

Detail in network models of epidemiology: are we there yet?

S. Eubank , C. Barrett , R. Beckman , K. Bisset , L. Durbeck , C. Kuhlman , B. Lewis , A. Marathe , M. Marathe & P. Stretz

To cite this article: S. Eubank , C. Barrett , R. Beckman , K. Bisset , L. Durbeck , C. Kuhlman , B. Lewis , A. Marathe , M. Marathe & P. Stretz (2010) Detail in network models of epidemiology: are we there yet?, Journal of Biological Dynamics, 4:5, 446-455, DOI: [10.1080/17513751003778687](https://doi.org/10.1080/17513751003778687)

To link to this article: <http://dx.doi.org/10.1080/17513751003778687>



Copyright Taylor and Francis Group, LLC



Published online: 05 May 2010.



Submit your article to this journal [↗](#)



Article views: 364



View related articles [↗](#)



Citing articles: 7 View citing articles [↗](#)

Detail in network models of epidemiology: are we there yet?

S. Eubank*, C. Barrett, R. Beckman, K. Bisset, L. Durbeck, C. Kuhlman, B. Lewis, A. Marathe, M. Marathe and P. Stretz

Network Dynamics and Simulation Sciences Laboratory, Virginia Bioinformatics Institute at Virginia Tech, 1880 Pratt Dr, Blacksburg, VA 24060, USA

(Received 1 October 2009; final version received 1 February 2010)

Network models of infectious disease epidemiology can potentially provide insight into how to tailor control strategies for specific regions, but only if the network adequately reflects the structure of the region's contact network. Typically, the network is produced by models that incorporate details about human interactions. Each detail added renders the models more complicated and more difficult to calibrate, but also more faithful to the actual contact network structure. We propose a statistical test to determine when sufficient detail has been added to the models and demonstrate its application to the models used to create a synthetic population and contact network for the USA.

Keywords: social network; mathematical modelling; epidemiology; mathematical biology; epidemic

1. Introduction

Network-based models for infectious disease epidemiology are natural generalizations of stochastic mass action models to stochastic dynamics on arbitrary interaction networks. Consider, for example, the following typical network model:

- a *network*, or graph G with labelled vertices and edges representing people and possible transmission paths, respectively;
- vertex labels, drawn from the set $\{S(\text{usceptible}), I(\text{nfected}), R(\text{ecovered})\}$, represent the corresponding person's state of health;
- edge labels, real numbers drawn from the interval $[0, 1]$, represent the conditional probability of transmission from one person to another;
- discrete-time dynamics corresponding to percolation across the graph, i.e. the label on a vertex in state S changes to I based on independent Bernoulli trials for each neighbour in state I with the probability specified by the edge between them; the label on a vertex in state I changes to R with some fixed probability.

In appropriate limits, specifically as the number of vertices tends to infinity and time becomes continuous, the mean behaviour of this model reduces to a standard Kermack and McKendrick

*Corresponding author. Email: seubank@vbi.vt.edu

[10] SIR model if the graph G is a complete graph and all edge labels are the same. In this paper, we will ignore the thermodynamic and continuum limits and focus on variability due to the graph structure. In our opinion, capturing the effects of realistic network structure is the most important motivation for introducing network models. Hence, although strictly speaking each of the items above could be considered a model, we will consider all of them to be given except the first. Thus by ‘model’ we will mean a method for selecting a graph G from among all possible graphs. Similarly, we will refer to ‘the’ dynamics as if we knew exactly how disease is transmitted from person to person.

It is clear that variations in network structure induce important variations in epidemiological outcomes, but it is difficult to estimate structural features of networks that represent millions of people. Several methods have been proposed for estimating and/or generating networks for use in epidemiological modelling [3,6–9,12]¹. These generative methods are stochastic and depend on parameters that affect structural properties of the generated networks, often in complicated ways. Associated with each method M and set of parameter values v , there is a set $\mathcal{G}(M, v)$ of constrained random graphs. That is, given a set of parameter values, each method can produce many possible networks.² In general, the more parameters, the more details of human behaviour are included and the more tightly constrained the set of possible networks. On this spectrum, less detailed methods are typically proposed as universal models of contact patterns; more detailed methods are typically used to estimate particular contact networks (Figure 1).

