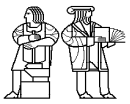

*Massachusetts Institute of Technology*

*Engineering Systems Division*

**Working Paper Series**

**ESD-WP-2004-05**

# HETEROGENEITY AND NETWORK STRUCTURE IN THE DYNAMICS OF DIFFUSION: COMPARING AGENT-BASED AND DIFFERENTIAL EQUATION MODELS

**Hazhir Rahmandad and John Sterman**
**Sloan School of Management**
**Massachusetts Institute of Technology**

**November, 2004**

# 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 [1]

*MIT Sloan School of Management, Cambridge MA 02142*

**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 in agent attributes and in the network of interactions among them. Using contagious disease as an example, we contrast the dynamics of AB models with those of the corresponding mean-field DE model, specifically, comparing the standard SEIR model—a nonlinear DE—to an explicit AB model of the same system. We examine both agent heterogeneity and the impact of different network structures, including fully connected, random, Watts-Strogatz small world, scale-free, and lattice networks. Surprisingly, in many conditions the AB and DE dynamics are quite similar. Differences between the DE and AB models are not statistically significant on key metrics relevant to public health, including diffusion speed, peak load on health services infrastructure and total disease burden. We explore the conditions under which the AB and DE dynamics differ, and consider implications for managing infectious disease. The results extend beyond epidemiology: from innovation adoption to the spread of rumor and riot 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

---

[1] We thank Rosanna Garcia, Nelson Repenning, seminar participants at MIT, and participants at the 2004 NAACSOS conference and 2004 International System Dynamics Conference for helpful comments. Financial support provided by the Project on Innovation in Markets and Organizations at the MIT Sloan School.

**Introduction:** Spurred by growing computational power, agent-based modeling (AB) is increasingly applied to problems previously modeled with nonlinear differential equations (DE). Scholars interested in physical, biological, social, and economic phenomena can now choose from highly disaggregate, AB representations to highly aggregated models. Both approaches have yielded important insights. In the social sciences, agent models explore phenomena from the emergence of segregation to organizational evolution and the emergence of hierarchy to market dynamics (Schelling 1978; Levinthal and March 1981; Carley 1992; Axelrod 1997; Axtell, Epstein et al. 2002; Epstein 2002; Tesfatsion 2002). Differential and difference equation models have an even longer history in the social sciences, from innovation diffusion (Bass 1969) and epidemiology (Andersson and Britton 2000) to option pricing (Black and Scholes 1973), business cycles (Samuelson 1939), and many others.

Each method has strengths and weaknesses. Nonlinear DE models often have a broad boundary encompassing a wide range of feedback effects but typically aggregate agents into a relatively small number of states (compartments). For example, models of innovation diffusion may aggregate the population into categories including unaware, aware, in the market, recent adopters, and former adopters (Urban, Hauser and Roberts, 1990; Mahajan, Muller and Wind, 2000). However, the agents within each compartment 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 agent attributes and in the network structure of their interactions; like DE models, these interactions can be deterministic or stochastic. However, the increased detail comes at the cost of introducing large numbers of parameters. It can be difficult to analyze the behavior of an AB model, and the computing resources required to carry out sensitivity tests can be prohibitive. Understanding where the agent-based approach yields additional insight and where such detail is unimportant is central to selecting appropriate methods for any problem at hand.

The stakes are large. Consider potential bioterror attacks. Kaplan, Craft, and Wein (2002) used a nonlinear DE model to examine smallpox attacks, 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 symptomatic individuals. They conclude MV outperforms TV in reducing casualties. In contrast, Eubank et al. (2004) and Halloran et al. (2002), using different AB models, conclude TV is superior. It is unclear whether the difference arises from relaxing the perfect mixing and homogeneity assumptions of the DE or from other assumptions such as the size of the population (ranging from 10 million for the DE model to 2000 for the Halloran et al. model), the implementation of vaccination strategy, parameters, etc. (Koopman 2002, Ferguson et al. 2003, Kaplan and Wein 2003).

We argue that AB and DE models are more productively viewed as points on a spectrum of aggregation assumptions rather than as fundamentally incompatible modeling paradigms. Further, in addressing a particular policy issue, whether bioterror attack or innovation diffusion, there may be elements in the system best modeled with explicit agents and some best modeled as aggregates. For example, in modeling the evolution of the personal computer industry, it may be helpful to model firms as individual agents while aggregating the millions of customers into a small number of different segments. However, despite calls for "model alignment" (e.g., Axtell, Axelrod, Epstein and Cohen 1996) few models integrate both AB and DE elements.

Here we compare AB and DE models of the diffusion of contagious disease. We choose the context of disease diffusion for three reasons. First, the effects of different network structures for contacts among individuals are important in the diffusion process in general (Davis 1991; Strang and Tuma 1993; Watts and Strogatz 1998; Barabasi 2002; Rogers 2003), providing a strong test for differences between the two approaches. Second, the dynamics of contagion involve many of the important characteristics of complex systems, including positive and negative feedbacks (e.g., contagion, recovery, immunization), time delays (e.g. latency in incubation and recovery), nonlinearities (the probability of infection depends nonlinearly on the size of the susceptible and infectious populations), stochastic events (random interactions between individuals), and agent heterogeneity (rates of contact, mortality and morbidity, etc.).

Finally, diffusion is a fundamental process manifesting in diverse physical, biological,

social, and economic settings. Many diffusion phenomena in human systems involve processes of social contagion analogous to infectious diseases, including word of mouth, imitation, and network externalities. From the diffusion of innovations, ideas, and new technologies within and across organizations, to rumors, financial panics and riots, contagion-like dynamics, and formal models of them, have a rich history in the social sciences (Bass 1969; Strang and Tuma 1993; Watts and Strogatz 1998; Mahajan 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 lumped nonlinear deterministic DE model (see e.g. Murray 2002). The DE version divides the population into four compartments: Susceptible (S), Exposed (E), Infected (I), and Recovered (R). In the AB model, each individual is separately represented and must be in one of the four states. To ensure comparability of the AB and DE models, we implement them in the same software environment and show how a stochastic AB model can be formulated in continuous time so that the same numerical integration procedure can be used in both. We set the (mean) values of parameters in the AB model equal to those of the DE. Therefore any differences in outcomes arise only from the relaxation of the mean-field aggregation assumptions of the DE model. We run the AB model under five different network structures, including fully connected, random, Watts-Strogatz small world, scale-free, and lattice. The fully connected network is closest to the perfect mixing assumption of the DE; the lattice, with connections solely to neighbors, is most different; the small world and scale free networks are widely used and characterize many real situations (Watts and Strogatz 1998; Barabasi and Albert 1999; Barabasi 2002). We test each network structure with homogeneous and heterogeneous agent attributes such as the rate at which each agent contacts others. We compare the DE and AB epidemics on a variety of key metrics relevant to public health, including the fraction of the population ultimately infected (the total burden of disease), the maximum prevalence of infectious cases (a measure of the peak load on public health infrastructure), and the time to the

peak of the epidemic (indicating how much time health officials have to respond).

As expected, different network structures alter the timing of the epidemic. The behavior of the AB model under the fully connected and random networks closely matches the DE model, as these networks approximate the perfect mixing assumption. Diffusion is fastest in the uniform network (where everyone can contact everyone else) and slowest in the lattice (where people contact near neighbors only). Also as expected, the stochastic AB models generate a distribution of outcomes, including some in which, due to random events, the epidemic never takes off despite an aggregate basic reproduction number greater than one. The deterministic DE model cannot generate this mode of behavior.

Surprisingly, however, the differences between the DE and AB models are not statistically significant for key metrics such as peak time, peak prevalence, and disease burden in any but the lattice network. Though the small-world and scale-free networks are highly clustered, their dynamics are close to the DE model: even a few long-range contacts and highly connected hubs seed the epidemic at multiple points in the network, enabling it to spread rapidly.

