusions regarding the utility of centrality measures for predicting probability of infection. However, the simplicity of the population networks and of the transmission algorithm induces limitations with regard to the generalizability of the results. These models do not incorporate inter- or intraindividual variation in the transmission parameter nor duration of infectivity. Importantly, for some diseases, these parameters are affected by an individual's social characteristics. For example, while increased and diverse social contacts increase the probability of contacting infectious agents by humans (31), there is evidence that increased sociability (incorporating both the diversity and number of contacts) reduces susceptibility to the common cold in most people (31–34), although this effect is reversed in stressed individuals (31). Evidently, for some diseases at least, increased social contact may increase the risk of exposure while, paradoxically, increasing resistance to infection. The populations used here also had constant probability of contact among all individuals connected by edges. Alternative models, for example, where individuals contact a fixed proportion of the individuals they are connected to by edges, may better represent some real populations and may elicit different results. Furthermore, our model does not permit behavioral alteration by noninfected individuals in the face of an outbreak or during clinical disease by infected individuals. However, many such behavioral modifications are reported in many species and following infection by numerous agents (35–37) because of both physiologic and social factors. For example, reductions in sexual activity and increased condom usage may occur following infection with sexually transmitted diseases (38). Similarly, improved hand cleaning and reduced outdoor activity were reported in areas affected by severe acute respiratory syndrome (39). Clearly, both theoretical and observational studies of specific diseases are required to better understand the complex relations between social characteristics and the risk of acquisition of infection.

In conclusion, measurement and analysis of contact networks may enable greater understanding of infection dynamics in populations and may inform surveillance and control procedures. Although transmission tends to occur more rapidly in small-world networks, the final outbreak size tends to be smaller. In the examples examined, the measures of network centrality were correlated with each other, and node degree (which is more readily measured) was at least as good as other network parameters in predicting risk of infection. Further work is required to determine the most appropriate centrality measure(s) for prediction of infection in models utilizing context-specific population contact structures and transmission parameters.

## ACKNOWLEDGMENTS

This work was supported by research grant VTRI VT0103 from the Higher Education Funding Council for England and the Department for Environment, Food, and Rural Affairs.

Conflict of interest: none declared.

## REFERENCES

1. Klovdahl AS, Potterat JJ, Woodhouse DE, et al. Social networks and infectious diseases: the Colorado Springs Study. Soc Sci Med 1994;38:79–99.
2. Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. Soc Sci Med 1985;21:1203–16.
3. Peiris JS, Yuen KY, Osterhaus AD, et al. The severe acute respiratory syndrome. N Engl J Med 2003;349:2431–41.
4. Svoboda T, Henry B, Shulman L, et al. Public health measures to control the spread of the severe acute respiratory syndrome during the outbreak in Toronto. N Engl J Med 2004;350:2352–61.
5. Lau JT, Tsui H, Lau M, et al. SARS transmission, risk factors, and prevention in Hong Kong. Emerg Infect Dis 2004;10:5875–92.
6. Mansley LM, Dunlop PJ, Whiteside SM, et al. Early dissemination of foot-and-mouth disease virus through sheep marketing in February 2001. Vet Rec 2003;153:43–50.



Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015

7. Kao RR. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. Trends Microbiol 2002; 16:279–86.
8. Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge, United Kingdom: Cambridge University Press, 1994.
9. Rothenberg RB, Potterat JJ, Woodhouse DE, et al. Choosing a centrality measure: epidemiologic correlates in the Colorado Springs Study of social networks. Soc Networks 1995;17: 273–97.
10. Bell DC, Atkinson JS, Carlson JW. Centrality measures for disease transmission networks. Soc Networks 1999;21:1–21.
11. Corner LA, Pfeiffer D, Morris RS. Social-network analysis of *Mycobacterium bovis* transmission among captive brushtail possums (*Trichosurus vulpecula*). Prev Vet Med 2003;59: 147–67.
12. Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2001;63:066117. Epub: May 22, 2001. (DOI: 10.1103/PhysRevE.63.066117).
13. Zanette DH, Kuperman M. Effects of immunization in small-world epidemics. Physica A 2002;309:445–52.
14. Lloyd AL, May RM. How viruses spread among computers and people. Science 2001;292:1316–17.
15. Liljeros F, Edling C, Amaral LAN, et al. The web of human sexual contacts. Nature 2001;411:907–8.
16. Bell DC, Montoya JM, Atkinson JS, et al. Social networks and forecasting the spread of HIV infection. J Acquir Immune Defic Syndr 2002;31:218–29.
17. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998;393:440–2.
18. Barabasi AL, Albert R. Emergence of scaling in random networks. Science 1999;286:509–12.
19. Newman ME, Watts DJ, Strogatz SH. Random graph models of social networks. Proc Natl Acad Sci U S A 2002;99: 2566–72.
20. Newman ME. A measure of betweenness centrality based on random walks. Condensed Matter 2003. (cond-mat/0309045).
21. Newman ME. Mixing patterns in networks. Phys Rev E Stat Nonlin Soft Matter Phys 2003;67:026126. Epub: Feb 27, 2003. (DOI: 10.1103/PhysRevE.67.026126).
22. Borgatti SP, Everett MG. Models of core/periphery structures. Soc Networks 1999;21:375–95.
23. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford, United Kingdom: Oxford University Press, 1992.
24. Hastie T, Tibshirani R. Generalised additive models. London, United Kingdom: Chapman and Hall, 1990.
25. Therneau TM, Grambsh PM. Modelling survival data. New York, NY: Springer, 2000.
26. Marsden PV. Egocentric and sociometric measures of network centrality. Soc Networks 2002;24:407–22.
27. Newman ME. Properties of highly clustered networks. Phys Rev E Stat Nonlin Soft Matter Phys 2003;68:026121. Epub: Aug 21, 2003. (DOI: 10.1103/PhysRevE.68.026121).
28. Keeling MJ. The effects of local spatial structure on epidemiological invasions. Proc Biol Sci 1999;266:859–67.
29. Boily MC, Anderson RM. Sexual contact patterns between men and women and the spread of HIV-1 in urban centres in Africa. IMA J Math Appl Med Biol 1991;8:221–47.
30. Potterat JJ, Zimmerman-Rogers H, Muth SQ, et al. *Chlamydia* transmission: concurrency, reproduction number, and the epidemic trajectory. Am J Epidemiol 1999;150:1331–9.
31. Hamrick N, Cohen S, Rodriguez MS. Being popular can be healthy or unhealthy: stress, social network diversity, and incidence of upper respiratory infection. Health Psychol 2002; 21:294–8.
32. Cohen S, Brissette I, Skoner DP, et al. Social integration and health: the case of the common cold. J Soc Struct 2000;1:1–10. (http://www.cmu.edu/joss/content/articles/volume1/cohen.html).
33. Cohen S, Doyle WJ, Skoner DP, et al. Social ties and susceptibility to the common cold. JAMA 1997;277:1940–4.
34. Cohen S, Doyle WJ, Turner R, et al. Sociability and susceptibility to the common cold. Psychol Sci 2003;14:389–95.
35. Hart BL. Biological basis of the behavior of sick animals. Neurosci Biobehav Rev 1988;12:123–37.
36. Hart BL. Behavioural adaptations to pathogens and parasites: five strategies. Neurosci Biobehav Rev 1990;14:273–94.
37. Hart BL. Behavioural defense against parasites: interaction with parasite invasiveness. Parasitology 1994;109:S139–51.
38. Payn B, Tanfer K, Billy JO, et al. Men's behavior change following infection with a sexually transmitted disease. Fam Plann Perpect 1997;29:152–7.
39. Tan X, Li S, Wang C, et al. Severe acute respiratory syndrome epidemic and change of people's health behaviour in China. Health Educ Res 2004;19:576–80.


## APPENDIX

### Glossary of Terms Used in Social Network Analysis

- *Node*—the units of interest in the network, often individuals, but may represent groups.
- *Edge*—an undirected link between nodes.
- *Network centrality*—Centrality measures a node's position in a network. Many measures of centrality exist, including the following:
  - *Degree*—the sum of contacts made by node *i* to other nodes.
  - *Farness*—the sum of the shortest network distances from node *i* to all other nodes in the network.
  - *Shortest-path betweenness*—the proportion of shortest distances between all pairs of nodes (excluding node *i*) that pass through node *i*.
  - *Random-walk betweenness*—the proportion of random walks between all pairs of nodes (excluding node *i*) that pass through node *i*.
- *Clustering*—Often, two nodes that contact a third node also contact one another, forming a triad. This is known as clustering. The clustering coefficient is the probability that nodes *i* and *j*, both of which connect to node *k*, also directly contact each other.
- *Path length*—The path length between nodes *i* and *j* is the number of contacts that separate the nodes. The average path length is the average number of contacts between all pairs of nodes in a network.

Downloaded from http://aje.oxfordjournals.org/ at Virginia Tech on April 24, 2015