Refining a stochastic model by introducing additional detail has two effects:

- (1) it reduces the variability in possible outcomes, i.e. it concentrates the probability distribution
- (2) it changes the mean outcome, i.e. it moves the concentrated probability.

That is, it increases its precision and, or so the modeller hopes, its accuracy. The amount of detail in a model should be necessary and sufficient to answer questions. It is sufficient if most of the variability in outcomes arise from inherent stochasticity in the problem, not controllable stochasticity in the model; it is necessary if models with less detail cannot be made acceptably accurate.

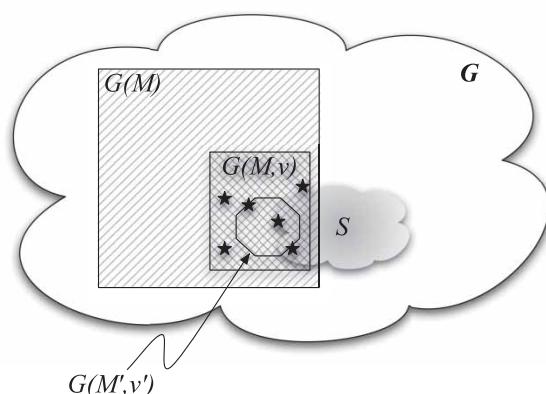


Figure 1. Model-induced variability in networks: social contact networks S are a subset of \mathcal{G} , the set of all possible graphs over a fixed number of vertices. A model M places a probability distribution $\mathcal{G}(M)$ on the space \mathcal{G} . For clarity, here the distribution $\mathcal{G}(M)$ is indicated as a subset with sharp boundaries. Calibrating M yields a particular set of parameter values v that the modeller expects will restrict networks the model generates to $\mathcal{G}(M, v)$, a subset of $\mathcal{G}(M)$ ‘close to’ S . If M' is a refinement of M that introduces additional constraints, the possible networks it generates will be a subset of $\mathcal{G}(M, v)$, as indicated by the octagon. Stars in this figure represent six different networks generated by the same model M with the same parameters v by using six different random seeds. We compare the dynamically induced variability of outcomes for a single network with the model-induced variability between networks to decide whether the model is sufficiently refined.

A natural question is *how much detail is necessary and sufficient* in a network model for understanding epidemiology in a particular population? On the one hand, *some* details of the network's structure must be represented correctly; on the other hand, if *every* detail must be correct, the task is hopeless. Without a standard for answering this question, it is difficult to argue that network models are in any sense better than mass action models, or that either can be applied to specific outbreaks and populations. We stress that our goal in this analysis is not to calibrate the algorithms, i.e. not to determine the best values for the parameters, nor to demonstrate the necessity of a certain amount of detail, but rather to develop methods to determine whether a given amount of detail is sufficient.

2. Approach

If we could organize the set of all possible networks into meaningful structural classes, then we could directly assess structural variability in \mathcal{G} . It may be that the most relevant properties are computationally hard to evaluate and thus impractical to use for categorizing graphs with millions of nodes. But worse, we do not even have an exhaustive classification scheme for structural features, much less an understanding of which features are most relevant for epidemic dynamics. This makes it difficult to determine what properties form a set of sufficient statistics for estimating features of the dynamics. For example, if degree distribution were the only important property, then the so-called 'scale-free' networks would be one category. However, although it is clearly important, it is easy to construct counter-examples in which degree distribution is not the sole, or even the most important property. This chicken-and-egg problem was pointed out by Rappaport decades ago:

It would seem offhand that a taxonomy of nets (the mathematical representations of sociograms) would arise naturally from the consideration of the statistical parameters, e.g. as a continuum of nets in the parameter space. But the statistical parameters themselves are singled out on the basis of taxonomic considerations, which have yet to be clarified [13].