We also examine the ability of the DE model to capture the dynamics of each network structure in the realistic situation where data on underlying parameters are not available. Key parameters are often poorly constrained by biological and clinical data, particularly for emerging diseases. When a novel pathogen such as SARS emerges and clinical data on key parameters such as infectivity, incubation time and disease duration are not available, epidemiologists estimate potential diffusion by fitting models to the aggregate data (Dye and Gay 2003; Lipsitch et al. 2003; Riley et al. 2003). Calibration of innovation diffusion and new product marketing models is similar (Mahajan et al. 2000). We mimic such practice by treating the AB simulations as the "real world" and fitting the deterministic SEIR model to them. Surprisingly, the fitted DE model matches the mean behavior of the AB model under all network structures and heterogeneity conditions tested. Departures from homogeneity and perfect mixing are incorporated into the best-fit values of parameters such as the transmission rate.

The parsimony and robustness of the DE model suggests these models remain useful and

appropriate in many situations, particularly where network structure is unknown or labile and where fast turnaround is required. The detail and flexibility of the AB models are likely to be most helpful where the structure of the contact network is known, stable, and highly localized, and where it is important to understand the impact of stochastic events on the range of likely outcomes. Further, since time and resources are always limited, modelers must trade off the data requirements and computational burden of disaggregation against the breadth of the model boundary. AB models will be most appropriate where results depend delicately on agent heterogeneity and random events. DE models will be most appropriate where results hinge on the incorporation of a wide range of feedbacks with other system elements (a broad model boundary). We suggest the complementary strengths and weaknesses of each model type can be used to advantage when DE and AB elements are integrated in a single model.

The next section discusses the differences between AB and DE models and reviews the literature. We then describe the structure of the models, the design of the simulation experiments, and the results. Finally we discuss 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 rather than incompatible modeling paradigms. Aggregation is one of the key dimensions of that space. Models can range from lumped deterministic differential equations to stochastic compartmental models, in which the continuous real-valued variables of the DE are replaced by discrete individuals, to event history models where the states of individuals are tracked but their network of relationships is ignored, to dynamic network models where networks of individual interactions are explicit (e.g., Koopman et al. 2001).

The degree of aggregation should be determined by the purpose of the model. Consider a model of world population. At the most aggregate level, both AB and DE models can treat population as a single compartment, lumped in the case of the DE, and representing (6.25 billion!) individuals in the case of the AB model. When disaggregation is necessary the population can be segmented by age, sex, region, and so on, by introducing more compartments

(in the DE model) and more agent types (in the AB model), continuing, if needed, until each person is represented as a unique individual with unique attributes. Indeed, why stop at the level of the individual? Individuals could be modeled as 'agents' in the form of organs, the organs as 'agents' of cells, the cells as consisting of organelles, and so on. Agent models, like all models, aggregate all structure and heterogeneity below some threshold. The level of aggregation must always be matched to the model purpose and is constrained by computer power, data availability, and the time available for the study.

Another common difference is the representation of time. In DE models time is continuous. DE models are typically nonlinear and of such complexity that no analytical solutions can be found and the equations are solved by numerical integration. The time step should be small enough so that results correspond, within the required tolerance, to the solution of the underlying differential equation. AB models are typically formulated in discrete time, with agents interacting at intervals, as for example in cellular automata models of the iterated prisoners' dilemma where agents on a lattice cooperate or defect based on the past moves of their neighbors (the literature is massive; see e.g. Wolfram 2002, Gotts, Polhill and Law 2003, Nowak and Sigmund 2004). However, the time interval between rounds is often undefined. Where the period between actions is defined, changing it requires manual reparameterization to ensure that the hazard rates for each state transition remain invariant to the length of the time step.

A third dimension is the breadth of the model boundary. Model boundary here stands for the richness of the feedback structure captured endogenously in the model. A highly disaggregated model may represent population at the level of 1-year age cohorts, sex, and zip code, but assume constant 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 depending on food per capita, access to health care, exposure to pollution, social norms for family size, and so on, each of which, in turn, may be endogenous; such a model has a broad boundary. DE and AB models may in principle fall anywhere on this dimension. In practice, where time, budget, and computation resources are limited, modelers

must trade off breadth of boundary and the level of disaggregate detail.

Research at the intersection of the two methodologies is limited. Axtell et al. (1996) call for "model alignment" or "docking" and illustrate with the Sugarscape model. Parunak, Savit and Riolo (1998) compare AB and DE models of a supply chain. 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. Picard and Franc (2001) compare aggregate and agent forest growth models, concluding that the AB model fits the data better; however, the parameters of the aggregate models were taken from the AB model, not estimated. In epidemiology, Keeling (1999) formulates a DE model that approximates the effects of spatial structure; the dynamics capture the full spatial model well even when contact networks are highly clustered. Chen et al. (2003) and Chen et al. (2004) develop AB models of smallpox and anthrax, finding the dynamics generally consistent with DE models. In contrast, Kaplan et al.'s (2002) DE smallpox model conflicts with the AB models of Halloran et al. (2002) and Eubank et al. (2004). In sum, AB and DE models of the same phenomenon sometimes agree and sometimes diverge. None of the studies systematically vary network structure and agent heterogeneity to determine the source of the divergence.

**Model Structure:** The SEIR model is a lumped nonlinear differential equation model in which all members of the population are in one of four stages of the disease—Susceptible, Exposed, Infected, and Recovered. Exposed (asymptomatic) and infected (symptomatic) individuals can transmit the disease to susceptibles before they recover or die. The exposed compartment captures latency between infection and the emergence of symptoms. Typically, the exposed have more contacts than the infectious because they are healthier and often unaware that they are potentially contagious, while the infectious are ill and self-quarantined.

SEIR models have been applied successfully to many diseases. Additional states are often introduced to capture more complex disease lifecycles, diagnostic categories and therapeutic protocols, heterogeneity in the population, birth/recruitment of new susceptibles, loss

of immunity, etc. In this study we maintain the boundary assumptions of the traditional SEIR model (four stages, fixed population, permanent immunity). The DE implementation of the model makes several additional assumptions, including perfect mixing and homogeneity of individuals within each compartment, Poisson exit probabilities for exposed and infectious individuals (the probability of exit is independent of the time spent in each state), and mean field aggregation (the flows between compartments equal the expected value of the sum of the underlying probabilistic rates for individual agents). To derive the differential equations, consider the number of new cases each infectious individual generates:

$$c_{IS}\text{*Prob}\textit{(Contact with Susceptible)}\text{*Prob}\textit{(Transmission|Contact with Susceptible),} \qquad (1)$$

where the contact frequency $c_{IS}$ is the number of contacts between infectious and susceptible individuals per time period (e.g., per day); homogeneity implies $c_{IS}$ is equal for all infectious individuals. 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 a contact with a susceptible, or infectivity, as $i_{IS}$, and summing over the infectious population yields the total flow of new cases generated by contacts between I and S populations, $c_{IS}*i_{IS}*I*(S/N)$. The number of new cases generated by contacts between the exposed and susceptibles is formulated analogously, yielding the total Infection Rate, IR,

$$IR = (c_{ES} * i_{ES} * E + c_{IS} * i_{IS} * I) * (S/N). \qquad (2)$$

Turning to the transitions from E to I and I to R (the Emergence and Recovery Rates), perfect mixing implies residence times in the E and I stages are distributed exponentially. Denoting the mean emergence (incubation) time and disease duration $\varepsilon$ and $\delta$, respectively, yields

$$ER = E/\varepsilon \text{ and } RR = I/\delta. \qquad (3)$$

The full model is thus:

$$\frac{dS}{dt} = -IR, \qquad \frac{dE}{dt} = IR - ER, \qquad \frac{dI}{dt} = ER - RR, \qquad \frac{dR}{dt} = RR. \qquad (4)$$

Anderson and May (1991) and Murray (2002) provide extensive discussion. In general, DE models like the SEIR model are nonlinear and must be integrated numerically. We use Euler

integration to facilitate mixing AB and DE elements in the same model.[2] Under Euler integration the infectious population evolves as:

$$I_{t+dt} = I_t + dt*(ER_t - RR_t) \tag{5}$$

where *dt* is the time step for the numerical integration. The number of people recovering each time step is then $RR_t = dt*(I_t/\delta)$. The other states are updated analogously.

