Heterogeneity and Network Structure in the Dynamics of Diffusion:
Comparing Agent-Based and Differential Equation Models

Hazhir Rahmandad hazhir@mit.edu
John Sterman jsterman@mit.edu

MIT Sloan School of Management, Cambridge MA 02142

Revision of March 2007

We thank Reka Albert, Rosanna Garcia, Ed Kaplan, David Krackhardt, Marc Lipsitch, Nelson Repenning, Perwez Shahabuddin, Steve Strogatz, Duncan Watts, Larry Wein, the associate editor and referees, and seminar participants at MIT, the 2004 NAACSOS conference and 2004 International System Dynamics Conference for helpful comments. Ventana Systems and XJ Technologies generously provided their simulation software and technical support. Financial support provided by the Project on Innovation in Markets and Organizations at the MIT Sloan School.
Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models

Abstract

When is it better to use agent-based (AB) models, and when should differential equation (DE) models be used? Where DE models assume homogeneity and perfect mixing within compartments, AB models can capture heterogeneity across individuals and in the network of interactions among them. AB models relax aggregation assumptions but entail computational and cognitive costs that may limit sensitivity analysis and model scope. Because time and resources are limited, the costs and benefits of such disaggregation should guide the choice of model in policy analysis. Using contagious disease as an example, we contrast the dynamics of AB models with those of the analogous deterministic compartment DE model. We examine the impact of individual heterogeneity and different network topologies, including fully connected, random, Watts-Strogatz small world, scale-free, and lattice networks. Differences between the DE and AB dynamics are documented for several metrics relevant to public health, including diffusion speed, peak load on health services infrastructure and total disease burden. For many conditions, however, these differences are small compared to variability caused by stochastic events, uncertainty in parameters, model boundary, and network structure. We discuss implications for the choice between AB and DE models. The results apply beyond epidemiology: from innovation adoption to financial panics, many important social phenomena involve analogous processes of diffusion and social contagion.

Keywords: Agent Based Models, Networks, Scale free, Small world, Heterogeneity, Epidemiology, Simulation, System Dynamics, Complex Adaptive Systems, SEIR model
Spurred by growing computational power, agent-based modeling (AB) is increasingly applied to physical, biological, social, and economic problems previously modeled with nonlinear differential equations (DE). Both approaches have yielded important insights. In the social sciences, agent models explore phenomena from the emergence of segregation to organizational evolution to market dynamics (Schelling 1978; Levinthal and March 1981; Carley 1992; Axelrod 1997; Lomi and Larsen 2001; Axtell, Epstein et al. 2002; Epstein 2006; Tesfatsion 2002). Differential and difference equation models, also known as compartmental models, have an even longer history in social science, including innovation diffusion (Mahajan, Muller et al. 2000) and epidemiology (Anderson and May 1991).

When should AB models be used, and when are DE models appropriate? Each method has strengths and weaknesses. The importance of each depends on the model purpose. Nonlinear DE models can easily encompass a wide range of feedback effects, but typically aggregate agents into a relatively small number of states (compartments). For example, innovation diffusion models may aggregate the population into categories including unaware, aware, in the market, adopters, and so on (Urban, Houser et al. 1990; Mahajan et al. 2000). However, within each compartment people are assumed to be homogeneous and well mixed; the transitions among states are modeled as their expected value, possibly perturbed by random events. In contrast, AB models can readily include heterogeneity in individual attributes and in the network structure of their interactions; like DE models, AB models can be deterministic or stochastic and capture feedback effects.

The granularity of AB models comes at some cost. First, the extra complexity significantly increases computational requirements, constraining the ability to conduct sensitivity analysis. A second cost of agent-level detail is the cognitive burden of understanding model behavior. Linking the behavior of a model to its structure becomes more difficult as model complexity grows. Finally, limited time and resources force modelers to trade off disaggregate detail and the scope of the model boundary. Model boundary here stands for the richness of the feedback structure captured endogenously in the model (Meadows and Robinson 1985, Sterman 2000). For example, an agent-based demographic model may portray each individual in a population separately but assume exogenous fertility and mortality; such a model has a narrow boundary. In contrast, an aggregate model may lump the entire population into a single compartment, but model fertility and
mortality as functions of food per capita, health care, pollution, norms for family size, etc., each of which, in turn, are modeled endogenously; such a model has a broad boundary. DE and AB models may in principle fall anywhere on these dimensions of disaggregation and scope. In particular, there is no intrinsic limitation that prevents AB models from incorporating behavioral feedback effects or encompassing a broad model boundary. In practice, however, where time, budget, and computational resources are limited, modelers must trade off disaggregate detail and breadth of boundary. Choosing wisely is central in selecting appropriate methods for any problem.

The stakes are large. Consider potential bioterror attacks. Kaplan, Craft, and Wein (2002) used a deterministic nonlinear DE model to examine smallpox attack in a large city, comparing mass vaccination (MV), in which essentially all people are vaccinated after an attack, to targeted vaccination (TV), in which health officials trace and immunize those contacted by potentially infectious individuals. Capturing vaccination capacity and logistics explicitly, they conclude MV significantly reduces casualties relative to TV. In contrast, Eubank et al. (2004) and Halloran et al. (2002), using different AB models, conclude TV is superior, while Epstein et al. (2004), using an AB model, favor a hybrid strategy. The many differences among these models make it difficult to determine whether the conflicting conclusions arise from relaxing the perfect mixing and homogeneity assumptions of the DE (as argued by Halloran et al. 2002) or from other assumptions such as the size of the population (ranging from 10 million for the DE model to 2000 in Halloran et al. to 800 in Epstein et al.), other parameters, or boundary differences such as capacity constraints on immunization (Koopman 2002; Ferguson, Keeling et al. 2003; Kaplan and Wein 2003). Kaplan and Wein (2003) and Kaplan et al. (2003) show that their DE model closely replicates the Halloran et al. AB results when simulated with the Halloran et al. parameters, including vaccination rates and scaled-down population and initial attack size, concluding that parameterization accounts for the different conclusions, not differences in mixing and homogeneity.

Here we carry out controlled experiments to compare AB and DE models in the context of contagious disease. We choose disease diffusion for four reasons. First, the dynamics of contagion involve important characteristics of complex systems, including positive and negative feedbacks, time delays, nonlinearities, stochastic events, and individual heterogeneity. Second, network topologies linking individuals are important in the diffusion process (Davis 1991; Watts and
Strogatz 1998; Barabasi 2002; Rogers 2003), providing a strong test for differences between the two approaches. Third, the DE paradigm is well developed in epidemiology (for reviews see Anderson and May 1991 and Hethcote 2000); AB models also have a long history (e.g. Abbey 1952) and have recently gained momentum (for reviews see Newman 2002, 2003 and Watts 2004).

Finally, diffusion is a fundamental process in diverse physical, biological, social, and economic settings. Many diffusion phenomena in human systems involve processes of social contagion analogous to infectious disease, including word of mouth, imitation, and network externalities. From the diffusion of innovations to rumors, financial panics and riots, contagion-like dynamics, and formal models of them, have a rich history in the social sciences (Bass 1969; Watts and Strogatz 1998; Mahajan, Muller et al. 2000; Barabasi 2002; Rogers 2003). Insights into the advantages and disadvantages of AB and DE models in epidemiology can inform understanding of diffusion in many domains of concern to social scientists and managers.

Our procedure is as follows. We develop an AB version of the classic SEIR model, a widely used nonlinear deterministic DE model. The DE divides the population into four compartments: Susceptible (S), Exposed (E), Infected (I), and Removed (R). In the AB model, each individual is separately represented and must be in one of the four states. The same parameters are used for both AB and DE models. Therefore any differences in outcomes arise only from the relaxation of the restrictive assumptions of the DE model. Realistic DE models add more compartments to capture heterogeneity among individuals, reducing the differences between the DE and AB model (in the limit, each compartment would contain at most one individual, and the DE model becomes equivalent to the AB model). We use the classic SEIR model to maximize potential differences between the two approaches. We run the AB model under five widely used network topologies (fully connected, random, small world, scale-free, and lattice) and test each with homogeneous and heterogeneous individuals. We compare the resulting diffusion dynamics on a variety of metrics relevant to public health, including cumulative cases, peak prevalence, and the speed the disease spreads (the time available for health officials to respond).