Our approach does not rely on characterizing network structure directly. Instead, we study the variability in practically relevant measures of the outcome of an epidemic due to structural differences in networks. The outcome measures are defined in terms of the number of infected people as a function of time for a simple artificial disease similar to influenza. These curves tend to have a single, easily recognizable peak. We consider three outcome measures: the total fraction of the population infected, the time to peak and height of peak. All are important in practice, as they determine overall impact as well as the surge of demand for health care and the time scale on which interventions must be implemented. Variability in these outcome measures may be sensitive to details of the artificial disease we model, but we ignore those effects for simplicity. It is easy to generalize the methods presented here to other features that are tailored for particular questions. For example, if we intend to use a model for studying the impact of closing schools, we might want to consider the variability of infections for households with students. Note, however, that the *results* presented here are specific to the models studied and the outcome measures used.

We compare two important sources of variability in the outcomes of this model:

- (1) *Dynamically induced*: because the dynamics of the model are stochastic, the outcomes of simulated outbreaks vary even when all initial conditions, including the network topology, are fixed;
- (2) *Model-induced*: because the networks are generated by a stochastic model, the network topology varies, even for fixed values of the parameters.

Dynamically induced variability is inherent in the nature of disease transmission and hence is not reducible; model-induced variability can be reduced by imposing additional constraints, i.e. adding more detail, to the models we use to estimate networks. Obviously, the outcomes also vary as initial conditions or model parameters are varied, but our goal here is to understand the importance of further constraining the network construction models in light of the irreducible dynamically induced variability. We will claim that a network generation method is sufficiently detailed if the variability in outcome measures across a sample of $\mathcal{G}(M, v)$ is comparable to the inherent variability in a single instance of a network from this set.

3. Methods

Do we need to add more detail into the process we use to generate networks to constrain the results more tightly? Our network estimates are compositions of several stochastic models, described briefly below, each with several parameters that have been calibrated to data. Because the models are stochastic, we can think of the resulting networks as members of the subset \mathcal{G} of random graphs that can be produced by the model composition; because the models do not explicitly constrain structural properties, there is no simple description of \mathcal{G} , e.g. small world, scale-free, etc. In principle, we could incorporate more details and more constraints into the network estimation process. For example, if we modelled the details of contacts within the workplace, we would further constrain the possible estimated networks into a subset of \mathcal{G} . The test we construct here can tell us whether we could significantly increase the *precision* (n.b. not the *accuracy*) of estimated outcomes by adding more detail.

We generate several estimated contact networks for a large region (the state of Washington). We compare the variability of a stochastic epidemic process for a single network with the variability across networks. We argue that, because the variability between networks is comparable to the variability within a single network, there is no need to refine the models that generated the networks.

3.1. Network generation

Following Beckman *et al.* [2], we estimate a contact network in four steps:

- (1) Population synthesis (P), in which a synthetic representation of each household in a region is created;
- (2) Activity assignment (A), in which each synthetic person in a household is assigned a set of activities to perform during the day, along with the times when the activities begin and end;
- (3) Location choice (L), in which an appropriate real location is chosen for each activity for every synthetic person;
- (4) Contact estimation (C), in which each synthetic person is deemed to have made contact with a subset of other synthetic people.

