# HIGH PERFORMANCE INFORMATICS FOR PANDEMIC PREPAREDNESS

Keith R. Bisset
Stephen Eubank
Madhav V. Marathe

Network Dynamics & Simulation Science Laboratory, Virginia Bioinformatics Institute
Department of Computer Science
Virginia Tech
Blacksburg, Virginia 24061, USA

## ABSTRACT

Pandemics such as H1N1 influenza are global outbreaks of infectious disease. Human behavior, social contact networks, and pandemics are closely intertwined. The ordinary behavior and daily activities of individuals create varied and dense social interactions that are characteristic of modern urban societies. They provide a perfect fabric for rapid, uncontrolled disease propagation. Individuals' changing behaviors in response to public policies and their evolving perception of how an infectious disease outbreak is unfolding can dramatically alter normal social interactions. Effective planning and response strategies must take these complicated interactions into account. Recent quantitative changes in high performance computing and networking have created new opportunities for collecting, integrating, analyzing and accessing information related to such large social contact networks and epidemic outbreaks. The paper will describe our efforts to build a Cyber Infrastructure for EPIdemics (CIEPI) – a high performance computing oriented decision-support environment to support planning and response in the event of epidemics.

## 1 INTRODUCTION

A pandemic is an outbreak of infectious disease affecting large populations simultaneously across the entire globe. For example, the recent H1N1 influenza pandemic spread to most parts of the world in a few weeks. Pandemics pose disease control challenges unmatched by any other infectious disease outbreaks.

Certain modern trends exacerbate the speed and severity of pandemics. First, the global population is larger and increasingly urbanized, providing greater opportunity for viruses to be transmitted within populations. Second, levels of international travel are much greater than in the past, allowing viruses to spread globally more quickly. Third, populations in many countries are increasingly elderly and immuno-compromised, thereby increasing the chances that infection will lead to more severe illness and death. It is important and worthwhile to prepare for these events. Indeed, over the last 50 years, public health agencies across the world have made remarkable strides in reducing the social, economic and health impacts of such pandemics.

Advances in computational, biological and information sciences have the potential to dramatically improve how public health authorities across the world plan for, detect, and respond to pandemics. New surveillance tools are being developed so that pandemics can be detected at onset. These methods include sentinel clinics and serological sampling as well as social media based data analytics. Similarly, new high performance computing models are being created to simulate the spread of a disease at regional, national and international scales. This scalability and ability to represent region specific detail permits cost-benefit analysis of various intervention strategies as well as improved prediction. Furthermore, recent models have begun incorporating detailed behavioral representation. Finally, informatic tools are being developed to support public health analysts carrying out detailed what-if analysis using these sophisticated tools. There

are two key advances: (i) the ability to develop simple and intuitive user interfaces so that the end user can interact with the sophisticated modeling environments without becoming a computer scientist; (ii) ubiquitous web access to these tools, allowing the analyst to carry out investigations at *any time, on any device and at any location*.

## 2 PERVASIVE INFORMATICS ENVIRONMENT

Current-generation High Performance Computing (HPC) based modeling environments for complex co-evolving networks use large-scale special-purpose systems. These systems are readily accessible only to a limited number of highly specialized technical personnel in narrowly defined software applications. In other words, web-based distributed information CyberInfrastructure (CI) for representing and analyzing large socio-technical systems simply do not currently exist. Developing such an infrastructure poses unique challenges, different, for example, from those faced by researchers developing the CI for physical systems:

**Scalability** The CI must be globally scalable. The scalability comes in three forms: (i) allowing multiple concurrent users; (ii) processing huge quantities of distributed data; and (iii) executing large, national-scale models.

**Coordination** The CI should allow computational steering of experiments. The systems needed by stakeholders are geographically distributed, controlled by multiple independent, sometimes competing, organizations and are occasionally dynamically assembled for a short period of time.

**Data and Information Processing** The CI should facilitate efficient data and information fusion and analysis. The Internet has enabled data sharing in a simple and cost effective way, from the producers' side. Consumers of the data must still locate the appropriate data and deal with multiple incompatible data formats.

**User Support** The CI should provide an appropriate analysis framework for users, including user interfaces, high level formalisms to set up experiments, and visual and data analytics that include methods for integrating heterogeneous databases to support multi-view visualization (e.g. disease spread in a geographic region and epidemic curves).
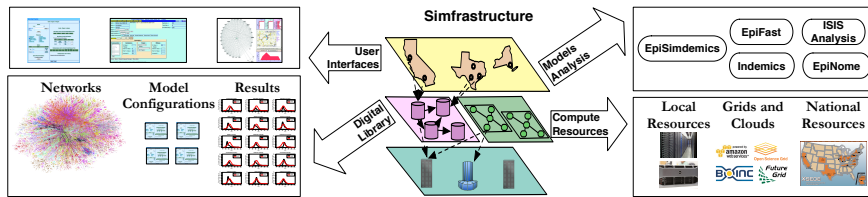


Figure 1: High level view of CIEPI. The geographically distributed middleware system is shown in the center, connecting the four main components of the system. Each of the individual resources may be geographically and organizationally distributed.

