

# An Integrated Modeling Environment to Study the Coevolution of Networks, Individual Behavior, and Epidemics

Chris Barrett, Keith Bisset, Jonathan Leidig,  
Achla Marathe, and Madhav Marathe

■ We discuss an interaction-based approach to study the coevolution between sociotechnical networks, individual behaviors, and contagion processes on these networks. We use epidemics in human populations as an example of this phenomenon. The methods consist of developing synthetic yet realistic national-scale networks using a first-principles approach. Unlike simple random graph techniques, these methods combine real-world data sources with behavioral and social theories to synthesize detailed social contact (proximity) networks. Individual-based models of within-host disease progression and interhost transmission are then used to model the contagion process. Finally, models of individual behaviors are composed with disease progression models to develop a realistic representation of the complex system in which individual behaviors and the social network adapt to the contagion. These methods are embodied within *Simdemics*, a general-purpose modeling environment to support pandemic planning and response. *Simdemics* is designed specifically to be scalable to networks with 300 million agents; the underlying algorithms and methods in *Simdemics* are all high-performance computing-oriented methods. New advances in network science, machine learning, high-performance computing, data mining, and behavioral modeling were necessary to develop *Simdemics*.

*Simdemics* is combined with two other environments, *Simfrastructure* and *Didactic*, to form an integrated cyber environment. The integrated cyber environment provides the end user with flexible and seamless Internet-based access to *Simdemics*. Service-oriented architectures play a critical role in delivering the desired services to the end user. *Simdemics*, in conjunction with the integrated cyber environment, has been used in more than a dozen user-defined case studies. These case studies were done to support specific policy questions that arose in the context of planning the response to pandemics (for example, H1N1, H5N1) and human-initiated bioterrorism events. These studies played a crucial role in the continual development and improvement of the cyber environment.

Contagions (or diffusion) over complex networks are pervasive in social and physical sciences. Three recent global-scale contagions that have received attention in the media as well as academic circles are (1) current and past financial contagions (Desai 2003),<sup>1</sup> (2) failure of the coupled infrastructure system caused by power-grid failure, for example, the Northeast blackout of 2003,<sup>2</sup> and (3) potential pandemics caused by influenzalike illness (Halloran et al. 2008; Germann et al. 1983). Individuals, institutions, and governments could not prevent the Northeast blackout. However, they are aggressively developing interventions to control the current financial contagion and responding to reduce the economic burden and human suffering of the current H1N1 outbreak. Developing high-resolution computational models to reason about these systems is complicated and scientifically challenging for at least three reasons. First, these systems are extremely large (for example, pandemic planning at the scale of the United States, requiring models with 300 million agents). Second,

the contagion, the underlying interaction network (consisting of both human and technical elements), the public policies, and the individual behaviors coevolve. This makes it nearly impossible to apply standard model-reduction techniques that have been successfully used to study physical systems. Finally, in practical situations, multiple contagion processes simultaneously coevolve.

Here we describe *Simdemics*, an interaction-based multiagent approach to study diffusion processes in very large sociotechnical networks. *Simdemics* is an example of a disaggregated network-based modeling approach in which interactions between every pair of individuals connected in the social contact network are represented. It uses a realistic, synthetic representation of the underlying social contact network. It is based on the idea that a better understanding of the characteristics of the underlying network and individual behavioral adaptation can give better insights into contagion dynamics and response strategies. It should be noted that *Simdemics* by itself does not prescribe a specific level of quality for the social contact networks. The necessary quality (in terms of accuracy, resolution, and fidelity) of the networks is determined by the questions that we aim to address.

*Simdemics* can be used to study a much larger class of diffusion processes. These include epidemic processes in ecologies; the spread of certain non-infectious diseases such as obesity and smoking; the spread of fads, conventions, norms, and information in social systems; the spread of worms and malware in communication networks (Channakeshava, et al. 2009). Here, we will confine our discussion to the spread of infectious diseases in human populations. Besides their obvious societal importance, epidemics serve as an excellent example of diffusion processes over interaction networks. Within the infectious disease context, *Simdemics* details the demographic and geographic distributions of contagion spread. It also provides decision makers with information about the consequences of an outbreak, the resulting demand for health services, and the feasibility and effectiveness of various response options. A unique feature of *Simdemics* is the size and scale of social and ecological systems that can be analyzed through its use. Planning and responding to the threat of pandemics presents an important societal and public health challenge. Public health authorities around the world are far more prepared to respond to pandemic threats now than they have ever been in the past. However, a number of modern trends continue to make this a vexing problem. These include (1) a larger global population and increased urbanization leading to a higher density of individuals within cities; (2) higher levels of long-distance travel, including international travel; and (3)

increased numbers of elderly individuals and individuals with chronic medical conditions.

## An AI Perspective

We first began to work on this class of problems 15 years ago. Our goal was to develop new computational and analytical techniques to understand and reason about very large complex sociotechnical systems. Solutions to these problems required us to borrow ideas from diverse areas in science and engineering, including network science, high-performance computing, urban planning, economics, nonlinear systems, algorithmics, statistics, and so on. Over time, it became increasingly clear that an AI perspective on complex sociotechnical systems was extremely valuable; several ideas and concepts in AI proved to be the conceptual glue that was critical in developing an integrated solution. These ideas include multiagent systems for representing large complex systems; causality, reasoning, and inference in complex systems; and human-computer interaction. We have tried to highlight this perspective throughout the article.

## Basic Approach

Traditional epidemiological research has focused on rate-based differential equation models of completely mixed populations, that is, all the individuals are allowed to interact with each other (see Bailey [1975] and Hethcote [2000] for a comprehensive discussion on this subject). An attractive feature of this modeling approach is that it allows one to obtain analytical expressions for a number of interesting parameters such as the numbers of infected, recovered, and susceptible individuals in a population. But such a modeling approach does not capture the complexity of human interactions that serve as a mechanism for disease transmission. Typically the number of different subpopulation types considered is small (for analytical tractability), and parameters such as mixing rate and reproductive number are either unknown or hard to observe. Over the past several years, large-scale, individual-based, disaggregated models have been studied (Carley et al. 2006; Eubank et al. 2004; Ferguson et al. 2005; Ferguson et al. 2006; Halloran et al. 2008). These new models use an endogenous representation of individual agents together with explicit interactions between these agents to generate and capture the disease spread across social networks.

*Simdemics* differs in important ways from recent individualized agent-based approaches. First, the methodology used to synthesize social contact networks in *Simdemics* generates far more realistic social contact networks. The networks display the spatial as well as demographic heterogeneity that is absent in the prior work of Ferguson et al. (2006), Germann et al. (1983), and Longini et

al. (2005). Networks lacking this heterogeneity use stylized templates for small communities that are then connected in a regular manner to form larger communities and cities. Thus, the spatial diversity of locations in these cities is absent in stylized networks. The important structural features exhibited by more realistic social contact networks crucially affect the disease dynamics and the effect of various response strategies (see Barrett et al. [2009b] for more discussion on this). The structure and the details of the social networks have important implications on the performance of the parallel simulations as well. Although Germann et al. (1983), Longini et al. (2005), and Parker (2007) claim national- and global-scale models for simulating pandemics on clusters, their ability to carry out these computations crucially exploits the spatial and demographic homogeneity present in their networks. Second, although in theory various interventions can be represented by Ferguson et al. (2006), Germann et al. (1983), and Longini et al. (2005), these representations are affected by the underlying models of networks. In Simdemics, a substantial effort has been made to find a realistic representation of various interventions and to develop a formal language structure for specifying and executing various contagions and interventions (Barrett et al. 2009b, Bisset et al. 2009b). The variety of interventions combined with the flexibility and naturalness of their specification is a unique feature of Simdemics. Behavioral representation for systems by Ferguson et al. (2006) and Germann et al. (1983) is not endogenous and does not appear to be suitable to study the coevolutionary dynamics. Finally, Simdemics, in conjunction with two additional computational systems, namely, *Didactic* and *Simfrastucture*, provides seamless access to the modeling capabilities of Simdemics. A number of important analyses can be done on the fly; this includes isolating the main effects, studying interaction among various interventions, and so on. This represents a radical departure in the field of computational epidemiology. Such a capability is, at present, not generally available in other individualized urban-scale models (see Eubank et al. [2004] for more details).