The contact network is a graph whose vertices are synthetic people, labelled by their demographics, and whose edges represent contacts determined in step four, labelled by conditional probability of disease transmission.³ The overall model for constructing a network, N , can be represented as the composition of these four steps, $N = C \circ L \circ A \circ P$, in a suggestive notation that can be made formal. Each of the steps L , A , and P add new details, based on data, to our representations of synthetic people. They refine the overall model N by introducing additional constraints. Note in particular that it would be possible to construct a network at any stage using the contact estimation

model, but that as more detail is added, the number of networks that could result decreases:

$$C \circ L \circ A \circ P \subseteq C \circ A \circ P \subseteq C \circ P \subseteq C.$$

Each of the four steps makes use of a stochastic model; the first three incorporate observed data for a particular region. The synthetic population is generated using an iterative proportional fit to joint distributions of demographics taken from a census [1]. It consists of a list of demographic variables – age, gender, income, household size, education, etc. – for each of the approximately 5.77 million people in the state of Washington. Activities – school, work, shopping, other, etc. – are assigned using a decision tree model based on household demographics fitted to activity or time-use survey data [1]. This step creates a sort of personal day planner for each person in the synthetic population. Activity locations are chosen based on a gravity model and land use data [1]. That is, every synthetic activity produced in step three is assigned to an actual location in Washington – office building, school, shopping mall, etc. – as appropriate based on its distance from the person’s previous activity and its *attractiveness*, a measure of how likely it is that the activity happens there – number of employees, school enrollment, square feet of retail shopping, etc.

Contacts are estimated for each synthetic person by choosing at random from a set of other synthetic people. At its most basic level – in the absence of population synthesis, activity assignment, and location choice – C must operate on a set of unlabelled vertices. For example, C_t might generate an Erdős–Rényi graph by placing an edge between every pair of vertices with probability p , and assigning each edge the same conditional transmission probability. With the addition of population synthesis, C could take into account details of mixing between age groups, household structures, etc. In this paper, contacts are chosen at random from among other people simultaneously at the same location, and conditional transmission probabilities are assigned based on the duration of contact. Contact networks estimated in this way incorporate many structural aspects of real contact networks in a non-prescriptive, non-parametric way. That is, the networks exhibit, for example, clustering patterns that are more similar to those observed in social contact networks than those in Erdős–Rényi random graphs [5]. There is, however, no ‘knob’ with which we can tune the clustering coefficient, or any other graph measure, to achieve a particular desired value. Or more correctly, as in the real world, tuning parameters in these models simultaneously affects many features of the resulting networks’ structures.

Population synthesis uses a non-parametric, data-driven model based on US census data for the year 2000, including the Public Use Microdata Sample [4]. The particular decision tree used for activity assignment in this paper contains 12 parameters. The actual activity templates assigned to households are data-driven, based on the National Household Transportation Survey [11]. The gravity model used for location choice contains nine parameters. The locations’ addresses and attractor values are derived from Dun & Bradstreet’s Strategic Database Marketing Records. The data sets used for all the networks generated in this experiment are appropriate for the state of Washington; all parameter values for all the models have been calibrated to several different sets of observations and are held fixed for this paper. Network models of epidemiology are often criticized for the enormous number of parameters they are supposed to require, compared with the amount of epidemiological data available for calibration. We note that decomposing the problem into models for network generation and disease diffusion makes it more tractable. Here, for example, the network generation model relies on 21 parameters – admittedly a large number, but not infeasible – and is calibrated to enormous amounts of data. The resulting model is flexible enough to represent many aspects of human behaviour, but obviously not flexible enough to specify separately the weight on every possible edge between millions of people.

However, the estimates are certainly not correct in the sense that each synthetic person represents an actual person, and edges in the estimated network have a one-to-one correspondence with edges in the real network. If it were the case that only such a one-to-one match reproduced

important dynamical behaviours, it would be futile to estimate the networks, or indeed to build any mathematical models of epidemiology for particular large populations. Because epidemic processes are inherently stochastic, we optimistically assume that the realism in the results are a sigmoidal function of the realism in the estimated network: the better the network estimate, the more realistic the results – up to a point determined by the inherent variability of epidemic processes. The question we pose is not where our current estimates lie on this sigmoid, but whether adding additional detail would result in significantly different results. Is it worth refining, for example, the gravity model used to choose activity locations? The answer, in general, depends on the particular results to be gleaned from the model. For studying the spread of influenza in the workplace, one would expect a more refined model of workplaces to be necessary. We illustrate the method here using three criteria sensitive to outcomes in the general population, but it could be applied just as easily to a particular subpopulation such as school children or workers.