The AB model relaxes the perfect mixing and homogeneity assumptions. Each individual is in one of the four states S[j], E[j], I[j], and R[j] for $j \in (1, \ldots, N)$. An individual's current state has value 1 while all others are 0. The aggregate rates IR, ER, and RR are the sum of the individual transitions IR[j], ER[j], and RR[j]. These transitions are one when each agent moves from the current to the next stage of the disease and zero otherwise. To integrate AB and DE elements in the same model it is necessary to formulate AB elements in continuous time so that their dynamics are invariant to changes in the time step for the numerical integration. To illustrate, consider the recovery of the *j*th individual. The Infectious state, I, evolves as:

$$I[j]_{t+dt} = I[j]_t + dt*(ER[j]_t - RR[j]_t) \tag{6}$$

$$RR[j]_t = \begin{cases} I[j]_t/dt & if \quad r_I[j]_t < dt/\delta[j] \\ 0 & otherwise \end{cases} \tag{7}$$

where $r_I[j]$ is a random number drawn from U(0, 1) and $\delta[j]$ is the expected duration of disease for the jth individual. If $r_I[j] < dt/\delta[j]$ and the individual is infectious (I[j] = 1), the individual recovers: R[j] becomes 1, while I[j] becomes zero (if the individual is infectious ER[j] must be zero). The instantaneous recovery probability each time step is $dt/\delta[j]$. Assuming, as in the DE model, that the expected duration of disease is equal for all individuals ($\delta[j] = \delta$ for all j) and summing over the population yields the expected value of the recovery rate in each time step: $dt*(I_t/\delta)$, identical to the mean-field DE model. The other state transitions in the AB model are

[2] We set dt = 0.5 days. Results are not sensitive to dt = 0.25 and 0.125, indicating 0.5 days is sufficiently short to limit integration error.

formulated analogously. AB elements formulated in this fashion can be integrated with lumped elements in a single model, and the modeler is free to vary the time step for the numerical integration to achieve the accuracy required.

We use parameters roughly characterizing smallpox (Halloran et al. 2002, Kaplan et al. 2002, Ferguson et al. 2003, Eubank et al. 2004).[3] The key parameter is the basic reproduction number, $R_0$, the expected number of new cases each contagious individual generates before recovering, assuming all others are susceptible. Smallpox has a mid range reproduction number, $R_0 \approx 3 - 10$, and therefore is a good choice to observe potential differences between DE and AB models: diseases with $R_0 << 1$ die out quickly for (nearly) any network structure, while for those with $R_0 >> 1$, e.g. chickenpox, measles, a severe epidemic is unavoidable in (nearly) any contact network. We set the basic reproduction rate $R_0 = 4.125$ for the DE (Table 1). The AB models use the same values for the infectivities and expected residence times, and we set the contact frequencies so that the mean contact rate in each network structure and heterogeneity condition is the same as the DE model. We set the population, N = 200, all susceptible except for two randomly chosen exposed individuals. While small compared to settings of interest in policy design, e.g., cities or nations, the effects of random events and network structure are likely to be more pronounced in small populations (Kaplan and Wein 2003), 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, independent of population size. The AB model has 4*N states, and must also track interactions among the N individuals. Depending on the density of the contact network, the AB model can grow at rates up to $O(N^2)$. With a population of only 200 the AB model contains 5 to 50 thousand variables, depending on the network structure, plus the associated parameters, initial conditions, and random number calls. The online material includes the model and full documentation.

---

[3] Smallpox-specific models often add a prodromal period between the incubation and infectious stages (e.g. Eubank et al. 2004); for generality we maintain the SEIR structure.

**Experimental design:** We vary both the network structure of contacts among individuals and the degree of agent heterogeneity in the AB model and compare the resulting dynamics to the DE version. We implement a full factorial design with five network structures and two heterogeneity conditions, yielding ten experimental conditions. In each condition we generate an ensemble of 1000 simulations of the AB model, each with different realizations of the random variables determining contacts, emergence, and recovery. Each simulation uses a different realization of the stochastic network structures (random, small world, and scale-free; the uniform and lattice networks are deterministic and equal in all cases). 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 structure, agent heterogeneity, or the discrete, stochastic treatment of individuals and state transitions in the AB model.

*Network structure:* The DE model assumes perfect mixing, implying anyone can meet anyone else with equal probability. Realistic networks are far more sparse, with most people contacting only a small fraction of the population. Further, most contacts are clustered among near neighbors.[4] We explore five different network structures: uniform (each person can meet any other with equal probability), random (Erdos and Renyi 1960), small-world (Watts and Strogatz 1998), scale-free (Barabasi and Albert 1999), and lattice (where contact only occurs between near neighbors). We parameterize the model so that all networks have the same mean number of links per node, k = 10 (Watts 1999).

The fully connected network corresponds most closely to the perfect mixing assumption of the DE model; the only difference is the discrete, stochastic character of state transitions. In the random network people are linked with a subset of the population; each node has the same probability of being linked to every other node. At the other extreme, the lattice, where contacts

---

[4] Clustering can be measured by the mean clustering coefficient $(1/N)\sum_i 2n_i/[k_i(k_i-1)]$ where $n_i$ is the total number of links among all nodes linked to $i$ and $k_i$ is the number of links of the $i$th node (Watts and Strogatz 1998).

occur only between near neighbors, is expected to be most different from the DE model. The small world and scale-free networks are intermediate cases with many local and some long-distance links. These networks are widely used and characterize a number of real systems (Watts 1999, Barabasi 2002). In a small-world network most links are local, while a few are long-range. We set the probability of long-range links to 0.05, in the range used by Watts (1999). We build the scale-free networks using the preferential attachment algorithm of Barabasi and Albert (1999) in which the probability a new node links to existing nodes is increasing in the number of links the existing node already has, creating a positive feedback through which popular nodes become even more popular. 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 are characterized by a number of "hubs" with a large number of links, while the rest of the population is mainly connected to these hubs. Initial diffusion in scale-free networks is expected to be faster than random networks, as the hubs get infected quickly and then spread the disease to remote parts of the network (Barthelemy et al. 2004). To test the network most different from the perfect mixing assumption, we also model a one-dimensional ring lattice with no long-range contacts. With k = 10 links per node, each person contacts only the five neighbors on each side. Diffusion is expected to be much slower than the other networks. The online material details the construction of each network.

*Agent Heterogeneity:* Each individual has four relevant characteristics: expected contact rate, infectivity, emergence time, and disease duration. In the homogeneous condition each agent is identical with parameters set to the values in the DE model. In the heterogeneous condition we vary individual contact frequencies. Focusing on contact frequencies reduces the dimensionality of required sensitivity analysis at little cost. First, only the product of the contact rate and infectivity matters to the spread of the disease. Second, the exponential distributions of emergence time and disease duration introduce significant variation in individual residence times in each disease stage even when mean incubation times and disease durations are equal.