### Brief Description of Methods

The overall approach consists of four distinct models: (1) a model for creating a set of synthetic individuals, (2) a model for generating (time varying) interaction networks, (3) a model for simulating the epidemic process, and (4) a model for representing and evaluating interventions and public policies. Mathematically, these steps can be represented by a combination of (1) a discrete dynamic system framework that captures the coevolution of disease dynamics, social network, and individual behavior (first three steps) and (2) a partially

observable Markov decision process (POMDP) that captures various control and optimization problems formulated on the phase space of this dynamic system. We describe these steps in detail below.

In step 1, a synthetic urban population is created by integrating a variety of databases from commercial and public sources into a common architecture for data exchange. The process preserves the confidentiality of the individuals in the original data sets, yet produces realistic attributes and demographics for the synthetic individuals. The synthetic population is a set of synthetic people and households, located geographically, each associated with demographic variables recorded in the census. Joint demographic distributions are reconstructed from the marginal distributions available in typical census data using an iterative proportional fitting (IPF) technique. Each synthetic individual is placed in a household with other synthetic individuals. Each household is located geographically using land-use data and data pertaining to transportation networks. The process guarantees that a census of our synthetic population is statistically indistinguishable from the original census. Since the population is synthetic, the privacy of individuals is protected. The basic process can be extended to assign other demographic attributes using additional data sources. For example, in recent work we assigned individuals' mobile devices using commercially available survey data.

In step 2, a set of activity templates for households is determined, based on several thousand responses to an activity or time-use survey. The modeling methodology is called activity-based travel-demand modeling and is now accepted as the de facto standard in transportation science (Beckman, Baggerly, and McKay 1996). Our early work in this area (Barrett et al. 2001) played an important role in the development of this methodology. The activity templates include the activities each household member performs and the time of day they are performed. Each synthetic household is then matched with one of the survey households, using a decision tree based on demographics such as the number of people in the household, number of children of various ages, and income. The synthetic household is assigned the activity template of its matching survey household. For each household and each activity performed by this household, a preliminary assignment of a location is made based on observed land-use patterns, tax data, and so on. A social network is formed out of these activities as agents mix together with some level of contact, at the locations visited, at various times throughout the day. The dynamic social contact network represented by a (vertex and edge) labeled bipartite graph  $G_{PL}$ , where  $P$  is the set of people and  $L$  is the set of locations. If a person  $p \in$

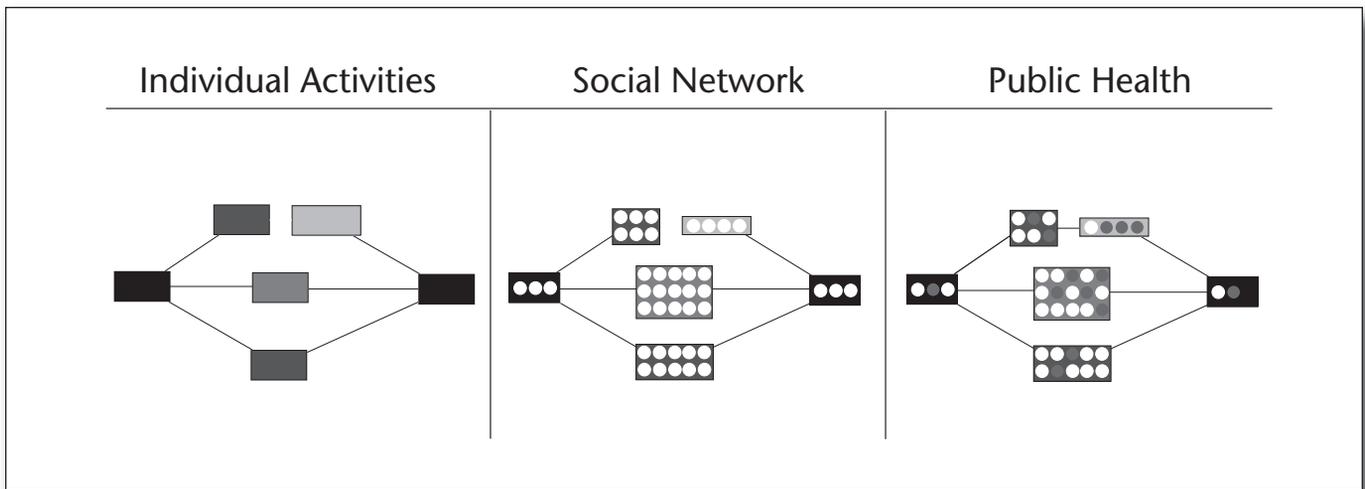


Figure 1.

Schematic diagram showing how the activities from three individuals in a household (left) are combined to obtain a social contact network derived from their human-human contacts throughout a daily routine (middle). A contagion process (right) is then laid down over this network to model the spread of the disease from an infected individual (shaded) to two other locations.

$P$  visits a location  $\ell \in L$ , there is an edge  $(p, \ell, label) \in E(G_{PL})$  between them, where *label* is a record of the type of activity of the visit and its start and end points. The edges between individuals capture physical proximity between the individuals. The existence of such edges and their attributes often depends on the disease we wish to study. For example, in a case of influenzalike illness, two individuals have an edge if they are within certain proximal distance of one another for a certain period of time. In case of sexually transmitted diseases, an edge implies intimate physical contact. See figure 1 for a representation on how a small family's activities during the course of a day are composed to form a dynamic social network. Please note that it is impossible to build such a network by simply collecting field data; the use of generative models to build such networks is a unique feature of this work.

Step 3 consists of developing a computational model for representing the disease within individuals and its transmission between them. The model can be viewed as a *networked finite probabilistic timed transition system* (PTTS). Each individual is associated with a timed probabilistic finite state machine. A PTTS can be represented using appropriate stochastic local functions (with possibly additional nodes to simulate a clock). A PTTS representing an individual interacts with or is coupled to PTTSs of other individuals that are neighbors in the social contact network. PTTSs are an extension of the well known finite state machines (FSMs) with two additional features: the state transitions are probabilistic or timed (or both). In other words, the end point of state transitions may be chosen probabilistically or deterministically; the transi-

tions may be timed, that is, they may occur at a specified time after the previous transition, or there may be a fixed probability of transition for each discrete time interval. In principle, each agent can have a different probabilistic timed transition system. More commonly, subsets of people determined by demographics such as age are assigned a single parameterized PTTS, and individuals are assigned values for the parameters at random from an appropriate distribution. The coupling between individual PTTSs is derived from the social contact network. The dynamics of disease diffusion processes can now be simulated using the above coupled model. We have developed fast high-performance computing-based algorithms to achieve this (Barrett et al. 2008, Bisset et al. 2009a, 2009b).