3.2. Epidemic process

We used a networked SEIR disease model that is similar to a strong influenza. In this model, people who are infected incubate disease for 1, 2, or 3 days with probabilities 0.3, 0.5, and 0.2, respectively. They become infectious only after the incubation period is over. They remain infectious for 3, 4, 5, or 6 days with probabilities 0.3, 0.4, 0.2, and 0.1, respectively. Incubation and infectious periods are assigned randomly to individuals who become infected. The conditional probability of transmission $p(B|A)$ from A to B is assigned based on the duration of contact Δ_{AB} between them, and is controlled with a parameter τ representing the rate of transmission:

$$p(B|A) \equiv 1 - (1 - \tau)^{\Delta_{AB}}.$$

The value for Δ between any pair of individuals is given by our network estimation algorithm. For the simulations conducted here, we chose a very strongly transmissible disease, with $\tau = 5 \times 10^{-5} \text{ min}^{-1}$, and τ is the same for every pair of people.

Each simulated outbreak must be initialized by seeding the population with infectives. The exact manner in which this is done affects the time to peak and the fraction of outbreaks that become epidemic. We have seeded these simulations with 30 infectives chosen uniformly at random from the entire population each day for five days. Thus the number, but not the identity, of the seeds is fixed for all simulations.

4. Experimental design and results

As shown in Figure 2, we have estimated twelve hierarchically nested instances of contact networks for Washington state: three instances of step one, population synthesis; two instances of step two, activity assignment, for each population; and two instances of step three, location choice, for each activity set. Each instance is a different stochastic realization of the same set of models with the same parameter values. We have not included step four, contact estimation, in this illustrative study for several reasons: it is different in character from the steps we study, because it does not introduce new data-dependent details into the synthetic individuals; it can be thought of as part of any network construction model, and so the variability it introduces cannot easily be reduced; and omitting it reduces the complexity of the analysis below. Clearly, 12 networks are not enough to sample the set of all possible random graphs over 5.77 million vertices. We have chosen this design by taking into account the effort required to produce a single network. In hindsight, perhaps because we are sampling a much smaller set, the variability across networks is so small that twelve is a reasonable number.

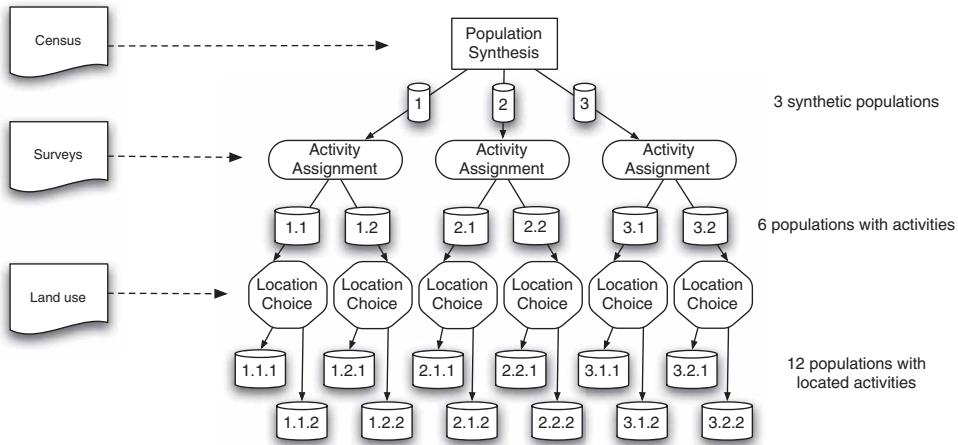


Figure 2. Creating an ensemble of contact networks. This shows the sequence of processes used to generate the networks used in this paper. Each process introduces more details about the synthetic individuals, which constrains more tightly the eventual network constructed. Edges leaving a process represent data sets generated by choosing a particular random number seed for that (stochastic) process. All other parameters and input data are held constant while generating the networks.