Heterogeneity in contacts between individuals is modeled as follows. Given that two

people are linked (that they can come into contact), the frequency of contacts between them depends on two factors. First, contact frequency depends on how often each uses their links, on average: some people are gregarious; others are shy. Second, time constraints may limit contacts. At one extreme, the probability of using a link 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 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 meet people is fixed, the chance of using a link is inversely proportional to the number of links, a reasonable assumption when transmission requires personal contact: the professor can only converse with a limited number of people each day. We capture these effects by assigning individuals different propensities to use their links, $\lambda[j]$, formulating the mean contact frequency for the link between individuals i and j, c[i,j], as

$$c[i,j]=\kappa*\lambda[i]*\lambda[j]/ (k[i]*k[j])^{\tau} \quad (8)$$

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 all individuals 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, $\lambda[j] \sim U(0.25, 1.75)$. Table 2 shows the standard deviation of individual contact rates c[i,j] for each network. In the homogeneous cases, deterministic link assignment means there is no variance in c[i,j] for the uniform and lattice networks. There is always some heterogeneity in the random, small world, and scale free networks because the number of links per node is a random variable. In the heterogeneous condition c[i,j] varies for all network structures, with the highest variance in the scale free case—hubs have many more contacts than hermits. The online documentation includes the histograms of each distribution.

*Calibrating the DE Model:* In real world applications the parameters determining the basic reproduction rate $R_0$ are often poorly constrained by biological or clinical data. In the case of new and emerging diseases such as SARS, vCJD, BSE, avian flu, etc., clinical data on incubation times, infectivities, contacts, etc. do not become available until the epidemic has already spread widely. In such cases model parameters are usually estimated by fitting SEIR-style models to aggregate data as an outbreak unfolds; SARS provides a typical example (Dye and Gay 2003; Lipsitch et al. 2003; Riley et al. 2003). To mimic this protocol we treat the AB model as the "real world". For each network type, we then fit the DE model to the mean of the ensemble of AB simulations, estimating the infectivities ($i_{ES}$ and $i_{IS}$), incubation time, ($\varepsilon$) and duration of illness ($\delta$) by non-linear least squares (see online documentation).

**Results:** For each of the ten experimental conditions we run the AB model 1000 times (each simulation of the random, small world, and scale-free networks uses a different realization of the network structure; the uniform and lattice networks are deterministic). We compare the overall patterns of diffusion and quantitatively assess the dynamics on three measures relevant to public health. 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 health workers, treatment and vaccines (if they exist). The maximum of the infectious population (the *peak value,* $I_{max}$) indicates the peak load on public health infrastructure including immunization resources, hospital and quarantine facilities, and health workers. Finally, the fraction of the population ultimately infected (the *diffusion fraction,* F) measures the total burden of morbidity and mortality. Figure 1 compares the base case DE model with a typical simulation of the AB model in the heterogeneous condition of the scale free network. Both epidemics follow the classic pattern in which prevalence grows rapidly through the positive feedback of contagion, then falls as the supply of susceptibles is depleted. The epidemic in the sample scale-free simulation grows slightly faster than in the DE ($T_p$ = 38 vs. 49 days), has similar peak prevalence ($I_{max}$ = 26% vs. 23%), and ultimately afflicts fewer people (F = 81% vs. 98%).

Figure 2 shows, for each network structure and heterogeneity condition, the symptomatic (infectious) population for an ensemble of 1000 simulations. Also shown are the mean of the ensemble and the trajectory of the DE model with both base case and best-fit parameters. Tables 4 and 5 report F, $T_p$, and $I_{max}$ for each condition, and compare them to the base and fitted DE models. Several important features are apparent. First, except for the lattice, the general mode of behavior is similar to that of the base case DE model and changes little across different network or heterogeneity conditions. In nearly all cases the epidemic generates the classic diffusion pattern where the epidemic initially spreads at an increasing rate, then gradually declines. The similarity is no surprise: The initial growth is driven by the positive contagion feedbacks where the exposed and infectious populations spread the infection thus further increasing the prevalence of infectious individuals. The epidemic ends when the susceptible population is sufficiently depleted that the (mean) number of new cases generated by the exposed and infected populations is less than the rate at which infected individuals recover.

There is some numerical sensitivity to network structure (Figure 2; tables 4, 5). Departures from the behavior of the base case DE model increase from the uniform to the random, scale free, small world, and lattice structures. The degree of clustering explains much of the variation: The fully connected and random networks have low clustering—the chance of contacting distant people is the same as that for neighbors. The positive contagion feedback is strongest in the uniform network because an infectious individual can contact everyone else, minimizing local contact overlap. The lattice has maximal clustering (all contacts are with neighbors). When most contacts are local, infectious people repeatedly contact the same neighbors. After these neighbors are infected, the chance of contacting a susceptible and generating a new case declines, even if the total susceptible population remains high. When clustering is high, the epidemic can burn out in a local patch of the network, slowing diffusion.

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 even when $R_0 > 1$ as the few initially contagious individuals recover before passing the

disease to enough new cases. As a measure of early burnout, table 4 reports the fraction of cases where the diffusion fraction remains below 10%.[5] The network structure strongly conditions the probability of takeoff, with the fraction of cases experiencing takeoff highest for the uniform and random networks, and lower in networks where most interactions among agents are local.

As expected, the DE model falls very close to the mean of the AB simulations for the uniform network. Note that diffusion in the base DE model is slightly faster than the mean of the AB simulations, due to the discrete, stochastic nature of contacts and state transitions in the AB case. However, the trajectory of the base DE model almost always remains within the 75% interval and is statistically indistinguishable from the mean AB path. Heterogeneity has little impact in the uniform case; the differences in diffusion fraction, peak time and peak prevalence are small and insignificant. Low clustering means few cases experience early burnout.

The random network closely follows the uniform network because clustering is low. Diffusion is somewhat slower than in the DE and uniform network, however. Since each person is linked to a subset of individuals, the chance that the next contact will be a susceptible person falls faster than in the fully connected case. The incidence of early burnout, while higher than the fully connected case, remains low. Heterogeneity again has little impact.

Diffusion in the scale free network is also quite similar to the DE, uniform and random cases. Once the infection reaches a hub it quickly spreads throughout the population. In the homogeneous case, diffusion on average is initially slower than the DE model, followed by a burst as the hubs are infected. Interestingly, heterogeneity in the scale-free case speeds diffusion because the hubs are not only well-connected but have higher contact rates (eq. 8). However, since the mean contact rate is the same in all conditions, higher rates for hubs imply lower rates for most other nodes. Consequently, after the hubs recover the remaining infectious nodes have lower average contact rates, causing the epidemic to burn out at lower levels of diffusion (F =

---

[5] The results are not sensitive to the 10% cutoff. Because $R_0 > 1$, early burnout is unlikely once more than a few percent become contagious (except in the lattice), causing F to rise beyond 10%.

80% vs. 90% in the homogeneous case; Table 4, Figure 2). For the same reason the incidence of early burnout, while still small, rises compared to the fully connected and random cases. The base DE trajectory remains within the 75% envelope most of the time and almost never falls outside the 95% interval; differences between the base DE and mean of the AB trajectories are not significant for mean diffusion fraction, peak time, or peak prevalence.

Small world networks show slower diffusion compared to the DE, uniform, random, and scale-free cases because their clustering coefficient is high (most links are local) and they lack highly connected hubs. Nevertheless, the few long-range links are sufficient to seed the epidemic throughout the population, consistent with Watts and Strogatz (1998). The main impact of agent heterogeneity is greater dispersion and a slight, but insignificant, reduction in $T_p$.

The lattice is most different from the base DE model, with far slower diffusion. In the one-dimensional ring lattice used here, individuals can only contact their neighbors (with 10 links per node, five on either side). Hence the epidemic advances nearly linearly in a well-defined wave-front of new cases trailed by symptomatic and then recovered individuals. Such waves are observed in the spread of wildfire and plant pathogens where transmission is mostly local, though in two dimensions more complex patterns are common (Bjornstad et al. 2002, Murray 2002). The fraction of cases exhibiting early burnout is highest for the lattice. Because the epidemic wave front reaches a (stochastic) steady state in which recovery balances new cases, the probability of burnout is roughly constant over time. In the other networks early burnout quickly becomes rare as long-range links seed the epidemic in multiple locations.