Step 4 consists of representing individual behaviors and their adaptations, public policies, and interventions. Individual behaviors are represented and analyzed using finite PTTSs and are based on well accepted and recently proposed socioeconomic theories of individual and collective behaviors. Public policies are represented and analyzed using partially observable Markov decision processes (POMDPs). This allows us to capture sequential decision-making processes related to studying the efficacy of various interventions and behaviors of individual agents in response to their perception of disease spread. The POMDP is exponentially larger than the problem specification, and, in general, determining an optimal solution is an intractable problem. We thus resort to efficient simulations. A key concept is that of *implementable policies*—policies or interventions that are implementable in the real world. Figure 2 shows a schematic of a multinet, multitheory model

that results at the end of these steps. In the real world, disease diffusion and information diffusion are based on different models and potentially use different social networks.

Although interventions and individual behavioral changes were studied in the past, their implementations were carried out in an ad hoc manner. Computational efforts concentrated first on developing fast methods for simulating disease progression. Fast methods for disease progression that do not consider the interaction with policies and individual behaviors are insufficient due to coevolution within the underlying system. A modeling environment is required in which simulation of disease spread is carried out in lockstep with the interventions that are instituted, and the resulting effects on the network structure, and individual attributes.

This coevolution is where the classical analogy between percolation and epidemics breaks down. To date, most of the research efforts in building large-scale models have represented this coevolution in an ad hoc manner. Recent advances in artificial intelligence and operations research are likely to be useful in representing and analyzing this aspect of the model. Figure 2 displays the interaction between multiple networks and the possibility of affecting contagion. Two mathematical models are at play: coevolving graphical discrete dynamic systems, or CGDDSs, form the basis for simulating the coevolving dynamics, while POMDPs and  $n$ -player games are suitable for representing and reasoning about interventions and individual behaviors. This requires computations over the configuration space of these dynamic systems. Note that the configuration space over which POMDPs have to reason is exponentially larger than their representations. Simulations that implement a CGDDS are computationally as well as conceptually much harder than simulations that simply implement basic percolation processes. In general, simulations implementing a CGDDS comprise three operations that are repeated: simulate a state transition of disease progression over the network; evaluate the state of the disease, and test whether one or more triggering conditions hold; and apply applicable pharmaceutical or non-pharmaceutical interventions that change the social network structure or individual disease models. The composite system in Simdemics facilitates this type of representation and this multistep process. The triggering condition can be based on an individual or on a subpopulation and may involve evaluating a complicated function. Interventions can be applied to individuals or to sets of individuals (for example, school-aged children).

### Model Validation

Usually the term *model validation* is associated with

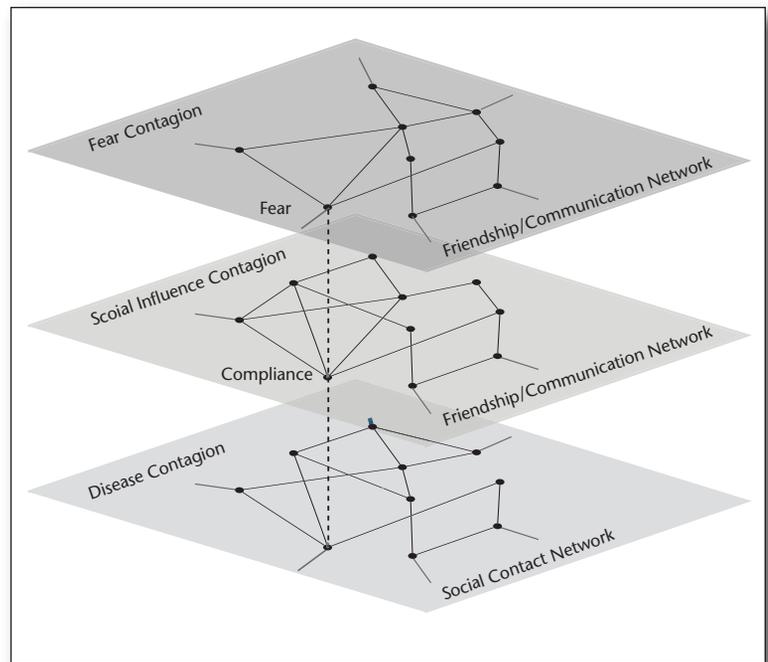


Figure 2.

Schematic diagram depicting multitheory, multinet network diffusion processes. The lowest level depicts a social proximity network over which aerosol-borne infectious disease spreads. The individual behavioral model captures the host disease progression model. The middle layer denotes the social influence contagion: individuals might comply with nonpharmaceutical interventions (for example, face masks) based on social influences, public information, and policies. The top layer denotes the fear of contagion that results when disease outbreak happens. Compliance in the social influence layer depends on the state of disease contagion as well as the level of fear. Behavioral change causes change in the networks as well as contagion dynamics.

a relatively simple-seeming view of predictive validity — state-space predictions by the model match the measured data. Furthermore, in large social simulations, these predictions are usually postdictions of historical information, for example, the infection time series of the 1968 flu season. Although this kind of examination of the model can be useful, it can also be misleading and is not adequate. One immediate implication of the previous section is that the configuration space (which is the essence of causal structure for the evolving system) is important to understanding the system being represented as well as the modeled representation of that system. However, any measured *real-world* data is incapable of capturing this structural range—only those modes that happened in the real world appear in the measured data. Thus the process of postdiction is, alone, inadequate. Additionally, usually while postdicting, insufficient information about the context is available to properly specify the initial and structural conditions that would allow the model to predict. As a result, high-dimensional models are often fit to relatively sparse data. In this sense, the

Urban Population Mobility Models	(Bowman et al. 1998; TRBC 1995 2003; TRB ; Barrett et al. 2009b)
Natural Disease History	(Elveback et al. 1976; Longini et al. 2005; Halloran et al. 2008; Bailey 1975; Hethcote 2000)
Transmission Models	(Longini et al. 2005; Halloran et al. 2008; Hethcote 2000)
Social Network Models	(Newman 2003; Eubank et al. 2004; Halloran et al. 2008)
Types of Interventions	(Halloran et al. 2002; 2008; Ferguson et al. 2005; 2006)

Table 1. Summary of Specific Models and Modeling Approaches Used in Simdemics.

The overall system is obtained by composing these models.

occurrence of a fit can be misleading because the inverse being solved is nonunique. In both cases, explanatory power is really at issue. In decision making a causal basis for the choice of the best option is more relevant than any particular kind of detailed prediction of state. This raises fundamental issues for the idea of validation and is the focus of ongoing work. Related concerns have been discussed by historians and philosophers (Oreskes, Shrader-Frechette, and Belitz 1994; Oreskes 2000). See table 1 for an overview of our modeling approaches.

Extensive efforts have been made to validate the overall approach and specific components of the model. This includes (1) structural validity of models, (2) matching the data produced to field data, and (3) functional validation and formal specifications of these models for software verification (Barrett 2009a). Several models used here are based on work done by other researchers (see table 1).

### Applications of AI Concepts

Simdemics uses a number of concepts studied in traditional and contemporary AI literature. This includes multiagent systems, social network analysis, Markov decision processes, and large  $n$ -player games (Barrett, Eubank, and Marathe 2006; Bailey-Kellogg, Ramakrishnan, and Marathe 2006). However, the practical use of this tool prompted the investigation of several new basic and applied research questions. For example, we had to develop new efficient algorithmic and machine-learning techniques to generate and analyze dynamic social networks and simulate contagion on these dynamic social networks (Eubank et al. 2004). These algorithms were designed and implemented so that they can scale to over 100 million node networks and can be mapped onto 100–1000 processor-distributed memory architectures (Barrett, Eubank, and Marathe 2006; Eubank et al. 2004). Similarly, scalable data-mining methods are being developed to analyze the vast data sets produced by Simdemics (Bailey-Kellogg, Ramakrishnan, and Marathe 2006). These scalable simulations and

mining algorithms form the basis of practical and usable decision-support systems that have been built and that are being continually enhanced. As presented by Barrett et al. (2009a), Simdemics uses and extends several topics studied in AI: (1) detailed multiagent models, (2) synthesis of large urban social and relational networks, (3) concise, realistic behavioral modeling, and (4) the theory of graphical dynamic systems and games. We give three examples to illustrate these ideas.