As expected, stochastic AB models alter the dynamics compared to the deterministic DE model. Most obviously, the stochastic AB models generate a distribution of outcomes, while the deterministic DE generates a single path representing the expected trajectory under the mean-field
approximation for contacts between infectious and susceptible individuals. Further, due to chance events, the epidemic never takes off in some realizations of the stochastic model. The deterministic model cannot generate this mode of behavior (stochastic compartment models can, however, and offer an intermediate method between deterministic models and the full AB representation). More interesting are differences due to network topology. Diffusion slows as contact networks become more tightly clustered. Heterogeneity accelerates the initial take-off, as highly connected individuals quickly spread the disease, but reduces overall diffusion as these same individuals quickly exit the infectious pool.

We also examine the ability of the DE model to capture the dynamics of each network structure in the realistic situation where parameters are poorly constrained by biological and clinical data. Epidemiologists often estimate potential diffusion, for both novel and established pathogens, by fitting models to the aggregate data as an outbreak unfolds (Dye and Gay 2003; Lipsitch, Cohen et al. 2003; Riley, Fraser et al. 2003). Calibration of innovation diffusion and new product marketing models is similar (Mahajan et al. 2000). We mimic this practice by treating the AB simulations as the “real world” and fitting the DE model to them. The fitted models closely match the behavior of the AB model under all network topologies and heterogeneity conditions tested. However, the ability to fit such data does not imply that the AB and calibrated DE models will respond to policy interventions in the same way, demanding caution in their use.

The implications of the differences among the models depend on the purpose of the analysis. Here we focus on the policy context. Policymakers face a world of time pressure, inadequate data and limited knowledge of parameters such as pathogen virulence, transmissibility, incubation latency, treatment efficacy, etc. Further, the appropriate boundary for analysis is often unclear: resources for vaccination and treatment may be limited; an outbreak, whether natural or triggered by bioterror, may alter the behavior of the public and first-responders, endogenously disrupting the contact networks that feed back to condition disease spread through processes of risk amplification and attenuation (Kasperson et al. 1988, Glass and Schoch-Spana 2002). Hence we consider whether the differences in mean behavior between DE and AB models are large relative to the uncertainties policymakers face. We also consider how these differences might affect the assessment of the costs and benefits, and hence the optimality, of policies.
The mean behavior of different models may be significantly different in the statistical sense, yet be small relative to the variation in output caused by uncertainty about parameters, model boundary, and stochastic events (McCloskey and Ziliak 1996). For example, consider the variability in outcomes generated by a stochastic AB model. Each realization of the model will differ: some exhibiting fast diffusion, some slow; some with many individuals afflicted, some with fewer, depending on the chance nature of contacts between infectious and susceptible individuals. An ensemble of many simulations generates the distribution of possible epidemics, but only one will be observed in a particular outbreak. Several questions may now be asked.

One important question is whether the expected values of key metrics differ in different models. For example, does the mean value of peak prevalence under a scale-free network differ from the value generated by the corresponding deterministic compartment model? By running a sufficiently large number of simulations sampling error can be made arbitrarily small, and any differences between the mean behavior of the models will be highly statistically significant.

Another question is whether the differences among means are significant from the point of view of policymakers facing the prospect of an epidemic. Policymakers responsible for a timely and effective response must assess the practical significance of each modeling assumption given the likely range of outcomes generated by all sources of uncertainty, including in parameters, contact network structure, individual attributes, public response to the disease, and so on.

The results document a number of differences between the AB and DE models. The results also show that the differences between the deterministic compartment model, with its assumptions of homogeneous individuals and perfect mixing, and the mean behavior of the stochastic AB models are often small compared to the variability in AB outcomes caused by chance encounters among individuals, even in a perfectly specified model with a known distribution of individual attributes and network structure. The results also show that cost/benefit assessments of policy interventions, and hence the optimal policy, can depend on network structure and model boundary, highlighting the importance of sensitivity analysis across these dimensions.

Policymakers typically face tradeoffs among model boundary, sensitivity analysis, and the level of detail due to time pressure, imperfect data and uncertainty about public behavior. The results suggest that in such settings the wise course may be to use a computationally efficient
deterministic compartment model with a broad model boundary rather than a computationally intensive individual-level model with a constrained boundary and limited ability to carry out parametric and structural sensitivity and policy tests. Policymakers might also commission the development of agent-based models well in advance of a threat to help characterize the uncertainty in outbreak dynamics and explore whether policy impacts depend on different contact networks and individual characteristics.

The next section reviews the literature comparing AB and DE approaches. We then describe the models, the design of the simulation experiments, and results, closing with implications and directions for future work.

**A spectrum of aggregation assumptions:** AB and DE models should be viewed as regions in a space of modeling assumptions, not as incompatible modeling paradigms. Aggregation is one dimension of that space. Models can range from lumped deterministic differential equations (also called deterministic compartmental models), to stochastic compartmental models, in which the continuous variables of the DE are replaced by counts of discrete individuals, to event history models, where the states of individuals are tracked but their network of relationships is ignored, to models where networks of individual interactions are explicit (e.g., Koopman et al. 2001).

A few studies compare AB and DE models. Axtell et al. (1996) call for “model alignment” or “docking” and illustrate with the Sugarscape model. Edwards et al. (2003) contrast an AB model of innovation diffusion with an aggregate model, finding that the two can diverge when multiple attractors exist in the deterministic model. In epidemiology, Jacquez and O'Neill (1991) and Jacquez and Simon (1993) analyze the effects of stochasticity in individual-level SIS and SI models, finding some differences in mean behavior for small populations. However, the differences practically vanish for populations above 100. Similarly, Gani and Yakowitz (1995) examine deterministic approximations to stochastic disease diffusion processes, and find a high correspondence between the two for larger populations. Greenhalgh and Lewis (2001) compare a stochastic model with the deterministic DE version in the case of AIDS spread through needle-sharing, and find similar behavior for those cases in which the epidemic takes off.

Heterogeneity has also been explored in models that assume different mixing sites for population subgroups. Anderson and May (1991, Chapter 12) show that the immunization fraction
required to quench an epidemic rises with heterogeneity if immunization is implemented uniformly but falls if those most likely to transmit the disease are the focus of immunization. Ball et al. (1997) and Koopman, Chick et al. (2002) find expressions for cumulative cases and global epidemic thresholds in stochastic SIR and SIS models with global and local contacts, finding that the behavior of deterministic and stochastic DE models can diverge for small populations, low basic reproduction rates ($R_0$), or highly clustered contact networks where transmission occurs in mixing sites such as schools and offices. Keeling (1999) formulates a DE model that approximates the effects of spatial structure when contact networks are highly clustered. Chen et al. (2004) develop AB models of smallpox, finding the dynamics generally consistent with DE models. In sum, AB and DE models of the same phenomenon sometimes agree and sometimes diverge, especially for smaller populations. Multiple network topologies and heterogeneity conditions have not been compared, and the practical significance of differences in mean behavior relative to uncertainties in stochastic events, parameters and model boundary has not been explored.

**Model Structure:** The SEIR model is a deterministic nonlinear differential equation model in which all members of a population are in one of four states—Susceptible, Exposed, Infected, or Removed. Infected individuals can transmit the disease to susceptibles before they are “removed” (i.e., recover or die). The exposed compartment captures latency between infection and the emergence of symptoms. Depending on the disease, exposed individuals may become infectious before symptoms emerge, and can be called early-stage infectious. Typically, such individuals have more contacts than those in later stages because they are asymptomatic.