We computed 25 instances of the epidemic process for each of the twelve estimated contact networks. Our experience suggests that this number of realizations is adequate for estimating the variance. Figure 3(a) shows the fraction of people newly infected as a function of time, or ‘epidemic curve’ for short, for a single instance of the epidemic process on one of the networks. Indicated on the curve are the three outcome measures we consider:

- total attack rate, the area under the epidemic curve;
- peak attack rate, the height of the peak in the epidemic curve;
- time to peak, the day on which the peak attack rate occurred.

The inherent variability of the stochastic epidemic process is indicated in Figure 3(b), which shows the epidemic curves for all 25 instances of a single contact network, 1.2.1. Figure 3(c) shows the variability among mean epidemic curves for the 12 different estimated networks. Note that the variability *within* a network is much greater than that *between* the means of networks. This result by itself indicates that the amount of detail in our network construction model is sufficient for analyses that depend on the epidemic curve. However, it is instructive to show how the simulation results can be used to understand the relative contribution of each new piece of information in the L, A, P hierarchy of processes.

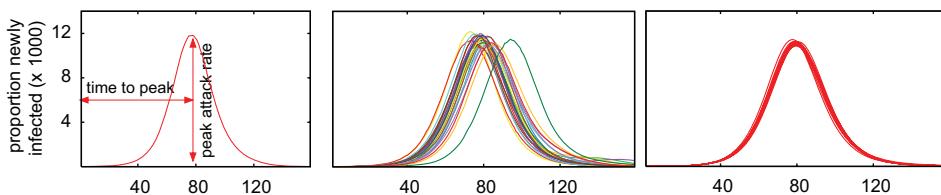


Figure 3. (a) An instance of an epidemic curve, illustrating the peak attack rate and time to peak outcome measures used in this paper. Total attack rate is the area under this curve. (b) Epidemic curves for the 25 instances of an epidemic process for contact network 1.2.1. (c) Mean epidemic curves for each of the 12 contact networks. The x-axis is in days; the y-axis is identical for all three plots.

The variability in total attack rate across all 12 estimated networks is shown in Figure 4. Results for peak attack rate and time to peak are similar. Because of the hierarchical relationship among the networks in this experiment, it is not a simple factorial design, but a nested design. The difference in the instances at any of the three steps is due to selecting a different random seed for the model. For example, networks labelled 1.1.1 and 1.2.1 used the same random number seeds in the activity location and population synthesis models, but a different seed in the activity assignment model. However, because the final network is a composition of all three steps, the difference between 1.1.1 and 1.2.1 is not isolated to activity assignment, but includes location choice. The best way to interpret the outcomes is an analysis of variance coming from three independent sources (the three steps), as shown in Table 1. The total variance of each of the outcome measures across all 25 instances/network · 12 networks = 300 simulated epidemic curves is taken to be the sum of the variance due to population synthesis, activity assignment, location choice, and the random variation due to changing the random seed in each simulation. The variances are estimated using a random effects analysis of variance linear model $y_{i,j,k,l} = P_i + A_{i,j} + L_{i,j,k} + r_{i,j,k,l}$, where $y_{i,j,k,l}$ is the measurement of interest, P_i is the effect attributed to population i , $A_{i,j}$ the effect of activity list j for population i , $L_{i,j,k}$ the effect of location k for activity list j of population i , and $r_{i,j,k,l}$ is the effect of the l th replicate.