*Scalable Multiagent Systems.* Developing scalable multiagent systems for studying a real-world problem remains an active area of research in AI (Wooldridge 2002; Shoham and Leyton-Brown 2008). We first note that the notion of agents as used here should be distinguished from the notion of entity and actor. Agents are actors that have intent or motive and thus require an individuated behavioral representation that is rich enough for the problem at hand and yet lightweight so it can be scaled to large populations of interacting individuals. We require three new ideas to achieve this: (1) Parametric representation of individual behaviors and local actions: A single basic algorithm is used for each agent and the behavioral variation is obtained by randomization and agent specific attributes. (2) Behavioral decomposition: We use automata theoretic techniques to represent each kind of local function associated with an agent by a separate automaton (algorithm) and then use generalized cross-product-like construction to obtain a composite behavior. Thus each agent is a multitheory intentional entity. (3) Finally, we introduce the notion of “unencapsulated agents”: In this notion of agency, intent and behavioral loci are distributed and do not necessarily reside within a single software object. For example, we have only one within-host disease progression model (represented using a probabilistic timed transition system). However the specific manifestation of a disease-within-host is a function of demographic variables associated with the individual, computed elsewhere asynchronously and used as parameters that affect the local state transition properties. Sim-

ilarly, behavioral models that are used for individual decision making in the event of epidemics are parameterized. An individual's overall representation is composed of a within-host disease model and the individualized behavioral model. The description of the agent, or what an agent does, is not confined to these local functions. The agent's interaction with other agents defines its overall behavioral description. In this sense, the idea bears certain resemblance to the notion of (non-)modularity of functions in cognitive science (Fodor 1983) and the concept of population coding (McIlwain 2001) in neuroscience. So while local functions and the state attributes associated with an agent determine its local dynamic evolution, the phase space encodes the system behavior and is necessary to understand the behavior of an individual agent.

*Coevolving Graphical Models.* Researchers in AI have been studying graphical models for solving problems pertaining to areas such as Bayesian inference, games, and constraint-satisfaction problems. Graphical models of Bayesian inference and games have been proposed and studied in AI to capture the network structure inherent in certain applications (Pearl 2000, Kearns 2007, Pearl 2009, Russell and Norvig 2009). Dynamic processes such as epidemics on social networks provide a new class of applications (Kearns 2007, Lauritzen 1996, Pearl 2009) in this regard. Two aspects make this a very challenging problem. First, in order to solve real-world problems we need to deal with graphs with 10 million to 300 million vertices and 500 million to 15 billion edges. Second, realistic social networks do not have a treelike property that has been exploited in earlier work to obtain tractable algorithms for computing Nash equilibria (Lauritzen and Spiegelhalter 1988, Pearl 1988, Kearns 2007). They are not even small-world networks or scale-free networks as defined in the current literature (Barrett, Eubank, and Marathe 2006; Eubank et al. 2004). Understanding the structure of these networks and exploiting this structure for designing efficient computational solutions is an important research question. Another interesting direction for further research is to extend the notion of graphical games and inference problems to coevolving graphical games and inference problems in which the underlying network is changing due to the decisions taken by individual agents, which in turn change due to the interaction predicated by the network. An example of an important application-motivated inference question is the following: Given sparse surveillance information about a small subset of infected individuals in a region, find the index cases, that is, the initial set of individuals who became infected.

*Intelligent Query-Processing Systems and Computational Steering of Simulation-Based Experiments.* As

we began to answer real-world policy questions, it became progressively clear that easy-to-use web-based systems that employ our models would provide an appropriate mechanism for delivering the results. As a result, we have been developing web-based services so that complex high-performance computing-based models can be used by analysts who are not computing experts. The web-based system is called Didactic. In essence, it provides users a way to specify a factorial style adaptive experimental design. Using Didactic, a public health analyst can carry out a number of what-if experiments, assess the trade-offs between various intervention strategies, and more generally develop a better understanding of how the outbreak is likely to spread over social networks. In many ways this effort is similar to the topic of cognitive prostheses (Ford 2008)—computational systems for analysts to obtain better situational awareness using Simdemics-like environments in conjunction with surveillance information. We are extending the system to become more interactive and to allow for computational steering of experiments. This requires methods for coordinating resource discovery of computing and data assets; AI-based techniques for translating user-level requests to efficient work flows; reusing data sets whenever possible and spawning computer models with required initial parameters and coordination of resources among various users. Our initial work on this topic is described in Atkins et al. (2008), Atkins, Barrett, and Marathe (2009), and Barrett et al. (2008). A natural formal environment to represent and assess graph dynamic systems, their composition, and coevolution in both algebraic and computational terms is necessary to deal with these questions.

### Theoretical Foundations of Graph Dynamic Systems and Their Coevolution

Our formalization consists of two parts: (1) a discrete graphical dynamic system framework that captures the coevolution of disease dynamics, social networks, and individual behavior, and (2) a partially observable Markov decision process that captures various control and optimization problems formulated on the phase space of this dynamic system. The basic framework consists of the following components: (1) a collection of entities with state values and local rules for state transitions, (2) an interaction graph capturing the local dependency of an entity on its neighboring entities, and (3) an update sequence or schedule such that the causality in the system is represented by the composition of local mappings. We formalize this as follows. A coevolving graphical discrete dynamic system  $\mathcal{S}$  over a given domain  $\mathbb{D}$  of state values is a triple  $(G, \mathcal{F}, W)$ , whose components are as follows:

```

Initialize  $t = 0$ 
Repeat Until  $W$  is empty
  (i) Let  $r$  be the the first symbol in  $W$ .
  (iii) If  $r = v_i(s)$ , update the state of the vertex  $v_i$ 
  as follows:
    (a) Let degree of node  $v_i$  in  $G_t$  be  $d_t$ . Node  $v_i$ 
    evaluates  $f_{v_i, d_t}$ . (This computation uses the
    current values of the state of  $v_i$  and those of
    the neighbors of  $v_i$  in  $G_t$ .) Let  $x$  denote the
    value computed.
    (b) Node  $v_i$  sets its state  $s_{v_i}$  to  $x$ .
  (iii) If  $r = v_i(g)$ , update the edges incident on
   $v_i$  as follows:
    (a) Use current graph  $G_t$  to compute  $g_{v_i}$ .
    (b) Let  $G_{temp}$  denote the new graph.
  (ii) Set  $t = t + 1$ ,  $G_{t+1} = G_{temp}$  and delete  $r$  from
  string  $W$ .
End Repeat

```

Figure 3. Pseudocode.

1. Let  $V = \{v_i\}_{i=1}^n$  be a set of vertices, and let  $(g_i)_i$  be a vertex indexed family of graph modification functions  $g_i: \{0,1\}^n \rightarrow \{0,1\}^n$ . The functions  $(g_i)_i$  through their applications, define an indexed sequence of graphs  $G = (G_r = G_r(V_r = V, E_r))_r$  with labeled edges and vertices. The graph  $G_r$  is the *underlying contact graph* of  $\mathcal{S}$  after  $r$  applications of functions  $g_i$ . It is assumed that the edge  $\{v_i, v_j\} \in E_r$  for all  $r$  and for all  $i$ . We set  $m_r = |E_r|$ .