SEIR models have been successfully applied to many diseases. Additional compartments are often introduced to capture more complex disease lifecycles, diagnostic categories, therapeutic protocols, population heterogeneity and mixing patterns, birth or recruitment of new susceptibles, loss of immunity, etc. (see Anderson and May 1991 and Murray 2002 for comprehensive discussion). In this study we maintain the standard assumptions of the classic SEIR model (four stages, fixed population, permanent immunity). The DE implementation of the model imposes several additional assumptions, including perfect mixing and homogeneity of individuals within each compartment and mean field aggregation (the flows between compartments equal the expected value of the sum of the underlying probabilistic rates for each individual). To derive the
differential equations, consider the rate at which each infectious individual generates new cases:

\[ c_{is} \times \text{Prob} \left( \text{Contact with Susceptible} \right) \times \text{Prob} \left( \text{Transmission} | \text{Contact with Susceptible} \right) \]  \hspace{1cm} (1)

where the contact frequency \( c_{is} \) is the expected number of contacts between infectious individual \( i \) and susceptible individual \( s \); homogeneity implies \( c_{is} \) is a constant, denoted \( c_{IS} \), for all individuals \( i, s \). If the population is well mixed, the probability of contacting a susceptible individual is simply the proportion of susceptibles in the total population, \( S/N \). Denoting the probability of transmission given contact between individuals \( i \) and \( s \), or infectivity, as \( i_{is} \) (which, under homogeneity, equals \( i_{IS} \) for all \( i, s \)), and summing over the infectious population yields the total flow of new cases generated by contacts between the I and S populations, \( c_{IS} \times i_{IS} \times I \times (S/N) \). The number of new cases generated by contacts between the exposed and susceptibles is formulated analogously, yielding the total Infection Rate, \( f \),

\[ f = (c_{ES} \times i_{ES} \times E + c_{IS} \times i_{IS} \times I)(S/N). \]  \hspace{1cm} (2)

To model emergence and recovery, consider these to be Markov processes with certain transition probabilities. In the classic SEIR model each compartment is assumed to be well mixed so that the probability of emergence (or recovery) is independent of how long an individual has been in the E (or I) state. Denoting the individual hazard rates for emergence and recovery as \( \varepsilon \) and \( \delta \), the mean emergence time and disease duration are then \( 1/\varepsilon \) and \( 1/\delta \), respectively. Summing over the E and I populations and taking expected values yields the flows of emergence and recovery:

\[ e = \varepsilon \times E \text{ and } r = \delta \times I. \]  \hspace{1cm} (3)

The full model is thus:

\[ \frac{dS}{dt} = -f, \quad \frac{dE}{dt} = f - e, \quad \frac{dI}{dt} = e - r, \quad \frac{dR}{dt} = r. \]  \hspace{1cm} (4)

Equation (3) implies the probabilities of emergence and recovery are independent of how long an individual has been in the E or I states, respectively, and results in exponential distributions for the residence times in these states. Exponential residence times are not realistic for most diseases, where the probability of emergence and recovery is initially low, then rises, peaks and falls. Note, however, that any lag distribution can be captured through the use of partial differential equations, approximated in the ODE paradigm by adding additional compartments within the exposed and infectious categories (Jacquez and Simon 2002). For simplicity we maintain the assumption of a
single compartment per disease stage of the classic SEIR model.

The AB model relaxes the perfect mixing and homogeneity assumptions of the DE. Each individual \( j \) is in one of the four states S, E, I, or R for \( j \in (1, \ldots, N) \). State transitions depend on the individual’s state, on individual attributes such as contact frequencies, and on the chances of interactions with others as specified by the contact network among them. The aggregate flows \( f, e, \) and \( r \) are the sum of the individual transitions \( f[j], e[j], \) and \( r[j] \), which equal 1 at the time of infection, emergence, and recovery of individual \( j \) and 0 otherwise. The supplementary material details the formulation of the AB model and shows how the DE model can be derived from it by assuming homogeneous agents and applying the mean-field approximation.

A key parameter in epidemic models is the basic reproduction number, \( R_0 \), the expected number of new cases each contagious individual generates before removal, assuming all others are susceptible. The base case transmission rates and stage durations we selected yield a mid-range \( R_0 \) of 4.125 (Table 1), similar to diseases like smallpox, \( R_0 \approx 3–6 \) (Gani and Leach 2001), and SARS, \( R_0 \approx 2-7 \) (Lipsitch, Cohen et al. 2003; Riley, Fraser et al. 2003). The base value provides a good opportunity to observe potential differences between DE and AB models: diseases with \( R_0 < 1 \) die out quickly and pose little risk to public health, while those with \( R_0 \gg 1 \), e.g. chickenpox and measles, generate a severe epidemic in (nearly) any network. The AB models use the same values for infectivity and expected residence times, and we choose individual contact frequencies so that mean total contact rates in each network and heterogeneity condition are the same as the DE model. We set the population \( N = 200 \), all susceptible except for two randomly chosen exposed individuals. Though small compared to settings of interest in policy design, e.g., cities, the effects of random events and network type are likely to be more pronounced in small populations (Gani and Yakowitz 1995), providing a stronger test for differences between the DE and AB models. A small population also reduces computation time, allowing more extensive sensitivity analysis. The DE model has 4 state variables; computation time is negligible for all \( N \). The AB model has \( 4*N \) states and must also track interactions among the \( N \) individuals, implying that computation time can grow at rates up to \( O(N^2) \), depending on the contact network. We report sensitivity analysis of \( R_0 \) and \( N \) below. The supplement includes the models and full documentation.

**Experimental design:** We vary both the network structure of contacts among individuals and the
degree of individual heterogeneity in the AB model and compare the resulting dynamics to the DE. We implement a full factorial design with five network structures and two heterogeneity conditions. In each of the ten conditions we generate an ensemble of 1000 simulations of the AB model, each with different realizations of the random variables determining contacts, emergence, and recovery. Since the expected values of parameters in each simulation are identical to the DE model, differences in outcomes can only be due to differences in network topology, heterogeneity among individuals, or the discrete, stochastic treatment of individuals in the AB model.

Network topology: The DE model assumes perfect mixing, implying anyone can meet anyone else with equal probability. Realistic networks are far more sparse and clustered. We explore five different network structures: fully connected, random (Erdos and Renyi 1960), small-world (Watts and Strogatz 1998), scale-free (Barabasi and Albert 1999), and lattice (where contact only occurs between neighbors on a ring). We parameterize the model so that all networks (other than the fully connected case) have the same mean number of links per node, \( k = 10 \) (Watts 1999).

The fully connected network corresponds to the perfect mixing assumption of the DE model. The random network is similar except people are linked with equal probability to a subset of the population. To test the network most different from the perfect mixing case, we also model a one-dimensional ring lattice with no long-range contacts. With \( k = 10 \) each person contacts only the five neighbors on each side. The small world and scale-free networks are intermediate cases with many local and some long-distance links. These widely-used networks characterize a number of real systems (Watts 1999; Barabasi 2002). We set the probability of long-range links in the small world networks to 0.05, in the range used by Watts (1999). We build the scale-free networks using the preferential attachment algorithm (Barabasi and Albert 1999) in which the probability a new node links to existing nodes is proportional to the number of links each node already has. Preferential attachment yields a power law for the probability that a node has \( k \) links, \( \text{Prob}(k) = \alpha k^{-\gamma} \). Empirically \( \gamma \) typically falls between 2 and 3; the mean value of \( \gamma \) in our experiments is 2.6. Scale free networks contain a few highly connected “hubs” and many nodes with few links.

Two of the networks are deterministic (the fully connected and lattice) so every simulation of these cases has the same network governing contacts among individuals. The Erdos-Renyi, small world, and scale-free cases are random networks. Each simulation of these cases uses a
different realization of the network structure. Though realistic networks change through time, we assume the network realization in each simulation is fixed to maximize the differences between the AB and DE conditions (a conservative assumption, since changing network connections during an epidemic introduces mixing that brings the AB model closer to the assumptions of the compartment model). The supplement details the construction of each network.