There are few significant differences between the mean of the AB trajectories and the deterministic DE model on the public health metrics. Except for the lattice, the mean diffusion fraction, measuring the total burden of disease, varies little with network structure or agent heterogeneity; though F falls as clustering increases because of the increasing incidence of early burnout, the differences between the DE model and mean of the AB models are not significant except in the lattice. Similarly, the peak time, determining how long public health officials have to respond to an epidemic, and peak prevalence, determining the maximum load on health

infrastructure, vary little among the uniform, random, and scale free networks, and none of the differences between these networks and the base DE model are significant. As clustering increases, peak times rise and peak prevalence falls; $T_p$ is substantially higher, and $I_{max}$ lower, for the small world and lattice. Clustering also increases the variability of outcomes. Intuitively, the placement of patient zero makes no difference in the uniform network, but matters more in highly clustered networks (if patient zero is well connected, diffusion is more likely and more rapid than if patient zero arrives in a poorly connected hinterland). The increase in variability with clustering means that the ability to distinguish network structure from a given trajectory does not increase as fast as the differences in means would suggest. Despite the large increase in mean peak times for the small world and lattice compared to the DE model, the increased variability of outcomes means these differences are not significant. Peak prevalence is significantly lower than the DE model, however, for the small world and lattice networks.

The differences between the AB and best-fit DE models are far smaller. Indeed, the best fitting DE model and mean of the AB simulations are hardly distinguishable ($R^2 \geq 0.98$ in all cases). None of the differences in the key metrics F, $T_p$ and $I_{max}$ are statistically significant at the 95% level; only one is different at the 90% level. As the network becomes increasingly clustered and diffusion slows, the estimated parameters adjust accordingly, yielding a trajectory nearly indistinguishable from the mean AB trajectory. The estimated parameters, however, do diverge from those in the underlying model. Specifically, the estimated basic reproduction rate falls as clustering increases and diffusion slows. However, the variability in AB outcomes and uncertainty in clinical data constraining $R_0$ are typically so great that it will often be difficult to determine how network structure distorts parameter values compared to the DE case.

**Discussion:** Before discussing the implications, we consider limitations and extensions. While we examined a wide range of network structures and heterogeneity in agent attributes, the agent models contain many parameters that could be subject to 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). The robustness of results to common elaborations of the classic SEIR model such as loss of immunity, other distributions for incubation and recovery, recruitment of new susceptibles, non-human disease reservoirs and vectors, and so on could also be assessed. Other dimensions of agent heterogeneity and other network structures could be examined, including networks derived from field study of organizations and communities (Coleman, Katz and Menzel 1957, Ahuja and Carley 1999). Nevertheless, varying network type and agent heterogeneity while holding all other parameters constant reveals much about the impact of network structure and agent attributes on the dynamics of diffusion.

The impact of relaxing the perfect mixing, homogeneity, and mean-field assumptions of the differential equation model depends on the network structure of interactions in the population. For both the pattern of behavior and the magnitudes of key metrics relevant to public health, the differences between the DE and AB models are small and insignificant for the uniform, random, and scale-free network, larger but mostly statistically insignificant for the small world networks, and large and often significant for the ring lattice. While a pure lattice is unrealistic in modeling human diseases due to the high mobility of modern society, it may be appropriate in other contexts such as the spread of diseases among immobile plant populations. The discrete, stochastic representation of individuals in the AB models offers additional insight: the epidemic fizzles out in a small fraction of cases even though the aggregate basic reproduction number exceeds one. The deterministic DE model cannot generate such behavior. The more highly clustered the network, the greater the incidence of early burnout.

The results also show little sensitivity to agent heterogeneity. Compared to the homogeneous case, heterogeneity in transmission rates tends to cause slightly earlier peak-times as high-contact individuals rapidly seed the epidemic, followed by slightly lower diffusion levels as the high-contact individuals are removed, leaving those with lower average transmission probability and a smaller reproduction rate. Such dynamics were observed in the HIV epidemic, where initial diffusion was rapid in subpopulations with high contact rates. Of course, heterogeneity must ultimately matter: in the limit where a population bifurcates into a highly

connected group and a group of near-hermits, diffusion will proceed at different rates in each subculture. Such heterogeneity is often captured in DE models by adding additional compartments to distinguish the highly connected and hermit types.

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 resemble the individual level parameters governing the micro-level interactions among agents. Aggregate parameter estimates not only capture the mean of individual attributes such as the rate of infectious contacts but also the impact of network structure and agent heterogeneity. Modelers often use both micro-level and aggregate data to parameterize both AB and DE models; the results suggest caution must be exercised in doing so, and in comparing parameter values in different models (Fahse et al. 1998).

Other considerations condition the choice of modeling method. Data availability is a major concern. Since parameters affecting $R_0$ are often highly uncertain, extensive sensitivity tests are required. Further, without data on the contact network and distribution of agent attributes, modelers must also conduct sensitivity tests over alternative networks and attribute distributions to ensure robust results. There is a tradeoff between model complexity and an analyst's ability to build confidence in and understand the model. AB models require significantly more computation than DE models, often precluding required sensitivity tests and limiting the size of the population that can be considered. Complete mapping of parameter space takes a few seconds of simulation time in the DE model; doing so in an AB model of, e.g., a medium size city is infeasible. On the other hand, if the network structure is known, one can examine the effect of creating and removing nodes and links to simulate random failures or targeted attacks. Such questions cannot be answered with aggregate models.

The purpose of the model determines the risks of different assumptions. The costs of avoidable morbidity and mortality due to inadequate public health response are high. Some researchers therefore suggest it is prudent to use the well-mixed DE model, where diffusion is fastest, as a worst case limit (Kaplan and Lee 1990; Kaplan 1991). The same argument may

suggest greater attention to network structure in studies of innovation diffusion where policymakers seek to promote, rather than suppress, diffusion of new products.

Model complexity can be expanded in different directions. Modelers can add detail, disaggregating populations by attributes such as location, decision rules, and relationship networks. Alternatively they can expand the model boundary to include feedbacks with other subsystems. Because resources are limited analysts must trade off these attributes. Kaplan et al. (2002) modeled limits on vaccination infrastructure, an assumption with substantial impact on the results, and carried out extensive sensitivity tests, but accepted the perfect mixing, homogeneity, and mean field assumptions of the DE paradigm. Halloran et al.'s (2002) AB model relaxed these assumptions while omitting capacity constraints on vaccination and giving up the ability to model a realistic population and conduct full sensitivity analysis.

Most important, the results reported here assume fixed parameters including network structure, contact rates and infectivities. All are actually endogenous. As prevalence (and therefore the perceived risk of infection) increases, people change their behavior. Self quarantine and the use of safe practices disrupts contact networks, cuts contact frequencies, and reduces the probability of transmission given a contact. From staying home, increased hand washing, and use of masks (as in SARS) to abstinence, condom use, and needle cleaning (as for HIV), endogenous behavior change can have first-order effects on the dynamics (Blower et al. 2000). In some cases behavior change tends to be stabilizing as, e.g., self-quarantine and use of safe practices reduce $R_0$. In others behavior change may amplify the epidemic, for example as people fleeing a bioterror attack seed outbreaks in remote areas and make contact tracing more difficult. Such feedback effects may swamp the impact of network structure, heterogeneity, and random events and should not be excluded in favor of greater detail.

**Conclusion:** We compared agent-based and differential equation models in the context of contagious disease. From SARS to HIV to the possibility of bioterrorism, understanding the spread and control of infectious diseases is critical to public health and (inter)national security.