2. For each vertex  $v_i$  there is a set of local transition functions  $\{f_{v_i, d}\}_d$  where  $f_{v_i, d}: \mathbb{D}^d \rightarrow \mathbb{D}$ . Let  $N(i, t)$  denote the set of vertices consisting of  $v_i$  and the neighbors of  $v_i$  at time  $t$ , and let  $d_t = |N(i, t)|$ . The function used to map the state of vertex  $v_i$  at time  $t$  to its state at time  $t + 1$  is  $f_{v_i, d_t}$  and the input to this function is the state subconfiguration induced by  $N(i, t)$ .

3. The final component is a string  $W$  over the alphabet  $\{v_1(s), v_2(s), \dots, v_n(s), v_1(g), \dots, v_n(g)\}$ . The string  $W$  is a schedule. It represents an order in which the state of a vertex or the possible edges incident on the vertex will be updated. Here  $v_i(s)$  intuitively specifies that the state of the vertex  $v_i$  is to be updated;  $v_i(g)$  specifies that one or more incident edges will be updated.

From a modeling perspective each vertex represents an agent. Here we will assume that the states of the agent come from a finite domain  $\mathbb{D}$ . The maps  $f_{v_i, j}$  are generally stochastic. Computationally, each step of a CGDDS (that is, the transition from one configuration to another), involves updating either a state associated with a vertex or modifying the set of incident edges on it. The

pseudocode in figure 3 shows the computations involved in one transition.

Let  $F_{\mathcal{S}}$  denote the global transition function associated with  $\mathcal{S}$ . This function can be viewed either as a function that maps  $\mathbb{D}^n$  into  $\mathbb{D}^n$  or as a function that maps  $\mathbb{D}^V$  into  $\mathbb{D}^V$ .  $F_{\mathcal{S}}$  represents the transitions between configurations, and can therefore be considered as defining the dynamic behavior of an CGDDS  $\mathcal{S}$ . We make several observations regarding the formal model just described.

First, we will assume that the local transition functions and local graph modification functions are both computable efficiently in polynomial time. In agent-based models used in the social sciences these are usually very simple functions. Furthermore, the functions  $g_{v_i}$  need to be specified using a succinct representation, rather than a complete table, which will be exponentially larger.

Second, the edge-modification function as defined can modify in one step a subset of edges simultaneously. An alternate model might have allowed a vertex to change exactly one edge at a time. We chose the former due to the specific application we had in mind. In all our applications, when an agent decides to not go to a location (either due to location closure as demanded by public policy or due to the fear of contracting the disease), its edges to all other individuals in that location are simultaneously removed while edges to all the individuals who might be at home are added.

Third, the model is Markovian in that the updates are based only on the current state of the system; it is possible to extend the model wherein updates are based on earlier state of the system. Finally, we have assumed that there is exactly one function for each arity for each node. This can be relaxed easily; similarly these functions will, in general, be stochastic.

In the next step, we overlay a partially observable Markov decision process framework over the discrete dynamic systems framework. This allows us to discuss control and optimization methods. The discrete dynamic system provides us with a computational view of how state transitions are made. We refer the reader to Mundhenk et al. (2000) for detailed definitions and complexity theoretic results on this topic. A POMDP  $M$  consists of a finite set of states ( $S$ ), actions ( $A$ ), and observations ( $O$ ).  $s_0 \in S$  is the initial state of the system.  $t$  is the local transition function from one state to another and is probabilistic.  $o$  is the observation function that assigns to each state an observation. Finally,  $r$  is the reward function that tells the reward received when action  $a \in A$  when in state  $s$ .

Using the terminology of Mundhenk et al. (2000), our POMDP is specified succinctly; we use a dynamic systems specification rather than a circuit representation to achieve this. The states of  $M$  are

all possible vectors of vertex states and edge states (present or absent). Each vector of the underlying Markov chain  $\mathcal{M}$  is specified by a vector of length  $\binom{n}{2} + n$ , representing all the possible edges and vertices in the graph. The state transitions are obtained by composing the local functions  $f_i$  and  $g_i$  as we discussed. If these functions are probabilistic then so is the transition function for the Markov chain. Thus the chain consists of  $2^{\binom{n}{2}+n}$  states. Actions should be thought of as interventions in our context. Policies map observations to actions, and actions in a state yield reward. The reward (cost) function can be a combination of number of infected individuals and the economic and social costs of the interventions. We have two possible classes of reward functions, as is common in game theory: a systemwide reward function and a local reward function associated with each individual. Individuals attempt to maximize their local reward function (for example, the probability of the individual or a family member becoming infected), while public policy attempts to maximize the systemwide reward function (for example, the total number of people not infected). An agent-based model serves as the corresponding computational model for the POMDP. Partial observability in our context will be captured through various triggering conditions and interventions that are instituted.

## Application Case Studies and Payoff

Our cyber environment is fully deployed and under continual development. Several studies involving human networks, livestock networks, and computing networks have been conducted using this cyber environment. In an epidemiological experiment, an application determines each agent's normative social network. A diffusion process is then run over this social network as agents probabilistically infect other agents connected to them in the social network.

At the start of the simulation, a set of individuals in the social network is initially seeded with an infection. Through the use of the scenario scripting language, a researcher can model specific scenarios by developing the set of decision triggers, individual and global interventions, and agents' behavior to be modeled. The scenario can dynamically modify the social network as the simulation progresses.

Simdemics has been developed continually in a spiral R&D process over the last 12 years. It has been deployed and used in a number of user-defined studies, including recent pandemic planning studies undertaken for the U.S. Department of Homeland Security (DHS), Department of Defense (DOD), and Department of Health and Human Services (DHHS). The studies have guided

the continued evolution of Simdemics both in terms of its usability and model development. The studies also helped us identify new research questions at the interface of multiagent modeling, data mining, network science, and high-performance computing. The following are notable case studies undertaken using Simdemics.

A study was conducted on behalf of the Office of Homeland Security (OHS) to develop planning and response strategies for a smallpox-based bioterror attack (Eubank et al. 2004). In contrast to earlier results, our experiments showed that early detection and targeted interventions can be quite promising in mitigating the effects of such an attack. The study pointed to the importance of developing realistic models of social contact networks but raised new questions on finding implementable policies for targeted interventions.

Three separate studies were conducted for the DOD regarding military preparedness and force readiness. The studies elucidated how protecting a small critical subset of a larger population is fundamentally different from public health epidemiology. The studies provided guidelines for military preparedness in the event of an epidemic outbreak. The results showed the importance of early detection in implementing effective sequestration and the apparently counterintuitive result that sequestration, if implemented late, might lead to more infections rather than fewer infections.

Another recent study (Barrett 2009b) was performed to determine the social and economic impact of public and private interventions, that are typically adopted, during a flulike epidemic. The economic costs included not only the loss in productivity due to sickness but also the indirect costs incurred through disease avoidance and caring for dependents. The results showed that the most important factor responsible for preventing income loss is the modification of individual behavior which reduced the total income lost by 62 percent compared to the base case. This result highlights the importance of behavioral modifications undertaken by the private citizens based on local and global disease prevalence levels.

## An Illustrative Use Case for Influenzalike Illness

We describe an example of a case study performed at the request of one of our sponsors. Further details of the study have been described by Bisset et al. (2009b). Although the example is slightly modified from the actual study in order to simplify it and present a more complete representation of the features present in Simdemics, it provides an accurate representation of multiple studies that have been done.