Individual Heterogeneity: Each individual has four relevant characteristics: expected contact rate, infectivity, emergence time, and disease duration. In the homogeneous condition \((H=)\) each individual is identical with parameters set to the values of the DE model. In the heterogeneous condition \((H\neq)\) we vary individual contact frequencies.

Heterogeneity in contacts is modeled as follows. Given that two people are linked (that they can come into contact), the frequency of contact between them depends on two factors. First, how often does each use their links, on average: some people are gregarious; others shy. Second, time constraints may limit contacts. At one extreme, the frequency of link use may be constant, so that people with more links have more total contacts per day, a reasonable approximation for some airborne diseases and easily communicated ideas: a professor may transmit an airborne virus or a simple concept to many people with a single sneeze or comment, (roughly) independent of class size. At the other extreme, if the time available to contact people is fixed, the chance of using a link is inversely proportional to the number of links, a reasonable assumption when transmission requires extended personal contact: the professor can only tutor a limited number of people each day. We capture these effects by assigning individuals different propensities to use their links, \(\lambda[j]\), yielding the expected contact frequency for the link between individuals \(i\) and \(j\), \(c[i,j]\):

\[
c[i,j]=\kappa*\lambda[i]*\lambda[j]/(k[i]*k[j])
\]

where \(k[j]\) is the total number of links individual \(j\) has, \(\tau\) captures the time constraint on contacts, and \(\kappa\) is a constant chosen to ensure that the expected contact frequency for the population equals the mean value used in the DE model. In the homogeneous condition \(\tau = 1\) and \(\lambda[j] = 1\) for all \(j\) so that in expectation all contact frequencies are equal, independent of how many links each individual has. In the heterogeneous condition, those with more links have more contacts per day \((\tau = 0)\), and individuals have different propensities to use their links. We use a uniform
distribution with a large range, $\lambda[j] \sim \text{U}[0.25, 1.75]$.

**Calibrating the DE Model:** In real world applications the parameters determining $R_0$ are often poorly constrained by biological and clinical data. For emerging diseases such as vCJD, BSE and avian flu data are not available until the epidemic has already spread. Parameters are usually estimated by fitting models to aggregate data as an outbreak unfolds; SARS provides a typical example (Dye and Gay 2003; Lipsitch, Cohen et al. 2003; Riley, Fraser et al. 2003). Because $R_0$ also depends on contact networks that are often poorly known, models of established diseases are commonly estimated the same way (e.g., Gani and Leach 2001). To mimic this protocol we treat each realization of the AB model as the “real world” and estimate the parameters of the DE to yield the best fit to the cumulative number of cases. We estimate infectivity ($i_{ES}$ and $i_{IS}$), and incubation time ($1/\varepsilon$) by nonlinear least squares in a large set of individual AB realizations (see the supplement). Results assess whether calibrated DE models can capture the behavior of heterogeneous individuals in realistic settings with different contact networks.

**Results:** For each experimental condition we examine three measures relevant to public health. The maximum symptomatic infectious population (**peak prevalence**, $I_{max}$) indicates the peak load on public health infrastructure including health workers, immunization resources, hospitals and quarantine facilities. The time from initial exposure to the maximum of the infected population (the **peak time**, $T_p$) measures how quickly the epidemic spreads and therefore how long officials have to deploy those resources. The fraction of the population ultimately infected (the **final size**, $F$) measures the total burden of morbidity and mortality. To illustrate, figure 1 compares the base case DE model with a typical simulation of the AB model (in the heterogeneous scale-free case). The sample scale-free epidemic grows faster than the DE ($T_p = 37$ vs. 48 days), has similar peak prevalence ($I_{max} = 27\%$), and ultimately afflicts fewer people ($F = 85\%$ vs. 98\%).

In this study we focus on the practical significance of differences between the mean output of AB and DE models. Specifically, we explore whether the differences among models are large relative to the variability in outcomes for which policymakers should plan and whether the differences alter the choice of optimal policies. To begin, we conservatively consider outcome variability arising only from stochastic interactions among individuals. Specifically, suppose
policymakers planning for a possible outbreak know with certainty mean infectivity, incubation period, disease duration, network type, and all other parameters conditioning contagion and diffusion, and that these characteristics are unaffected by the course of the epidemic. In short, assume policymakers possess a perfect agent-based model of the situation, and lack only knowledge of which individuals will, by chance, encounter each other at any moment and transmit the disease. As an example, suppose the contact network is characterized by a scale-free degree distribution with known parameters, and that individuals are highly heterogeneous in their behavior (but with known distribution). For the hypothetical disease we examine, prevalence peaks on average after 44 days at a mean of 23.9% of the population. In the deterministic compartment model with the same disease parameters, prevalence peaks after 48 days at 27.1% of the population. Given the large sample of AB realizations, these differences are statistically significant (p < .001), but they are not practically significant. Unobservable stochastic interactions among individuals means policymakers, to be, for example, 95% confident resources will be sufficient, must plan to handle an epidemic peaking between 4 and 75 days after introduction, with peak prevalence between 4% and 31.5% of the population.

Policymakers should also consider how model assumptions affect the optimality of interventions. Consider, for example, a quarantine policy. Quarantine should be implemented if its benefit/cost ratio (e.g. the value of QALYs or DALYs saved and avoided health costs relative to the costs of quarantine implementation), is favorable and higher than that of other policy options (including no action). Two models may yield similar estimates of epidemic diffusion, yet respond differently to policies. In such cases the differences between the models may be of great practical significance even if their base case behavior is similar. We provide an example below.

Figure 2 shows the symptomatic infectious population, I, in 1000 AB simulations for each network and heterogeneity condition. Also shown are the mean of the ensemble and the trajectory of the base case DE model. Table 2 reports the results for the fitted DE models; Tables 3-4 report $T_p$, $I_{\text{max}}$, and F for each condition and compare them to the base and fitted DE models. Except for the lattice, the AB dynamics are qualitatively similar to the DE. Initial diffusion is driven by positive feedback as contagious individuals spread the infection. The epidemic peaks when the susceptible population is sufficiently depleted that the (mean) number of new cases generated by
contagious individuals is less than the rate at which they are removed from the contagious pool.

Departures from the DE model increase from the connected to the random, scale free, small world, and lattice structures (Figure 2; tables 3-4). The degree of clustering explains some of these variations. In the fully connected and random networks the chance of contacts in distal regions is the same as for neighbors. The positive contagion feedback is strongest in the connected network because an infectious individual can contact everyone else, minimizing local contact overlap. In contrast, the lattice has maximal clustering. When contacts are localized in a small region of the network, infectious individuals contact their common neighbors repeatedly. As these people become infected the chance of contacting a susceptible and generating a new case declines, slowing diffusion, even if the total susceptible population remains high.

In the deterministic DE model there is always an epidemic if $R_0 > 1$. Due to the stochastic nature of interactions in the AB model, it is possible that no epidemic occurs or that it ends early if, by chance, the few initially contagious individuals recover before generating new cases. As a measure of early burnout, table 3 reports the fraction of cases where cumulative cases remain below 10%. (Except for the lattice, the results are not sensitive to the 10% cutoff. The online appendix shows the histogram of final size for each network and heterogeneity condition.) Early burnout ranges from 1.8% in the homogeneous connected case to 6.8% in the heterogeneous scale-free case. Heterogeneity raises the incidence of early burnout in each network since there is a higher chance that the first cases will have few contacts and recover before spreading the disease. Network structure also affects early burnout. Greater contact clustering increases the probability that the epidemic burns out in a local patch of the network before it can jump to other regions, slowing diffusion and increasing the probability of global quenching.

Heterogeneity results in slightly smaller final size, $F$, in all conditions: the mean reduction over all ten conditions is 0.10, compared to a mean standard deviation in $F$ across all conditions of 0.19. Similarly, heterogeneity slightly reduces $T_p$ in all conditions (by a mean of 9.5 days, compared to a mean standard deviation of 26 days). Maximum prevalence also falls in all conditions (by 1.5%, compared to a mean standard deviation of 5.1%). In heterogeneous cases the more active individuals tend to contract the disease sooner, causing a faster take-off compared to the homogeneous case (hence earlier peak times). These more active individuals also recover
sooner, reducing mean contact frequency, and hence the reproduction rate, among those who remain compared to the homogeneous case. Subsequent diffusion is slower, peak prevalence is smaller, and the epidemic ends sooner, yielding fewer cumulative cases.