Further, analogous processes of social contagion (imitation, word of mouth, etc.) play important roles in many social and economic phenomena, from innovation diffusion to crowd behavior.

We showed how AB structures can be formulated in continuous time so that AB and DE elements can be integrated in a single model. We then compared the classic SEIR model to equivalent agent-based models with identical (mean) parameters. We tested network structures spanning a wide range of clustering, including uniform, random, scale-free, small world, and pure lattice. Each was tested with homogeneous and heterogeneous agents. Surprisingly, the differential equation model often captures the dynamics well, even in networks that diverge significantly from perfect mixing and homogeneity. The presence of even a few long-range links or highly connected hubs can seed an epidemic throughout the population, leading to diffusion patterns that, in nearly all cases, are not significantly different from the dynamics of the differential equation model. The results suggest extensive disaggregation may not be warranted unless detailed data characterizing network structure are available, that structure is stable, and the computational burden does not prevent sensitivity analysis or inclusion of other key feedbacks that may condition the dynamics.

Each method has strengths and weaknesses. Both types of model can incorporate random events, nonlinearities, delays, accumulations, feedback processes, behavioral decision processes, learning, and other structures common in complex adaptive systems. In principle, AB and DE models can fall most anywhere in the space defined by the degree of aggregation, the breadth of the model boundary, and the inclusion of random events. However, in practice the two types often differ. AB models can readily include heterogeneity in the attributes of the agents and in the network structure of their interactions, but the computation required for sensitivity analysis can be prohibitive. In contrast, DE models often have a broad boundary but typically disaggregate a population into a comparatively small number of states, unrealistically assuming perfect mixing and homogeneity within compartments. No matter how powerful computers become, limited time, budget, cognitive capabilities and client attention mean modelers must always choose whether to disaggregate to capture the attributes of individual agents, expand the

model boundary to capture additional feedback processes, or keep the model simple so that it can be analyzed thoroughly. Mixing DE and AB elements in a single model can enhance the ability of analysts to model important policy issues efficiently and effectively.

**References:**


Ahuja, M. K. and K. M. Carley. 1999. Network structure in virtual organizations. *Organization Science* **10**(6): 741-757.

Anderson, R. M. and R. M. May. 1991. *Infectious diseases of humans: Dynamics and control*. Oxford; New York, Oxford University Press.

Andersson, H. and T. Britton. 2000. *Stochastic epidemic models and their statistical analysis*. New York, Springer.

Axelrod, R. 1997. The dissemination of culture - a model with local convergence and global polarization. *Journal of Conflict Resolution* **41**(2): 203-226.

Axtell, R. L., J. Axelrod, J. M. Epstein and M. Cohen. 1996. Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory* **1**(2): 123-141.

Axtell, R. L., J. M. Epstein, J. S. Dean, G. J. Gumerman, et al. 2002. Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House valley. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 7275-7279.

Barabasi, A. L. 2002. *Linked: How everything is connected to everything else and what it means*. Cambridge, Massachusetts, Perseus.

Barabasi, A. L. and R. Albert. 1999. Emergence of scaling in random networks. *Science* **286**(5439): 509-512.

Barthelemy, M., A. Barrat, R. Pastor-Satorras and A. Vespignani. 2004. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters* **92**(17): 178701-1.

Bass, F. 1969. A new product growth model for consumer durables. *Management Science* **15**: 215-227.

Bjornstad, O. N., M. Peltonen, A. M. Liebhold and W. Baltensweiler. 2002. Waves of larch budmoth outbreaks in the european alps. *Science* **298**(5595): 1020-1023.

Black, F. and M. Scholes. 1973. Pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3): 637-654.

Blower, S. M., H. B. Gershengorn and R. M. Grant. 2000. A tale of two futures: HIV and antiretroviral therapy in San Francisco. *Science* **287**(5453): 650-654.

Carley, K. 1992. Organizational learning and personnel turnover. *Organization Science* **3**(1): 20-46.

Chen, L.-C., K. Carley, D. Fridsma, B. Kaminsky and A. Yahja. 2003. Model alignment of anthrax attack simulations. Working paper, Institute for Software Research, Carnegie Mellon University**.**

Chen, L.-C., B. Kaminsky, T. Tummino, K. M. Carley, E. Casman, D. Fridsma and A. Yahja. 2004. *Aligning simulation models of smallpox outbreaks*. Proceedings of the Second Symposium on Intelligence and Security Informatics, Tucson, AZ.

Coleman, J. S., E. Katz and H. Menzel. 1957. The diffusion of an innovation among physicians. *Sociometry* **20**: 253-270.

Davis, G. 1991. Agents without principles? The spread of the poison pill through the intercorporate network. *Administrative Science Quarterly* **36**: 583-613.

Dye, C. and N. Gay. 2003. Modeling the SARS epidemic. *Science* **300**(5627): 1884-1885.

Edwards, M., S. Huet, F. Goreaud and G. Deffuant. 2003. Comparing an individual-based model of behavior diffusion with its mean field aggregate approximation. *Journal of Artificial Societies and Social Simulation* **6**(4), http://jasss.soc.surrey.ac.uk/6/4/9.html.

Epstein, J. M. 2002. Modeling civil violence: An agent-based computational approach. *Proc. of the National Academy of Sciences of the United States of America* **99**: 7243-7250.

Erdos, P. and A. Renyi. 1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**: 17-61.

Eubank, S., H. Guclu, V. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature*. 429: 180-184.

Fahse, L., C. Wissel and V. Grimm. 1998. Reconciling classical and individual-based approaches in theoretical population ecology: A protocol for extracting population parameters from individual-based models. *The American Naturalist* **152**(6): 838-852.

Ferguson, N. M., M. J. Keeling, J. W. Edmunds, R. Gani, B. T. Grenfell, R. M. Anderson and S. Leach. 2003. Planning for smallpox outbreaks. *Nature* **425**: 681-685.

Gotts, N. M., J. G. Polhill and A. N. R. Law. 2003. Agent-based simulation in the study of social dilemmas. *Artificial Intelligence Review* **19**(1): 3-92.

Halloran, E. M., I. M. Longini, N. Azhar and Y. Yang. 2002. Containing bioterrorist smallpox. *Science* **298**: 1428-1432.

Kaplan, E. H. 1991. Mean-max bounds for worst-case endemic mixing models. *Mathematical Biosciences* **105**(1): 97-109.

Kaplan, E. H., D. L. Craft and L. M. Wein. 2002. Emergency response to a smallpox attack: The case for mass vaccination. *Proc. of the National Academy of Sciences* **99**(16): 10935-10940.

Kaplan, E. H. and Y. S. Lee. 1990. How bad can it get? Bounding worst case endemic heterogeneous mixing models of HIV/AIDS. *Mathematical Biosciences* **99**: 157-180.

Kaplan, E. H. and L. M. Wein. 2003. Smallpox bioterror response. *science* **300**: 1503.

Keeling, M. J. 1999. The effects of local spatial structure on epidemiological invasions. *Proc. of the Royal Society of London Series B-Biological Sciences* **266**(1421): 859-867.

Koopman, J. S. 2002. Contolling smallpox. *Science* **298**: 1342-1344.

Koopman, J. S., G. Jacquez and S. E. Chick. 2001. New data and tools for integrating discrete and continuous population modeling strategies. *Population health and aging*. **954:** 268-294.

Levinthal, D. and J. G. March. 1981. A model of adaptive organizational search. *Journal of Economic Behavior and Organization* **2**: 307-333.

Lipsitch, M., T. Cohen, B. Cooper, et al. 2003. Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**(5627): 1966-1970.

Mahajan, V., E. Muller and Y. Wind. 2000. *New-product diffusion models*. Boston, Kluwer Academic.

Murray, J. D. 2002. *Mathematical biology*. 3rd ed. New York, Springer.