The population of the state of Alabama (4.3 mil-

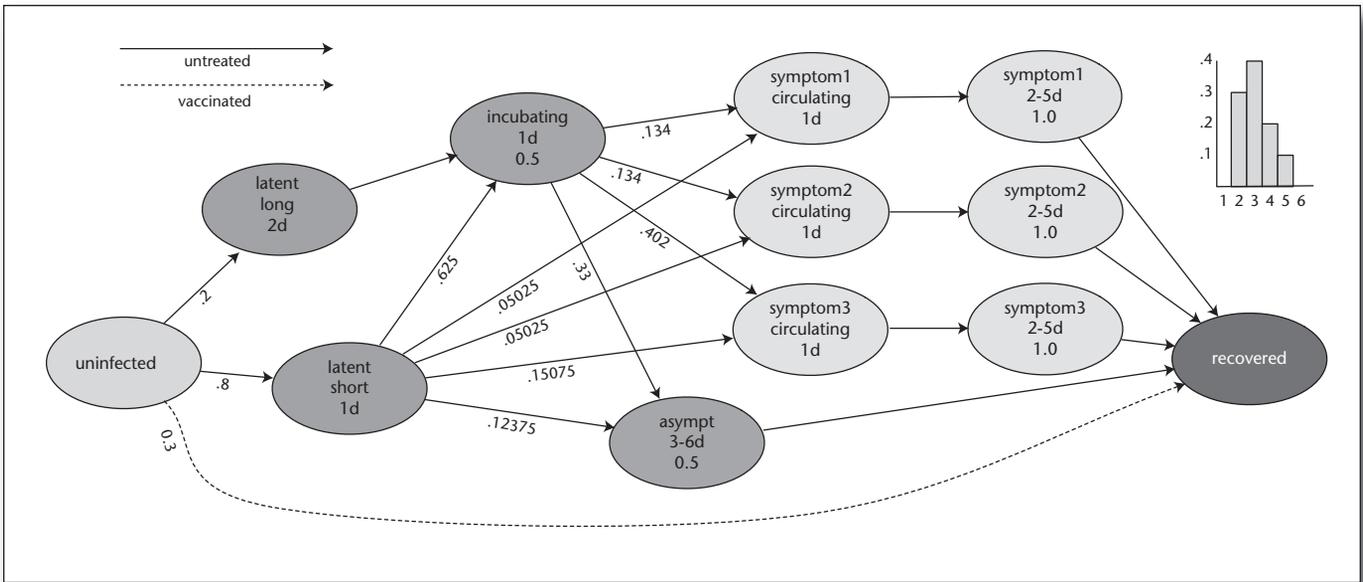


Figure 4.

PTTS for the H5N1 disease model. Ovals represent disease state and lines represent the transition between states labeled with the transition probabilities. The line type represents the treatment applied to an individual. The states contain a label and the dwell time within the state, and the infectivity if other than 1.

Label	Intervention	Compliance	Trigger	Description
V	Vaccinate Adults Vaccinate Children Vaccinate Critical Workers	25% 60% 100%	Day0	Prevent 30% of treated individuals from becoming infected.
S	School Closure School Reopen	60%	1.0% of children diagnosed (by county) 0.1% of children diagnosed (by county)	Children stay home during school hours (with adult if under 13).
Q	Quarantine Critical Workers	100%	1.0% of adults diagnosed	Critical workers removed from general population and isolated in small groups.
I	Self Isolation	20%	2.5% of adults diagnosed	Eliminate all activities outside of the home.

Table 2.

Description of the experimental setup showing the sets of interventions, their compliance rates, and when the intervention takes effect. The full experimental design of 16 cells consists of every combination of the four interventions.

lion residents, 1.1 million locations, and 58 million activities of six types) is simulated. About 14,000 of the residents have been marked as critical workers based on the demographics of the actual critical worker subpopulation. A critical worker is someone whose work is essential for the health and welfare of the general population (for example, first responders, health-care workers, power-plant operators, and so on). The contagion that will be spread is the H5N1 influenza virus (that is, Avian flu). Figure 4 shows the PTTS associated with the H5N1 disease model.

There are several different interventions that may be applied to different groups at different times: vaccination of adults, children, and critical workers (V), school closure (S), quarantine of criti-

cal workers (Q), and self-isolation (I). The vaccine is assumed to be low efficacy, meaning that only 30 percent of the individuals vaccinated will be protected. This efficacy is representative of a vaccine for a newly emerging strain of influenza. When schools are closed, an adult caregiver is required to remain home with any children under 13 years of age. Sixty percent of the school children are considered to be compliant, meaning that they remain at home for the entire day. The other 40 percent participate in their normal after-school activities. Critical workers will be isolated in some type of group quarters in small groups of 16 where they do not come in contact with people outside of their subgroup. When an individual chooses to self-isolate, he stays home and does not participate in any

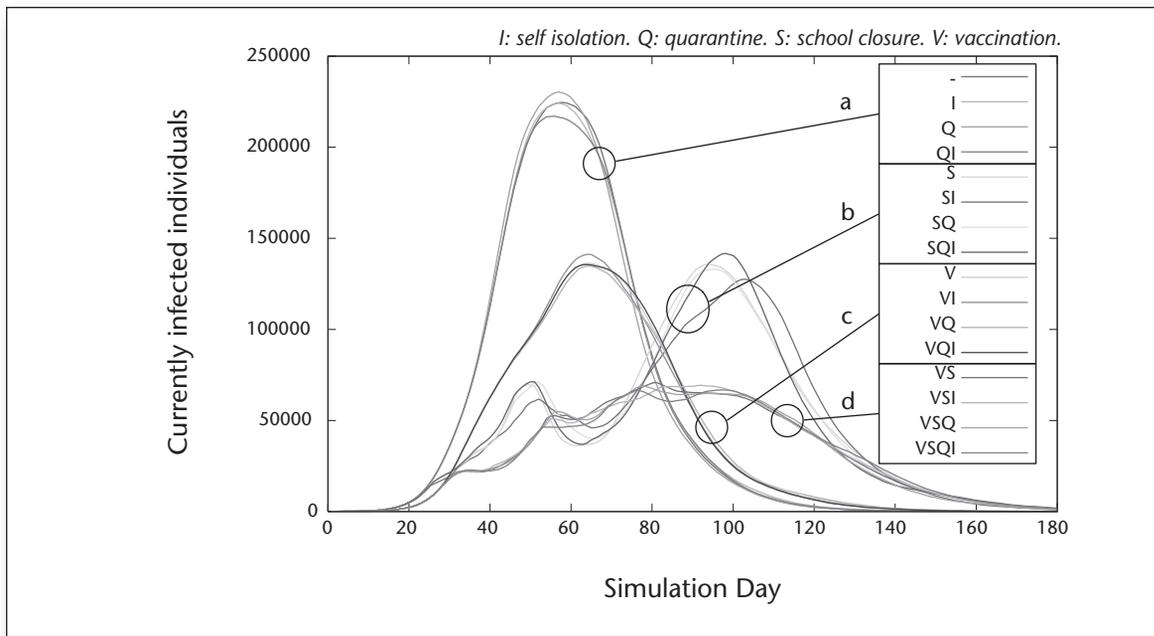


Figure 5.

Currently infected individuals by day for all combinations of interventions. The interventions can be divided into four groups based on the shape of the epidemic curves. Both school closure (b) and vaccination (c) are a significant improvement over doing neither (a), with school closure having the additional preparation and response benefit of delaying the peak by about a month. Combining school closure and vaccination (d) leads to further improvement.

activities away from home, although the person remains in contact with other people in the house. Table 2 describes the four sets of interventions.