Consider now the differences between the DE and AB cases by network type.

**Fully Connected:** The fully connected network corresponds closely to the perfect mixing assumption of the DE. As expected, the base DE model closely tracks the mean of the AB simulations. In the $H=0$ condition, $T_p$, $I_{\text{max}}$, and $F$ in the base DE model fall well within the 95% confidence interval defined by the ensemble of AB simulations. In the $H \neq 0$ case, $T_p$ and $I_{\text{max}}$ also fall within the 95% range, but $F$ lies just outside the range encompassing 95% of the ensemble.

**Random:** The random network behaves much like the connected case. The DE values of $T_p$ and $I_{\text{max}}$ fall within the 95% outcome range for both heterogeneity conditions. The value of $F$ in the DE falls outside the 95% range for both $H=0$ and $H \neq 0$, because the sparse contact network means more people escape contact with infectious individuals compared to the perfect mixing case.

**Scale-Free:** The scale free network departs substantially from perfect mixing. Most nodes have few links, so initial takeoff is slower, but once the infection reaches a hub it spreads quickly. The base DE values of $T_p$ and $I_{\text{max}}$ fall well within the 95% outcome interval for both heterogeneity conditions. However, as the hubs are removed from the infectious pool, the remaining nodes have lower average contact rates, causing the epidemic to burn out at lower levels of diffusion; the 95% range for final size is 2% to 98% for $H=0$ and 1% to 92% for $H \neq 0$, while the base DE value is 98%.

**Small World:** Small world networks are highly clustered and lack highly-connected hubs. Nevertheless, the presence of a few long-range links is sufficient to seed the epidemic throughout the population (Watts and Strogatz 1998). Diffusion is slower on average compared to the DE and the connected, random, and scale-free networks. The existence of a few randomly placed long-range links also increases the variability in outcomes. The 95% range for $T_p$ is 22 to 154 days for $H=0$ (7 to 176 days for $H \neq 0$), easily encompassing the base DE value. Slower diffusion relative to the DE causes peak prevalence in the DE to fall outside the 95% interval of AB outcomes for both $H=0$ and $H \neq 0$. The main impact of heterogeneity is greater dispersion and a reduction in final size.

**Lattice:** In the lattice individuals only contact their k nearest neighbors, so the epidemic advances
roughly linearly in a well-defined wave of new cases trailed by symptomatic and then recovered individuals. Such waves are observed in the spread of plant pathogens, where transmission is mostly local, though in two dimensions more complex patterns are common (Bjornstad et al. 2002; Murray 2002). Because the epidemic wave front reaches a stochastic steady state in which removal balances new cases, the probability of burnout is roughly constant over time, and $I_{\text{max}}$ is lower, with the base DE value falling outside the 99% range. For the same reason, mean final size is much lower and peak time longer than the base DE. Interestingly, the variance is higher as well, so that in the $H_\text{=}^{-}$ condition the DE values of $F$ and $T_p$ fall within the 95% range of AB outcomes.

In sum, peak time in the uncalibrated base DE model falls within the envelope encompassing 95% of the AB simulations in all ten network and heterogeneity conditions. Peak prevalence falls within the 95% range in all but the small world and lattice, and within the 99% range for all but the lattice. Final size, however, is sensitive to clustering and heterogeneity, falling within the 95% range in only three cases.

*Calibrated DE Model:* In practice parameters such as $R_0$ and incubation times are poorly constrained and are estimated by fitting models to aggregate data. Table 2 summarizes the results of fitting the DE model to 200 randomly selected AB simulations in each experimental condition, a total of 2000 calibrations. The differences between the AB and best-fit DE models are very small. The median $R^2$ for the fit to cumulative cases exceeds 0.985 in all scenarios. The mean values of $F$, $T_p$, and $I_{\text{max}}$ in the calibrated DE all fall within the range encompassing 95% of the AB outcomes in all network and heterogeneity conditions. The DE model fits well even though it is clearly mis-specified in all but the homogeneous fully connected network. Why? As the network becomes increasingly clustered and diffusion slows, the estimated parameters adjust accordingly.

Specifically, in deterministic SEIR compartment models, $R_0$ and final size are related by $R_0 = -\ln(1 - F)/F$ (Anderson and May 1991). Consequently, in heterogeneous or clustered networks, where $F$ is smaller, the estimated $R_0$ is also smaller as the estimated infectivity or incubation time fall. The calibration results suggest that very simple structures can capture the behavior produced by diverse network and heterogeneity conditions. The close fit of the compartment model, however, does not imply that its response to policies will be the same as that of the underlying clustered and heterogeneous network. The supplement provides further details.
**Sensitivity to Population Size:** We repeated the analysis for N = 50 and 800 (see the supplement). The results change little over this factor of 16. For most conditions, the rate of early burnout falls in the larger population, so the final fraction of the population infected is slightly larger (and therefore closer to the value in the DE). Population size has little impact on the other metrics.

**Sensitivity to $R_0$:** We varied $R_0$ from 0.5 to 2 times the base value; detailed results are reported in the supplement. Naturally, diffusion is strongly affected by $R_0$. Somewhat surprisingly, however, the practical significance of the differences between the DE and AB models is not highly sensitive to $R_0$. Changes in $R_0$ have two offsetting effects. First, the smaller the value of $R_0$, the larger are the differences between the DE and means of the AB trajectories. Second, however, the smaller the value of $R_0$ the greater the variation in outcomes within a given network and heterogeneity condition caused by chance encounters among individuals. Small values of $R_0$ reduce the expected number of new cases each infectious individual generates before recovering. In effect, the fraction of the contact network sampled by each infectious individual is smaller, so the probability that the epidemic will be seeded at multiple points in the network decreases. In highly clustered and heterogeneous networks, the lower representativeness of these small samples increases the difference between the DE and the mean of the AB trajectories (for example, more cases of early quenching will be observed). For the same reason, however, individual realizations of the same network and heterogeneity condition will differ more with small $R_0$, increasing the variance in outcomes for which policy makers must prepare. Similarly, larger $R_0$ reduces the differences between the DE model and the means of the AB models but also reduces variability in outcomes because each infectious individual samples the network many times before recovering. These offsetting effects imply that, over the range examined here, the differences between the DE and the mean behavior of the AB models are relatively insensitive to variations in $R_0$.

**Sensitivity to disease natural history:** In many diseases the exposed gradually become more infectious prior to becoming symptomatic. This progression can be modeled by adding additional compartments to the exposed stage with different infectivities in each. In the classic SEIR model used here, with only one compartment per disease stage, pre-symptomatic infectivity is approximated by assuming the exposed can also transmit the disease, though with $i_{ES} < i_{IS}$. To
examine the impact of this assumption we set \( i_{ES} = 0 \), adjusting \( i_{IS} \) to keep \( R_0 \) at its base value. Results are reported in the supplement. As expected, diffusion slows and the probability of early quenching increases. However, the practical significance of the differences in the mean values of the metrics across models is essentially unchanged compared to the base case. The results do not appear to be sensitive to the assumption that exposed individuals can transmit the disease.