Nowak, M. A. and K. Sigmund. 2004. Evolutionary dynamics of biological games. *Science* **303**(5659): 793-799.

Parunak, H., R. Savit and R. Riolo. 1998. Agent-based modeling vs. Equation-based modeling: A case study and users' guide. *Multi-Agent Systems and Agent-Based Simulation* **1534**: 10-25.

Picard, N. and A. Franc. 2001. Aggregation of an individual-based space-dependent model of forest dynamics into distribution-based and space-independent models. *Ecological Modeling* **145**(1): 69-84.

Riley, S., C. Fraser, C. Donnelly, et al. 2003. Transmission dynamics of the etiological agent of sars in hong kong: Impact of public health interventions. *Science* **300**(5627): 1961-1966.

Rogers, E. M. 2003. *Diffusion of innovations*. New York, Free Press.

Samuelson, P. 1939. Interactions between the multiplier analysis and the principle of acceleration. *Review of Economic Statistics* **21**: 75-79.

Schelling, T. C. 1978. *Micromotives and macrobehavior*. New York, Norton.

Strang, D. and N. B. Tuma. 1993. Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology* **99**(3): 614-639.

Tesfatsion, L. 2002. Economic agents and markets as emergent phenomena. *Proc. of the National Academy of Sciences of the United States of America* **99**: 7191-7192.

Urban, G. L., J. R. Hauser and H. Roberts. 1990. Prelaunch forecasting of new automobiles. *Management Science* **36**(4): 401-421.

Watts, D. J. 1999. *Small worlds*. Princeton, Princeton University Press.

Watts, D. J. and S. H. Strogatz. 1998. Collective dynamics of "small-world" networks. *Nature* **393**(4 June): 440-442.

Wolfram, S. 2002. *A new kind of science*. Champaign, IL, Wolfram Media.

Table 1 Base case parameters. The Online material documents the construction of each type of network used in the AB simulations.

| Parameter | | Value | Units |
|---|---|---|---|
| Total population | N | 200 | Person |
| Contact rate, Exposed | $c_{ES}$ | 4 | 1/Day |
| Contact rate, Infectious | $c_{IS}$ | 1.25 | 1/Day |
| Infectivity, Exposed | $i_{ES}$ | 0.05 | Dimensionless |
| Infectivity, Infectious | $i_{IS}$ | 0.06 | Dimensionless |
| Average incubation time | $\varepsilon$ | 15 | Day |
| Average duration of illness | $\delta$ | 15 | Day |
| Basic reproduction rate | $R_0 = (c_{ES}*i_{ES}*\varepsilon + c_{IS}*i_{IS}*\delta)$ | 4.125 | Dimensionless |
| Average links per node | k | 10 | Dimensionless |
| Prob of long-range links (SW network) | $p_{sw}$ | 0.05 | Dimensionless |
| Scaling exponent (scale-free network) | $\gamma$ | 2.61 | Dimensionless |

Table 2. Standard deviation of expected contacts for each network and heterogeneity condition. $H_{=}$, Homogeneous; $H_{\neq}$, Heterogeneous case. Mean contact rate is equal for all settings.

| Standard Deviation of Expected Contacts (contacts/person/day) | | Network Structure | | | | |
|---|---|---|---|---|---|---|
| | | Uniform | Random | Scale Free | Small World | Lattice |
| | $H_{=}$ | 0 | 0.59 | 1.58 | 0.17 | 0 |
| | $H_{\neq}$ | 1.95 | 2.58 | 6.42 | 2.13 | 2.09 |

Table 3. Estimated parameters for the DE model fitted to the mean of the symptomatic population in each AB scenario. Compare to base case parameters in Table 1.

| | Uniform | | Random | | Small World | | Scale-free | | Lattice | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $H_{=}$ | $H_{\neq}$ | $H_{=}$ | $H_{\neq}$ | $H_{=}$ | $H_{\neq}$ | $H_{=}$ | $H_{\neq}$ | $H_{=}$ | $H_{\neq}$ |
| **Infectivity of Exposed $i_{ES}$** | 0.050 | 0.054 | 0.045 | 0.056 | 0.042 | 0.076 | 0.034 | 0.039 | 0.038 | 0.036 |
| **Infectivity of Infectious $i_{IS}$** | 0.024 | 0 | 0 | 0.01 | 0 | 0.011 | 0.0001 | 0.0037 | 0.0074 | 0.0038 |
| **Average Incubation Time $\varepsilon$** | 13.8 | 17.4 | 12.5 | 7.4 | 11.7 | 4.5 | 10.0 | 8.3 | 5.6 | 6.5 |
| **Average Duration of Illness $\delta$** | 15.3 | 14.3 | 16.6 | 17.2 | 17.1 | 18.3 | 26.4 | 23.7 | 30.0 | 30.0 |
| **Implied $R_0 = c_{ES}*i_{ES}*\varepsilon + c_{IS}*i_{IS}*\delta$** | 3.21 | 3.78 | 2.25 | 1.87 | 1.96 | 1.63 | 1.37 | 1.40 | 1.13 | 1.07 |

Figure 1. Top: DE model with base case parameters (Table 1), showing the prevalence fraction of each disease stage. Bottom: Typical simulation of the equivalent AB model in the heterogeneous condition of the scale free network.

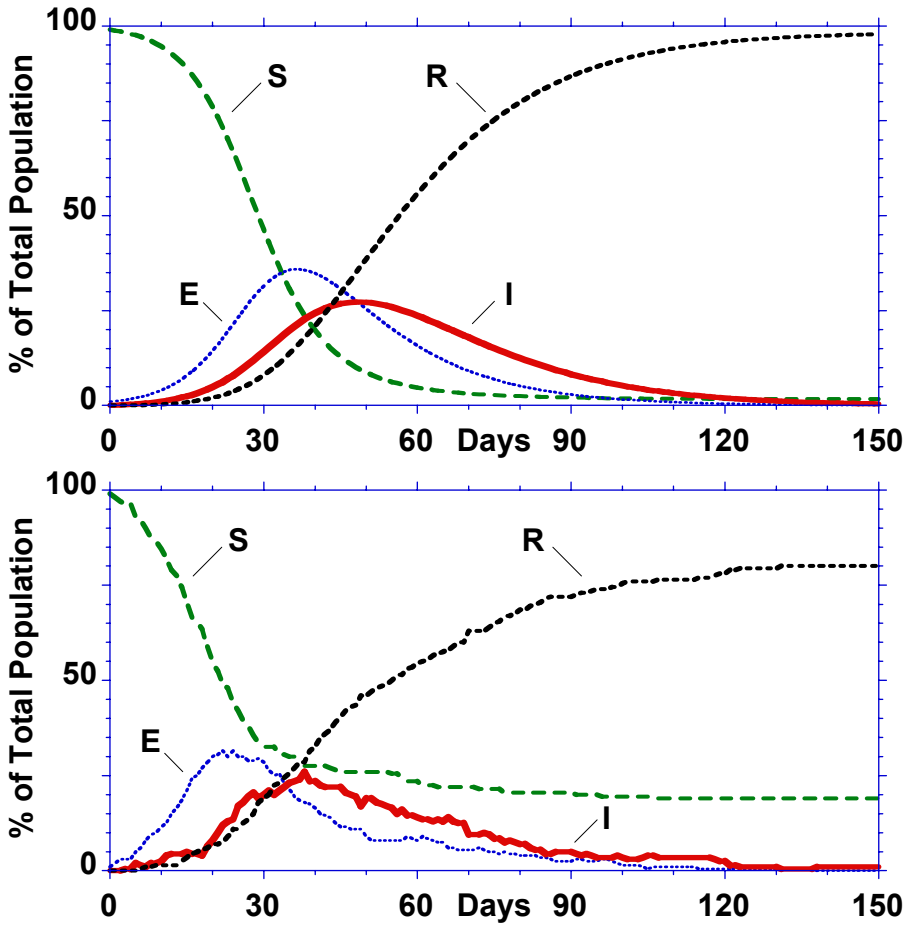


Figure 2 (next page): Comparison of the DE SEIR model to the corresponding AB model. Graphs show symptomatic cases as % of population (I/N). Each panel shows 1000 simulations of the AB model for each network and heterogeneity condition. The thick line ($DE_B$) is the base case DE model with parameters as in Table 1. $AB_M$ shows the mean of the AB simulations; $DE_F$ denotes the best fitting DE model ($AB_M$ and $DE_F$ are very close in all cases). Also shown are the envelopes encompassing 50%, 75% and 95% of the AB simulations. Vertical scales differ.