The intrahost disease progression passes through several stages (National Academies Press 2006; Hal-loran et al. 2008). When a person is exposed to the disease, it starts off in a latent stage, where the individual is neither infectious nor symptomatic. This is potentially followed by an incubating stage during which the individual is partially infectious but still does not exhibit symptoms. In the final stage of the disease an individual is fully infectious and displays one of four levels of symptoms from asymptomatic to fully symptomatic. The probability of a person staying home instead of participating in his normal activities increases with the severity of the symptoms. Once the disease has run its course, the infected individual is considered recovered and cannot be reinfected.

Vaccination takes effect at the start of the simulation and the other interventions are triggered when a certain percentage of a subpopulation is diagnosed with the virus. It is assumed that 60 percent of those who are symptomatic will be diagnosed. School closure is done on a county by county basis, based on the number of sick children who reside in each county. Each person who enters one of the symptomatic  $X$  states has a probability of withdrawing to home, depending on the severity of the symptoms (symptom 1—20 percent, symptom 2—50 percent, symptom 3—95 percent).

Sixteen scenarios were simulated, specifying all combinations of the four sets of interventions ( $I$ ,  $Q$ ,  $S$ ,  $V$ ). Figure 4 shows the number of individuals that are infected for each cell. They can be grouped into four categories: those without vaccination or school closure (labeled a in the figure), vaccination (labeled b), school closure (labeled c), and both vaccinations and school closure (labeled d). Self-isolation and quarantine of critical workers have little impact on the overall infection rates. Self-isolation happens late in the epidemic (2.5 percent of the adult population is infected on a single day), limiting its effectiveness. In fact, when school closure or vaccination is included, the trigger level is never reached. Quarantine of critical workers affects such a small portion of the total population (about a third of a percent) that its effects are not apparent in the total population. However, quarantine reduces the percentage of critical workers who are infected from 40 percent without quarantine to 18 percent when critical workers are both quarantined and vaccinated, which may be vitally important for maintaining a functioning society.

Another interesting phenomenon happens with school closure. Schools are closed when 1 percent of school children are diagnosed as ill on a particular day in a county. The schools are reopened, on a county by county basis, when the number of diagnosed children falls below 0.1 percent. Even at that low level, there is enough residual infection to cause another wave of infections. This can be

observed in the dual peaks of the group of epicones labeled b in figure 5.

## Concluding Remarks

As our society becomes more connected, there will be an increasing need to develop innovative computational tools that will help policy makers and analysts grapple with complex questions. The advances in computing and information science on one hand will likely make our society even more connected and reduce the amount of time we have for decision making. On the other hand, the same technology will form the basis of new modeling and data processing environments that will generate new kinds of synthetic data sets that cannot be created in any other way (for example, direct measurement). This will enable social scientists to investigate entirely new research questions about functioning societal infrastructures and the individuals interacting with these systems. The tools will also allow policy makers, planners, and emergency responders unprecedented opportunities for multi-perspective reasoning leading to improved situation assessment and consequence management.

## Acknowledgments

We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSL); the work reported here is a joint effort of all the team members. We thank Dr. Karen Haigh and the anonymous reviewers for their valuable comments that greatly improved the presentation. This work is partially supported by NSF Nets Award CNS-0626964, NSF Award CNS-0845700, NSF HSD Award SES-0729441, CDC Center of Excellence in Public Health Informatics Award 2506055-01, NIH-NIGMS MIDAS project 5 U01 GM070694-05, DTRA CNIMS Grant HDTRA1-07-C-0113, and NSF NETS CNS-0831633.

## Notes

1. See also the 2010 IG24 workshop presentation "Financial Contagion: What Do We Mean? What Do We Know?" by M. Dungey and D. Thambakis ([www.g24.org/Dungey-Thambakis2003.pdf](http://www.g24.org/Dungey-Thambakis2003.pdf)).
2. See the 2003 Network Reliability and Interoperability Council report, Review of Power Blackout on Telecom, by P. Aduskevicz, K. Condello, and C. Burton.

## References

- Atkins, K.; Barrett, C.; Beckman, R.; Bisset, K.; Chen, J.; Eubank, S.; Feng, A.; Feng, X.; Harris, S.; Lewis, B.; Kumar, A.; Marathe, M.; Marathe, A.; Mortveit, H.; and Stretz, P. 2008. An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems. *CT Watch Quarterly* 4(1): 46–53.
- Atkins, K.; Barrett, C.; and Marathe, A. 2009. A Web Service-Based Artificial Market. Paper presented at the 2009 Winter Simulation Conference, Austin, TX, 13–16 December.
- Bailey, N. 1975. *The Mathematical Theory of Infectious Diseases and Its Applications*. New York: Hafner Press.
- Bailey-Kellogg, C.; Ramakrishnan, N.; and Marathe, M. 2006. Spatial Data Mining to Support Pandemic Preparedness. *SIGKDD Explorations* 8(1): 80–82.
- Barrett, C.; Beckman, R.; Berkgigler, K.; Bisset, K.; Bush, B.; Campbell, K.; Eubank, S.; Henson, K.; Hurford, J.; Kubicek, D.; Marathe, M.; Romero, P.; Smith, J.; Smith, L.; Speckman, P.; Stretz, P.; Thayer, G.; Eeckhout, E.; and Williams, M. 2001. Transims: Transportation Analysis Simulation System. Technical Report LA-UR-99-1658, Los Alamos National Laboratory Unclassified Report, Los Alamos, NM.
- Barrett, C.; Bisset, K.; Eubank, S.; Feng, X.; and Marathe, M. 2008. Episimdemics: An Efficient Algorithm for Simulating the Spread of Infectious Disease over Large Realistic Social Networks. In *Proceedings of 22nd ACM International Conference on Supercomputing*, 1–12. New York: Association for Computing Machinery.
- Barrett, C.; Bisset, K.; Leidig, J.; Marathe, A.; and Marathe, M. 2009a. Estimating the Impact of Public and Private Strategies for Controlling an Epidemic: A Multiagent Approach. In *Proceedings of the Twenty-First Conference on Innovative Applications of Artificial Intelligence*. Menlo Park, CA: AAAI Press.
- Barrett, C.; Beckman, R.; Khan, M.; Kumar, V.; Marathe, M.; Stretz, P.; Dutta, T.; and Lewis, B. 2009b. Generation and Analysis of Large Synthetic Social Contact Networks. Paper presented at the 2009 Winter Simulation Conference, Austin, TX, 13–16 December.
- Barrett, C.; Eubank, S.; and Marathe, M. 2006. Modeling and Simulation of Large Biological, Information and Socio-Technical Systems: An Interaction Based Approach. In *Interactive Computing: A New Paradigm*, ed. D. Goldin, S. Smolka, and P. Wegner, 353–394. Berlin: Springer Verlag.
- Beckman, R. J.; Baggerly, K. A.; and McKay, M. D. 1996. Creating Synthetic Base-Line Populations. *Transportation Research Part A—Policy and Practice* 30(6): 415–429.
- Bisset, K.; Chen, J.; Feng, X.; Vullikanti, A.; and Marathe, M. 2009a. Epifast: A Fast Algorithm for Large Scale Realistic Epidemic Simulations on Distributed Memory Systems. In *Proceedings of 23rd ACM International Conference on Supercomputing*. New York: Association for Computing Machinery.
- Bisset, K.; Feng, X.; Marathe, M.; and Yardi, S. 2009b. Modeling Interaction between Individuals, Social Networks and Public Policy to Support Public Health Epidemiology. Paper presented at the 2009 Winter Simulation Conference, Austin, TX, 13–16 December.
- Bowman, J.; Bradley, M.; Shiftan, M.; Lawton, K.; and Ben-Akiva, M. 1998. Demonstration of an Activity Based Model System in Portland. Paper presented at the 8th World Conference on Transportation Research, Antwerp, Belgium, 23–16 July.
- Carley, K.; Fridsma, D.; Casman, E.; Yahja, A.; Altman, N.; Chen, L.-C.; Kaminsky, B.; and Nave, D. 2006. Biowar: Scalable Agent-Based Model of Bioattacks. *IEEE Transactions on Systems, Man, and Cybernetics*, Part A 36(2): 252–265.
- Channakeshava, K.; Chafekar, D.; Bisset, K.; Vullikanti,