Policy Analysis and Sensitivity to Model Boundary: Another important question is whether the behavior of the models differs in response to policy interventions and expansion of the model boundary. While comprehensive policy analysis is beyond the scope of this paper, we illustrate by examining the impact of actions that reduce contact rates. For example, the 2003 SARS epidemic appears to have been quenched through contact reduction (Wallinga and Teunis 2004). Contact reduction can arise from both policies, e.g., quarantine (including mandatory isolation and travel restrictions), and from behavioral feedbacks, e.g., social distancing, where individuals who fear infection voluntarily reduce contacts with others. For simplicity we assume contact rates fall linearly to a minimum value as the total number of confirmed cases (cumulative prevalence \( P = I + R \)) rises.\(^1\) Specifically, we model the contact frequency \( c_{js} \) between infectious persons, \( j \in \{E, I\} \), and susceptibles, \( s \in \{S\} \), as a weighted average of the initial frequency, \( c_{js}^* \), and the minimum achieved under quarantine, \( c_{js}^q \):

\[
c_{js} = (1 - q)c_{js}^* + qc_{js}^q
\]

\[
q = \text{MIN}\{1, \text{MAX}(0, (P - P_0)/(P_q - P_0))\}
\]

The impact of contact reduction, \( q \), rises linearly as cumulative prevalence, \( P \), rises from a threshold, \( P_0 \), to the level at which the effect saturates, \( P_q \). We set \( P_0 = 2 \) and \( P_q = 10 \) cases. Neither social distancing nor quarantine are perfect; we set the minimum contact frequency, \( c_{js}^q = 0.15c_{js}^* \). This value gradually reduces \( R_0 \) in the DE model from 4.125 to \( \approx 0.6 \), roughly similar to the reduction Wallinga and Teunis (2004) estimate for the SARS epidemic.

As expected, contact reduction quenches the epidemic earlier. In the DE model, prevalence

\(^1\) Other policies, such as targeted immunization, can exploit the structure of the contact network, if it is known, and generally require an AB model, though some such policies can be approximated in DE models (e.g., Kaplan, Craft and Wein 2003).
peaks 17 days sooner, \( I_{\text{max}} \) falls from 27% to 4.4%, and \( F \) falls from 98% to 19% of the population, greatly easing the burden on public health resources. Contact reduction has similar benefits in the AB cases. The differences between the means of the metrics in the AB models and their DE value are small relative to the variation in outcomes caused by stochastic interactions in the AB models. The DE results fall within the 95% outcome range for all three metrics in all network and heterogeneity conditions, with one exception: the value of \( F \) in the lattice (Table 5). Note however, that, as in the base case, clustering and heterogeneity cause some differences between the DE and AB models. In the base case heterogeneity lowers \( F \) because high-contact individuals tend to become infected early and are thus removed early, lowering the reproduction rate. In contrast, under contact reduction \( F \) increases with heterogeneity because high-contact individuals increase the number of exposed individuals before contact reduction is triggered. Therefore the reduction in \( F \) under contact reduction smaller in the heterogeneous cases.

The impact of heterogeneity illustrates an important aspect of policy analysis. Policies should be implemented if their cost-benefit ratio is favorable compared to other options, including no action. As a simple illustration, suppose the per-capita costs of mandatory contact reduction policies (quarantine, travel restrictions), denoted \( C \), are fixed and that the benefits are proportional to avoided cases, \( \Delta F = F_{\text{no policy}} - F_{\text{policy}} \). Ignoring uncertainty, mandatory measures should be implemented if \( b\Delta F > C \), where \( b \) is the benefit per avoided case. In the scale-free case, \( \Delta F = 0.75 \) for the \( H_= \) case but falls to 0.59 in the \( H_\neq \) condition (see supplement exhibit 12). For \( 0.59 \leq C/b \leq 0.75 \), whether mandatory measures are indicated depends on whether the population is homogeneous or not. The width of this interval of policy sensitivity depends also on the network type (and of course on parameters such as \( R_0 \)).

The size of the region of policy sensitivity also depends on the model boundary. For example, if awareness of the epidemic arising from, e.g., media reports causes individuals to engage in social distancing spontaneously, contacts will fall even without quarantine and travel restrictions, reducing the benefits of mandatory measures. If spontaneous social distancing reduces \( R_0 \) persistently below one, mandatory measures would not be needed to quench the epidemic and would not be justified on cost-benefit grounds. At the other extreme, if the public’s reaction to media reports were panic and flight, increasing the risk of long-range transmission,
mandatory control measures would have even higher benefits relative to their costs. Thus policymakers should carry out sensitivity analysis not only over uncertainty in parameters, network topology and individual characteristics, but over variations in the strength of behavioral feedback effects (that is, over a wide model boundary). Additional resources for empirical work to reduce uncertainty and to improve the model should be allocated where they have the greatest value. Making such judgments rationally requires the resources and time to carry out sensitivity analysis for each policy option across all relevant dimensions of uncertainty.

**Discussion and conclusions:** Stimulated by advances in computation, agent-based simulation is growing in popularity. Still, no matter how powerful computers become, limited time, budget, cognitive capabilities, and decision-maker attention mean modelers always face tradeoffs: should they disaggregate to capture the diverse attributes of individuals, expand the model boundary to capture additional feedback processes, or keep the model simple so that it can be analyzed thoroughly? Agent-based models allow detailed representation of individuals and the networks of relationships among them, but increase computational requirements. Traditional compartment models are computationally efficient, but assume perfect mixing and homogeneity within compartments. By contrasting agent-based and deterministic compartment models of epidemic diffusion we assess the importance of relaxing the perfect mixing and homogeneity assumptions.

As expected, network topology and individual heterogeneity affect the dynamics. Higher clustering increases the overlap in contacts among neighbors and therefore slows diffusion to other regions, leading to lower peak prevalence and higher peak times in the small-world and lattice networks. Heterogeneity in individual contact rates causes slightly earlier peak times as high-contact individuals rapidly seed the epidemic, followed by lower diffusion levels as the high-contact individuals are removed, leaving those with lower average transmission probability and a smaller reproduction rate. These results are consistent with analysis of heterogeneity for SI and SIS models (Veliov 2005). Such dynamics were also observed in the HIV epidemic, where initial diffusion was rapid in subpopulations with high contact rates. Such heterogeneity is often captured in DE models by adding additional compartments to distinguish highly and weakly connected types. Finally, in the AB models the epidemic fizzles out in a small fraction of cases even though the underlying parameters yield an expected value for the basic reproduction rate greater than one.
The more highly clustered and heterogeneous the population, the greater is the incidence of early quenching. The deterministic DE model cannot generate such behavior.

Before turning to implications, we consider limitations and extensions. The experiments examined the classic SEIR model. Further work should address the robustness of results to common elaborations such as loss of immunity, nonexponential distributions for emergence and recovery, recruitment of new susceptibles, non-human disease reservoirs and vectors, etc. Note, however, that by using the classic SEIR model, with only four compartments, we maximize the difference between the aggregation assumptions of the DE and AB representations. In practical applications DE models often disaggregate the population more finely to account for heterogeneity arising from, e.g., sex, age, behavior, location, contact frequencies, mixing and other attributes that vary across population segments and cause clustering in contact networks. Such disaggregation further reduces the differences between the DE and AB representations.

Though we examined a wide range of networks, the AB models contain many parameters that could be subject to additional sensitivity analysis, including the mean number of links per node, the probability of long-range links (in the small world network), and the scaling exponent (in the scale-free case). Other dimensions of heterogeneity and other networks could be examined, including networks derived from field study (Ahuja and Carley 1999). The number and distribution of the initially infectious individuals can be varied. The robustness of other policies with respect to network type, heterogeneity and model boundary should be examined. The boundary could be expanded to include endogenously the many effects that alter contact rates and network structure as an epidemic progresses (relaxing the fixed network assumption).

The deterministic DE model does not capture the variability in outcomes caused by stochastic events and yields a point value for any metric (for a given set of parameters). The costs and benefits of options facing policymakers, however, often depend on the distribution of possible outcomes, not only expected values. It is not appropriate to use a single parameter set in a deterministic model to answer policy questions such as “is hospital capacity sufficient to handle an outbreak?” Stochastic compartment models, along with individual level AB models, can capture the distribution of outcomes generated by random interactions among individuals and should be tested against the full AB models. Indeed, given the many sources of uncertainty in realistic
settings it would be irresponsible to use a single set of assumptions in any model, deterministic or stochastic, compartment or agent-based. The outcome distribution in the AB models only captures uncertainty arising from stochastic events; sensitivity to parameters may be larger and should be examined. For example, $R_0$ for smallpox is estimated to be in the range $3 - 6$ (Gani and Leach 2001). Varying $R_0$ over that range in the DE (by scaling both contact rates, $c_{ES}$ and $c_{IS}$, proportionately) causes $F$ to vary from 94.1% to 99.7%, $I_{\text{max}}$ to vary from 21.7% to 31.8%, and $T_p$ to vary from 36 to 64 days, comparable to the differences caused by relaxing the perfect mixing and homogeneity assumptions. Such uncertainty further highlights the importance of wide-ranging parametric and structural sensitivity tests for all models.