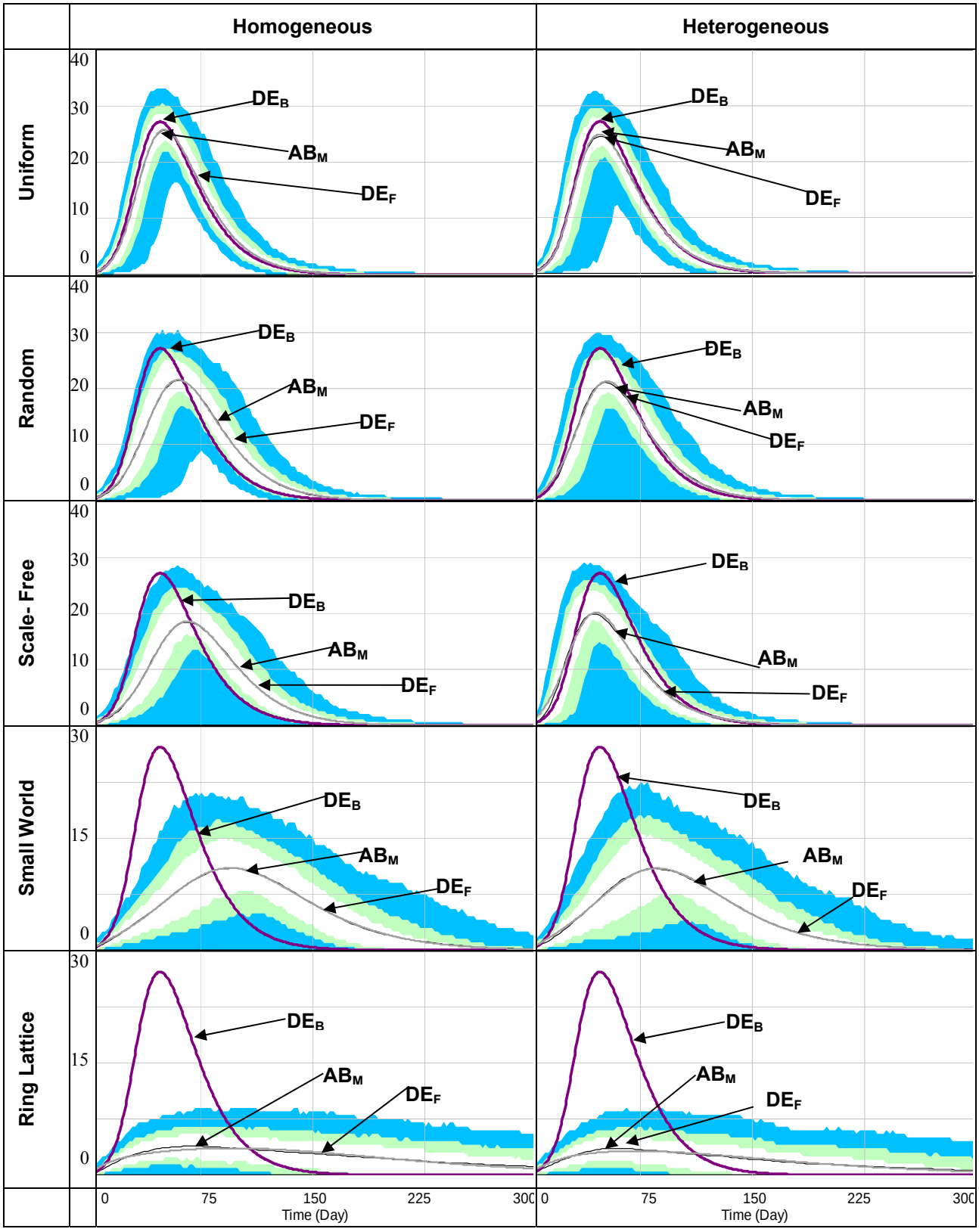
Homogeneous
Heterogeneous
Uniform
Random
Scale- Free
Small World
Ring Lattice
DE_B
AB_M
DE_F
40
30
20
15
10
0
0
75
150
225
300
Time (Day)

Table 4. Diffusion fraction F (final fraction of population infected). In the base case DE model F = 0.984. */**/*** indicates the DE simulation falls outside the 90/95/99% confidence bound defined by the ensemble of AB simulations. The fraction of simulations with F < 0.10 indicates early burnout of the epidemic.

| | | Uniform | Random | Scale Free | Small World | Lattice |
|---|---|---|---|---|---|---|
| Homo-geneous | % F < 0.10 | 1.0 | 1.8 | 3.9 | 5.6 | 9.3 |
| | Mean (σ) | 0.933 (0.131) | 0.870 (0.194) | 0.796 (0.201) | 0.825 (0.252) | 0.423** (0.277) |
| | Fitted DE | 0.954 | 0.854 | 0.787 | 0.502 | 0.262 |
| Hetero-geneous | % F < 0.10 | 1.9 | 4.7 | 5.9 | 8.8 | 15.4 |
| | Mean (σ) | 0.973 (0.098) | 0.947 (0.129) | 0.896 (0.182) | 0.893 (0.220) | 0.558 (0.311) |
| | Fitted DE | 0.975 | 0.761 | 0.660 | 0.519* | 0.200 |

Table 5. Mean and standard deviation of peak time $T_p$ and peak prevalence $I_{max}$ (% of population) for different networks scenarios, across 1000 simulations of the AB model. For the DE model, $T_p$ = 49 days and $I_{max}$ = 22.5%. */**/*** indicates the base DE falls outside the 90/95/99% confidence bounds for the AB simulations. Also reported are $T_p$ and $I_{max}$ for the DE model fitted to the mean of each AB ensemble.

| | | | Uniform | Random | Scale Free | Small World | Lattice |
|---|---|---|---|---|---|---|---|
| Peak Time, $T_p$ (Days) | Homogeneous | Mean (σ) | 50.2 (9.3) | 59.6 (14.0) | 65.0 (19.0) | 101.2 (41.9) | 105.4 (78.9) |
| | | Fitted DE | 51.0 | 60.5 | 66.5 | 95.4 | 95.6 |
| | Heterogeneous | Mean (σ) | 47.0 (10.6) | 50.1 (14.1) | 42.1 (14.3) | 88.9 (40.8) | 89.7 (73.9) |
| | | Fitted DE | 48.0 | 53.1 | 45.8 | 86.5 | 72.0 |
| Peak Value, $I_{max}$ (%) | Homogeneous | Mean (σ) | 29.6 (4.0) | 26.5 (4.6) | 23.5 (5.5) | 16.4** (5.2) | 7.3*** (2.7) |
| | | Fitted DE | 25.8 | 21.5 | 18.6 | 11.0 | 3.6 |
| | Heterogeneous | Mean (σ) | 28.8 (4.8) | 25.6 (6.2) | 24.6 (6.6) | 16.4* (5.8) | 6.7*** (2.7) |
| | | Fitted DE | 24.9 | 21.3 | 20.1 | 11.0 | 3.2 |