The total variance in final attack rate is 1.932×10^{-6} , giving a standard deviation of 1.39×10^{-3} . The analysis of variance on the attack rates shows significant differences between

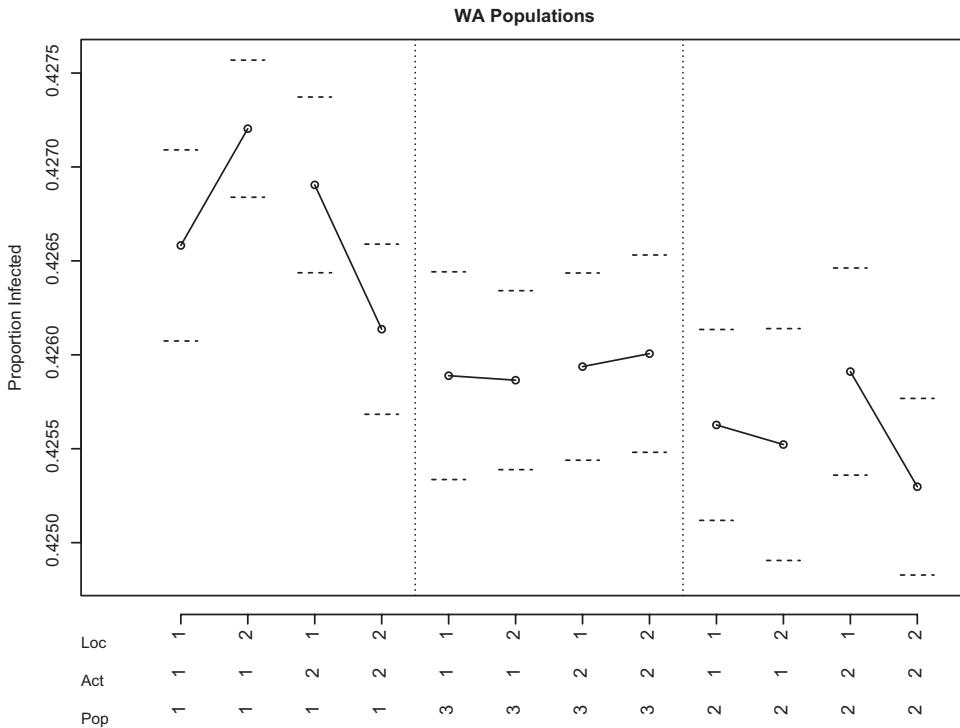


Figure 4. Variability in total attack rate across all 12 instances of contact networks. Results for peak infection rate and time to peak are similar. Each point represents the mean total attack rate across 25 epidemic curves such as are shown in Figure 3(b) for the particular network labelled on the x -axis. Error bars indicate the probable error as estimated from the 25 simulated epidemics. Thus the 12 points represent total attack rates for the 12 epidemic curves shown in Figure 3(c). This figure suggests both that the main effect on total attack rate is due to the population synthesis model P and that the size of the effect is small compared with the probable error inherent to outcomes on a single network. Both these observations are confirmed in the more formal analysis shown in the second column of Table 1.

Table 1. Analysis of contributions to the variance of attack rate, peak infection rate, and time to peak in the epidemic curve.

Factor	Attack rate	Peak infection rate	Days to peak
Total variance	1.932×10^{-6}	1.44×10^{-7}	11.53
Population synthesis	0.155	0.02	0.018
Activity assignment	0.000	0.00	0.035
Location choice	0.026	0.00	0.000
Residual variance	0.818	0.98	0.945

Note: The first row shows the total variance in each of the three outcome measures; the remaining rows show the fraction of the total variance accounted for by each factor.

populations and locations but no significant difference between activity sets. However, the standard deviation is a very small percentage of the final attack rate (about 0.3%); so these differences, while *statistically* significant, are probably not *practically* significant. The standard deviation in peak attack rate is about 3% of the mean. There are no significant differences between the populations, activities, and locations. The standard deviation in day of peak infections is about 4% of the mean. There are no significant differences in time to peak between the populations and locations, but the activity sets are significantly different ($p < 0.05$).

5. Conclusions

It is evident from Table 1 that variance in all three of the outcome measures is dominated by the inherent stochasticity of the epidemic process. Thus we conclude that we cannot significantly reduce the variability in these outcomes by further refining our network generation algorithm. At least for applications that depend on the disease model and the three outcome measures studied here, the networks produced by composing models for synthesizing a population, assigning activities, and locating those activities are sufficiently precise. As we have pointed out above, these methods do not address accuracy or necessity.