Finally, the results may inform phenomena beyond epidemics. Processes of social contagion (imitation, word of mouth, etc.) play important roles in many social and economic phenomena, from marketing to crowd behavior (Strang and Soule 1998; Rogers 2003). Models of diffusion in such contexts are similar to the SEIR family, most notably the Bass (1969) model and its extensions (e.g. Mahajan, Muller et al. 2000). Moreover, modelers tackling policy issues related to innovation and product diffusion face tradeoffs in the choice of modeling assumptions similar to those studying epidemics (Gibbons 2004). We do not explicitly address the differences between AB and DE models in these contexts; such issues are an important arena for future work.

The results demonstrate a number of differences between the deterministic compartment model and the individual-level models, and across the AB models with different network and heterogeneity assumptions. The significance of these differences depends on the purpose of the model. Here we focus on policy analysis where resources are limited and policymakers must make tradeoffs among the level of detail, the breadth of the model boundary, and the ability to carry out sensitivity analysis. As expected, the differences between the mean behavior of the agent-based models and the deterministic compartment model are statistically significant when the homogeneity and perfect mixing assumptions of the compartment model are strongly violated. However, these differences may have little practical significance in some settings. The differences in the peak burden on public health resources and the time available to deploy those resources between the DE and the mean of the AB models are small relative to the variability in outcomes caused by unobservable stochastic interactions among individuals for the connected, random, small
world, and scale-free networks. The DE values of these metrics generally fall within the envelope capturing 95% of the AB realizations, not only in the base case but across a range of assumptions about $R_0$, population size, and disease lifecycle. The main exception is the ring lattice, where there are no long-range contacts. However, a pure lattice is unrealistic in modeling human diseases due to the high mobility of modern society (though it may be appropriate in modeling immobile plant populations or where transmission to humans arises only from geographically constrained vectors). The results show that a policymaker with a perfectly specified model, including exact knowledge of disease parameters, the distribution of individual attributes and the topology of the contact network, would nevertheless find that the variation in outcomes caused by unobservable stochastic interactions among individuals is often larger than the differences between the deterministic compartment model and the mean behavior of the agent-based models, at least for the public health metrics considered here.

Of course models are not perfectly specified. Since parameters characterizing emerging (and some established) diseases are poorly constrained, epidemiologists typically fit models to aggregate data. We tested the impact of this protocol by treating the realizations of the AB model as the “real world” and fitting the DE model to them. The calibrated compartment model captures the dynamics well, with the median $R^2$ exceeding 0.985 in all conditions. The means of the public health metrics for the calibrated models fell within the 95% confidence range defined by the ensemble of AB simulations in all network and heterogeneity conditions tested. These results suggest that simple DE models can capture a wide range of variation in network structure and individual attributes. The results are potentially important because the DE model is computationally efficient. Where sensitivity analyses over uncertainties in network topology, individual heterogeneity, parameters, social and behavioral feedbacks and other extensions to the model boundary are prohibitively time consuming in full AB models with realistic populations, DE models, disaggregated to capture key aspects of clustering and heterogeneity and calibrated to past outbreaks, might be a computationally efficient approximate method to carry out policy and sensitivity analysis. Caution must be exercised, however: as shown in the supplement, the ability to reproduce historical data does not mean the simple compartment models will respond to policies the same way the AB models do. A major challenge for future work is optimally choosing the
number and definitions of compartments to capture the impact of clustering and heterogeneity. AB models may be used effectively to design DE models that capture heterogeneity and network effects using the fewest additional compartments—if the network structure is reasonably well known and stable. Testing this proposal is beyond the scope of this paper.

The calibration results also highlight an important methodological issue. The parameter values obtained by fitting the aggregate model to the data from an AB simulation (and therefore from the real world) do not necessarily equal the mean of the individual-level parameters governing the interactions among individuals. Aggregate parameter estimates not only capture the mean of individual attributes such as contact rates but also the impact of heterogeneity and network structure. Modelers often use both micro and aggregate data to parameterize both AB and DE models. The estimation results suggest caution must be exercised in doing so, and in comparing parameter values across different models (Fahse, Wissel et al. 1998).

How, then, should modelers choose among AB and compartment models (either deterministic or stochastic)? The purpose, situation, time, and resources available condition the choice. AB models enable analysts to examine questions not easily modeled in the DE paradigm, e.g. creating and removing nodes and links to simulate random failures or targeted attacks. AB models can show how aggregate behavior emerges from interactions among the elements of the system (e.g. Reynolds 1987), allow for more realistic representation and analysis of stochastic behavior in a population, and extend theoretical understanding by identifying instances where DE representations cannot generate certain behaviors (e.g. Shnerb, Louzoun et al. 2000).

Data availability significantly affects model choice. AB models will be useful when data or the underlying “physics” of a situation specify the network structure, suggest it is critical in the results, and that structure is stable over the time horizon of interest. Often, though, data on contact networks and the distribution of individual attributes are often hard to obtain and highly uncertain, requiring extensive sensitivity analysis to ensure robust results. Disaggregating to an AB representation without data to support the choice of network structure and individual heterogeneity increases the computational cost and dimensionality of sensitivity analysis but for many cases adds little value.

2 We thank an anonymous reviewer for this suggestion.
In this study we focused on the practical significance of difference among models. We compared differences in the mean values of important public health metrics in the different models relative to the uncertainty in outcomes for which policy makers should prepare. Another important measure of practical significance is the impact of model type on the determination of optimal policy. Two models may generate similar base-case behavior yet respond differently to policies; cost-benefit analysis may therefore lead to different policy choices in different models. For example, contact reduction, whether resulting from mandatory quarantine or voluntary social distancing, reduces the size and impact of the outbreak. However, the results show that network type and individual heterogeneity affects the benefits of contact reduction. Thus, for certain values of the costs and benefits, the decision to implement mandatory control measures such as quarantine will depend on the network structure of and contact heterogeneity within the population. Similarly, the costs and benefits of policies will depend on behavioral feedbacks such as the degree of endogenous social distancing. Sensitivity analysis over these and other sources of uncertainty is required to assess the robustness of policy choices to model assumptions.

Model complexity can be expanded in different directions. Modelers can add detail, disaggregating populations by location, individual attributes, and relationship networks. Alternatively they can expand the model boundary to include feedbacks with other subsystems. For example, the results reported here assume fixed network structure, contact rates and infectivities. All are actually endogenous. As prevalence increases, people change their behavior. Social distancing and safe practices disrupt contact networks, reduce contact frequencies and cut the probability of transmission. From staying home, increased hand washing, and use of masks (for SARS) to abstinence, condom use, and needle exchange (for HIV), endogenous behavior change lowers \( R_0 \) and can have large effects on disease diffusion (Blower, Gershengorn et al. 2000). Alternatively, behavior change may worsen an epidemic: people fleeing a disease make contact tracing more difficult and may seed outbreaks in remote areas; more effective treatments for HIV increase risky behaviors for some people (Lightfoot, Swendeman et al. 2005). The impact of such feedback effects may be larger than the impact of network structure and individual heterogeneity and should not be excluded in favor of greater detail. The policy test above illustrates: the reduction in \( R_0 \) with cumulative cases can be interpreted as endogenous social
distancing. This feedback has a large impact on model behavior compared to the differences between the DE and AB models. Expanding the boundary of a model can have effects much greater than those introduced by disaggregation from compartments to individuals.