- A.; and Marathe, M. 2009. Epinet: A Simulation Framework to Study the Spread of Malware in Wireless Networks. In *Proceedings of the Second International Conference on Simulation Tools and Techniques for Communications, Networks, and Systems, SIMU-Tools09*. Rome: ICST Press.
- Desai, P. 2003. *Financial Crisis, Contagion, and Containment: From Asia to Argentina*. Princeton, NJ: Princeton University Press.
- Elveback, L.; Fox, J.; Ackerman, E.; Langworthy, A.; Boyd, M.; and Gatewood, L. 1976. An Influenza Simulation Model for Immunization Studies. *American Journal of Epidemiology* 103(2): 152–165.
- Eubank, S.; Guclu, H.; Kumar, V. A.; Marathe, M.; Srinivasan, A.; Toroczkai, Z.; and Wang, N. 2004. Modeling Disease Outbreaks in Realistic Urban Social Networks. *Nature* 429(6988): 180–184.
- Ferguson, N.; Cummings, D.; Cauchemez, S.; Fraser, C.; Riley, S.; Meeyai, A.; Lam-sirithaworn, S.; and Burke, D. 2005. Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia. *Nature* 437(7056): 2009–14.
- Ferguson, N.; Cummings, D.; Fraser, C.; Cajka, J. C.; Cooley, P. C.; and Burke, D. 2006. Strategies for Mitigating an Influenza Pandemic. *Nature* 442(7101): 448–452.
- Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: The MIT Press.
- Ford, K. 2008. Towards Cognitive Prostheses. AAAI Englemore Memorial Award Lecture presented at the Twenty-Third AAAI Conference on Artificial Intelligence. July 13–17, Chicago, IL.
- Germann, T.; Kadau, K.; Longini, I.; and Macken, C. 1983. Mitigation Strategies for Pandemic Influenza in the United States. In *Proceedings of the National Academy of Sciences*, 103(15): 5935–5940. Washington, DC: National Academy of Sciences.
- Halloran, M.; Longini, I.; Cowart, D.; and Nizam, A. 2002. Community Interventions and the Epidemic Prevention Potential. *Vaccine* 20(27–28): 3254–62.
- Halloran, M.; Ferguson, N.; Eubank, S.; Longini, I. M. Jr.; Cummings, D.; Lewis, B.; Xu, S.; Fraser, C.; Vullikanti, A.; Germann, T. C.; Wagener, D.; Beckman, R.; Kadau, K.; Barrett, C.; Macken, C.; Burke, D.; and Cooley, P. 2008. Modeling Targeted Layered Containment of an Influenza Pandemic in the United States. In *Proceedings of the National Academy of Sciences*, 105(12). Washington, DC: National Academy of Sciences.
- Hethcote, H. 2000. The Mathematics of Infectious Diseases. *SIAM Review* 42(4): 599–653.
- Kearns, M. 2007. *Graphical Games, Algorithmic Game Theory*, ed. N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, 159–180. New York: Cambridge University Press.
- Lauritzen, S., and Spiegelhalter, D. 1988. Local Computation with Probabilities on Graphical Structures and Their Application to Expert Systems. *Journal of the Royal Statistical Society B* 50: 157–224.
- Lauritzen, S. 1996. *Graphical Models*. New York: Oxford University Press.
- Longini, I.; Nizam, A.; Xu, S.; Ungchusak, K.; Hanshaworakul, W.; Cummings, D.; and Hollaran, E. 2005. Containing Pandemic Influenza at the Source. *Science* 309(5737):1083–1087.
- McIlwain, J. 2001. Population Coding: A Historical Sketch. In *Advances in Neural Population Coding*, 3–7. Amsterdam: Elsevier.
- Mundhenk, M.; Goldsmith, J.; Lusena, C.; and Allender, E. 2000. Complexity of Finite-Horizon Markov Decision Process Problems. *Journal of the ACM* 47(4): 681–720.
- National Academies Press 2006. Modeling Community Containment for Pandemic Influenza: A Letter Report. Washington, DC: National Academic Press.
- Newman, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review* 45(2): 167–256.
- Oreskes, N.; Shrader-Frechette, K.; and Belitz, K. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science* 263(5147): 641–646.
- Oreskes, N. 2000. Why Believe a Computer? Models, Measures, and Meaning in the Natural World. In *The Earth Around Us: Maintaining a Livable Planet*, ed. Jill S. Schneiderman, 70–82. San Francisco: W. H. Freeman and Co.
- Parker, J. 2007. A Flexible, Large-Scale, Distributed Agent-Based Epidemic Model. Paper presented at the 39th Conference on Winter Simulation, Washington, D.C., 9–12 December.
- Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kaufmann.
- Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.
- Pearl, J. 2009. Causality in the Social and Behavioral Sciences. Technical report, UCLA Cognitive Systems Laboratory (R-355). Los Angeles, CA.
- Russell, S., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.
- Shoham, Y., and Leyton-Brown, K. 2008. *Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations*. New York: Cambridge University Press.
- Wooldridge, M. 2002. *An Introduction to MultiAgent Systems*. New York: John Wiley and Sons Ltd.
- Chris Barrett** is the director of the Network Dynamics and Simulation Science Laboratory and professor in the Virginia Bioinformatics Institute and Department of Computer Science at Virginia Tech. His research interests are in the simulation of very large systems; theoretical foundations of simulation; interaction-based systems, computing, and dynamic systems; computational and systems biology; computational problems in epidemiology; cognitive science and computationally aided reasoning; computational economics; and infrastructure simulation.
- Keith Bisset** is a senior research associate in the Network Dynamics and Simulation Science Laboratory in the Virginia Bioinformatics Institute at Virginia Tech. His research is in parallel discrete event simulation, parallel and distributed architectures, object-oriented design, component frameworks, interaction-based computing, and grid computing.
- Jonathan Leidig** is a research assistant at the Network Dynamics and Simulation Science Laboratory in the Virginia Bioinformatics Institute at Virginia Tech and doctoral student in the Department of Computer Science at Virginia Tech. His research and dissertation involve digital libraries, information management, storage of simulation content, e-science, and information visualization.
- Achla Marathe** is the lead economist at the Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, and an associate professor at the Department of Agricultural and Applied Economics, Virginia Tech. Her research interests include modeling and simulation of socioeconomic systems; experimental and computational economics; data mining and knowledge discovery; and networked markets.
- Madhav Marathe** (corresponding author, mmarathe@vbi.vt.edu) is the deputy director of the Network Dynamics and Simulation Science Laboratory and professor in the Department of Computer Science, Virginia Tech. His research is in interaction-based modeling and simulation of large, complex biological, information, social, and technical networks; design and analysis of algorithms and computational complexity; network science; communication networks; high-performance and grid computing systems; and computational epidemiology.