It should be noted that different aspects of the network estimation process influence the outcome measures differently. For example, final attack rate is more sensitive to activity location, L , than is time to peak. Likewise, population synthesis, P , influences final attack rate more than activity assignment, A , does. We expect that any criterion for accepting a contact network as a useful representation will be sensitive to both the exact disease model as well as the outcome measures. However, this study has demonstrated that it is possible to define (and meet) such a criterion.

Acknowledgements

We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their suggestions and comments. This work has been partially supported by NSF HSD Grant SES-0729441, NIH MIDAS project 2U01GM070694, DTRA R&D Grant HDTRA1-0901-0017, and DTRA CNIMS Grant HDTRA1-07-C-0113.

Notes

1. In addition, there is a long history of inferring structural properties of contact networks from observations of small samples.
2. \mathcal{G} is an enormous but finite set, and there is a natural measure on \mathcal{G} given by the probability that M generates an element $G \in \mathcal{G}$.
3. The contact networks we use in this work represent those of a single normative day, but this restriction is not inherent to the methodology.

References

- [1] C.L. Barrett, R. Beckman, K. Berkgigler, K. Bisset, B. Bush, K. Campbell, S. Eubank, K. Henson, J. Hurford, D. Kubicek, M. Marathe, P. Romero, J. Smith, L. Smith, P. Speckman, P. Stretz, G. Thayer, E. Van Eeckhout and M. Williams, *TRANSIMS: Transportation analysis simulation system*, Tech. Rep. LA-UR-00-1725, Los Alamos National Laboratory Unclassified Report, 2001, Vol. 3. Available at <http://ndssl.vbi.vt.edu/transims.php>.
- [2] R.J. Beckman, K.A. Baggerly, and M.D. McKay, *Creating synthetic baseline populations*, *Transp. Res. A* 30(6) (1996), pp. 415–429.
- [3] D. Brockmann, L. Hufnagel, and T. Geisel, *The scaling laws of human travel*, *Nature* 439(7075) (2006), pp. 462–465.
- [4] Census of Population and Housing, Tech. Rep. US Census Bureau, 2000. Available at <http://www.census.gov/prod/cen2000/>.
- [5] P. Erdős and A. Rényi, *On the evolution of random graphs*, in *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 1960, pp. 17–61.
- [6] S. Eubank, H. Guclu, V.S. Anil Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wan, *Modelling disease outbreaks in realistic urban social networks*, *Nature* 429 (2004), pp. 180–184.
- [7] N.M. Ferguson, D.A.T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D.S. Burke, *Strategies for containing an emerging influenza pandemic in Southeast Asia*, *Nature* 437(7056) (2005), pp. 209–214.
- [8] M.C. González, C.A. Hidalgo, and A.L. Barabási, *Understanding individual human mobility patterns*, *Nature* 453(7196) (2008), pp. 779–782.
- [9] M.E. Halloran, I.M. Longini, A. Nizam, and Y. Yang, *Containing bioterrorist smallpox*, *Science* 298(5597) (2002), pp. 1428–1432.
- [10] W.O. Kermack and A.G. McKendrick, *A contribution to the mathematical theory of epidemics*, *Proc. R. Soc. Lond.* 115 (1927), pp. 700–721.
- [11] National Household Transportation Survey, Tech. Rep., US Department of Transportation, Federal Highway Administration, 2001. Available at <http://nhts.ornl.gov/>.
- [12] R. Pastor-Satorras and A. Vespignani, *Epidemic dynamics and endemic states in complex networks*, *Phys. Rev. E* 63 (2001), 066117.
- [13] A. Rapoport and W.J. Horvath, *A study of a large sociogram*, *Behav. Sci.* 6 (1961), pp. 279–91.