In a review entitled “Uses and abuses of mathematics in biology,” May (2004, p. 793) calls for balance in model development:

Perhaps most common among abuses, and not always easy to recognize, are situations where mathematical models are constructed with an excruciating abundance of detail in some aspects, whilst other important facets of the problem are misty or a vital parameter is uncertain to within, at best, an order of magnitude. It makes no sense to convey a beguiling sense of “reality” with irrelevant detail, when other equally important factors can only be guessed at.”

While further work is needed, the results reported here may be useful to modelers seeking the appropriate balance among detail, scope, and the ability to carry out sensitivity analysis over the inevitable uncertainties we all face.

References


Table 1. Base case parameters. The supplement provides full details for the AB simulations.

<table>
<thead>
<tr>
<th>Parameter (dimensionless)</th>
<th>Parameter</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infectivity, Exposed $i_{ES}$</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Infectivity, Infectious $i_{IS}$</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Basic reproduction rate $R_0$</td>
<td>4.125</td>
<td></td>
</tr>
<tr>
<td>Average links per node $k$</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Prob of long-range links (SW) $p_{sw}$</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Scaling exponent (scale-free) $\gamma$</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>Total population $N$</td>
<td>200</td>
<td>Person</td>
</tr>
<tr>
<td>Contact rate, Exposed $c_{ES}$</td>
<td>4</td>
<td>1/Day</td>
</tr>
<tr>
<td>Contact rate, Infectious $c_{IS}$</td>
<td>1.25</td>
<td>1/Day</td>
</tr>
<tr>
<td>Average incubation time $1/\epsilon$</td>
<td>15</td>
<td>Day</td>
</tr>
<tr>
<td>Average duration of illness $1/\delta$</td>
<td>15</td>
<td>Day</td>
</tr>
</tbody>
</table>

Table 2. Median estimated value of $R_0$ for the calibrated DE models. $R^2$, the square of the Pearson correlation coefficient, measures goodness of fit between each AB and calibrated DE simulation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Connected</th>
<th>Random</th>
<th>Scale-free</th>
<th>Small World</th>
<th>Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
</tr>
<tr>
<td>Median</td>
<td>4.21</td>
<td>2.96</td>
<td>3.10</td>
<td>2.54</td>
<td>3.15</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.71</td>
<td>0.62</td>
<td>0.58</td>
<td>0.52</td>
<td>0.66</td>
</tr>
</tbody>
</table>

| Implied $R_0$ = $c_{ES}i_{ES}/\epsilon + c_{IS}i_{IS}/\delta$ | Median | $\sigma$ | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.998 | 0.998 | 0.985 | 0.987 |
| $R^2$ | Median | $\sigma$ | 0.025 | 0.049 | 0.017 | 0.050 | 0.016 | 0.039 | 0.040 | 0.059 | 0.056 | 0.043 |

Table 3. Mean and standard deviation of Final Size, F, in (1) the AB and (2) fitted DE simulations. (3) % of AB and (4) fitted DE simulations with F < 0.10. */** indicates F in the base DE (0.98) falls outside the range encompassing 95/99% of the AB simulations.

<table>
<thead>
<tr>
<th></th>
<th>Connected</th>
<th>Random</th>
<th>Scale-free</th>
<th>Small World</th>
<th>Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
<td>$H_0$</td>
</tr>
<tr>
<td>1</td>
<td>AB Mean ($\sigma$)</td>
<td>0.97</td>
<td>(0.13)</td>
<td>0.90*</td>
<td>(0.19)</td>
</tr>
<tr>
<td>2</td>
<td>Fitted DE $\mu$ ($\sigma$)</td>
<td>0.98</td>
<td>(0.07)</td>
<td>0.91</td>
<td>(0.17)</td>
</tr>
<tr>
<td>3</td>
<td>AB %F &lt; 0.10</td>
<td>1.8</td>
<td>4.4</td>
<td>2.7</td>
<td>3.8</td>
</tr>
<tr>
<td>4</td>
<td>Fitted DE %F &lt; 0.10</td>
<td>0.5</td>
<td>3.5</td>
<td>2.5</td>
<td>4.0</td>
</tr>
</tbody>
</table>
Table 4. Peak time and peak prevalence in 1000 simulations of the AB model and calibrated DE models for each experimental condition. */** indicates the base DE values (T\textsubscript{p} = 48 days and I\textsubscript{max} = 27.1%) fall outside the 95/99% confidence bounds defined by the ensemble of AB simulations.

<table>
<thead>
<tr>
<th></th>
<th>Connected</th>
<th>Random</th>
<th>Scale-free</th>
<th>Small World</th>
<th>Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peak Time, T\textsubscript{p} (Days)</strong></td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
</tr>
<tr>
<td>AB (\mu)</td>
<td>49.8</td>
<td>44.9</td>
<td>52.8</td>
<td>49.5</td>
<td>60.6</td>
</tr>
<tr>
<td>AB (\sigma)</td>
<td>10.8</td>
<td>12.4</td>
<td>13.4</td>
<td>14.2</td>
<td>16.7</td>
</tr>
<tr>
<td>Fitted DE (\mu)</td>
<td>51.3</td>
<td>49.6</td>
<td>56.6</td>
<td>58</td>
<td>62.9</td>
</tr>
<tr>
<td>Fitted DE (\sigma)</td>
<td>9</td>
<td>21.3</td>
<td>21.3</td>
<td>37.7</td>
<td>23.2</td>
</tr>
<tr>
<td><strong>Peak Prev I\textsubscript{max} (%)</strong></td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
</tr>
<tr>
<td>AB (\mu)</td>
<td>29.1</td>
<td>27.1</td>
<td>26.5</td>
<td>25.1</td>
<td>24.6</td>
</tr>
<tr>
<td>AB (\sigma)</td>
<td>4.9</td>
<td>6.3</td>
<td>5.2</td>
<td>5.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Fitted DE (\mu)</td>
<td>26.9</td>
<td>26.7</td>
<td>24.6</td>
<td>24.2</td>
<td>22.8</td>
</tr>
<tr>
<td>Fitted DE (\sigma)</td>
<td>2.8</td>
<td>13.2</td>
<td>9.6</td>
<td>18.1</td>
<td>6.1</td>
</tr>
</tbody>
</table>

Table 5. Public health metrics under contact reduction. */** indicates the DE simulation falls outside the 95/99% confidence bound defined by the ensemble of AB simulations. The results for the DE model are F = 0.190, T\textsubscript{p} = 31.3 days, and I\textsubscript{max} = 4.43%.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Final Size F</th>
<th>Peak Time T\textsubscript{p}</th>
<th>Peak Prev I\textsubscript{max}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
<td>H\textsubscript{s}</td>
</tr>
<tr>
<td>Final Size F (\mu)</td>
<td>0.215</td>
<td>0.249</td>
<td>0.157</td>
</tr>
<tr>
<td>Final Size F (\sigma)</td>
<td>0.084</td>
<td>0.091</td>
<td>0.064</td>
</tr>
<tr>
<td>Peak Time T\textsubscript{p} (\mu)</td>
<td>35.0</td>
<td>36.1</td>
<td>33.1</td>
</tr>
<tr>
<td>Peak Time T\textsubscript{p} (\sigma)</td>
<td>15.3</td>
<td>15.9</td>
<td>14.8</td>
</tr>
<tr>
<td>Peak Prev I\textsubscript{max} (\mu)</td>
<td>6.42</td>
<td>7.28</td>
<td>5.17</td>
</tr>
<tr>
<td>Peak Prev I\textsubscript{max} (\sigma)</td>
<td>2.40</td>
<td>2.66</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Figure 1. Left: DE model with base parameters (Table 1). Right: Typical simulation of the equivalent AB model with the heterogeneous condition of the scale free network.
**Figure 2.** Comparison of DE and AB models. Graphs show prevalence of symptomatic infectious individuals (I/N, %). Panels show the envelopes comprising 50%, 75% and 95% of 1000 AB simulations for each network and heterogeneity condition. The thin black line is the base case DE with parameters as in Table 1; the thick red line shows the mean of the AB simulations.