We will describe *Cyber Infrastructure for EPIdemics (CIEPI)*, a pervasive computing environment for global modeling, information and decision support, that will provide scientists, educators and public health analysts ways to detect, plan for and respond to pandemics (see Section 5 for recent use of CIEPI). The goal of the CIEPI is to make HPC resources seamless, invisible, and indispensable in routine analytical efforts by demonstrating that HPC resources should be organized as an evolving commodity, and made accessible in a fashion as ubiquitous as Google's home page. CIEPI was motivated by the need to provide an integrated platform to support epidemiological science and public health analysis.

CIEPI is comprised of: (i) high-resolution scalable models of disease transmission and behavioral adaptation; (ii) a service-oriented architecture and delivery mechanism for facilitating the use of these models by domain experts; (iii) a distributed coordinating architecture for information fusion, model execution and data processing; (iv) a scalable data management architecture to support model execution

and analytics; and (v) scalable methods for visual and data analytics to support analysts. See (Barrett et al. 2010; Barrett et al. 2006) for further details.

## 3 DATA SYNTHESIS AND MODELING

Traditionally mathematical and computational modeling of epidemics has focused on coupled rate equations – differential equation models for completely mixing populations, in which large groups of people interact with each other; see (Bailey 1975; Hethcote 2000; Kaplan et al. 2002; Newman 2003) for comprehensive reviews of this approach. Over the years, these models have proved to be immensely successful. They yield analytical expressions for a number of important epidemic parameters, including number infected, deaths, etc. But they can not capture the complexity of human behavior and interactions that serve as a mechanism for disease transmission.

An important factor influencing the progress of an infectious disease outbreak is the structure of the contact network across which it spreads. This structure changes with time, both endogenously and in response to disease prevalence and public health interventions. The resulting phenomena are irreducibly complex, i.e., the simplifying assumptions required by aggregate models inevitably render them incapable of reproducing important phenomena. Over the last few years, researchers have begun to investigate a new approach to epidemiological modeling that can naturally handle this co-evolution. It is based on a combination of network theory and individual based simulation (Halloran et al. 2002; Carley et al. 2006; Eubank et al. 2004; Ferguson et al. 2006; Ferguson et al. 2004; Ferguson et al. 2005; Longini et al. 2005; Germann et al. 2006; Meyers et al. 2003).

"Interaction-based" or "network-based" epidemiology, as this approach has come to be known, addresses some of the challenges faced by aggregate models. It seeks to understand the complicated interplay between the three components that contribute to irreducible complexity: (i) individual behaviors of agents (people and infrastructure components), (ii) unstructured, heterogeneous multi-scale socio-technical networks, and (iii) the dynamical processes on these networks.

Although computationally expensive and data intensive, network-based epidemiology fundamentally alters the types of questions that can be posed, providing qualitatively different insights into disease dynamics and public health policies. More importantly, the modeling approach allows policy makers to formulate and investigate potentially novel and context specific interventions. A good example of this is the school closure policy adopted during the recent H1N1 influenza outbreak http://www.time.com/time/health/article/0,8599,1915244,00.html and http://www.sciencedaily.com/releases/2012/02/120206174205.htm.

### 3.1 Network Synthesis

We describe first-principle based methods for generating synthetic regional and national scale social contact networks. Unlike simple random graph techniques, these methods use real world data sources and combine them with behavioral and social theories to synthesize networks. The work began almost 20 years ago as a part of the TRANSIMS project we developed with colleagues at the Los Alamos National Laboratory (Barrett et al. 2001; Beckman et al. 1996) and has continued with new methods introduced under the Simfrastructure project (Barrett et al. 2009).

Figure 2 illustrates our approach to synthesizing urban scale social contact networks, as discussed in detail in (Eubank et al. 2004; Barrett et al. 2009). First, we create a synthetic population by integrating a variety of databases from commercial and public sources. The process preserves the confidentiality of the original data sets, yet produces realistic attributes and demographics for the synthetic individuals. This process is not simply a de-identification or anonymization of an existing population, but the creation of an entirely new population with the same statistical properties as the original population. A census of the two populations is statistically identical at a slightly aggregated level (the US Census block group, typically 600-3000 people). The process has successfully been employed in other countries such as India and China. The synthetic population is a set of synthetic people and households, geographically located,
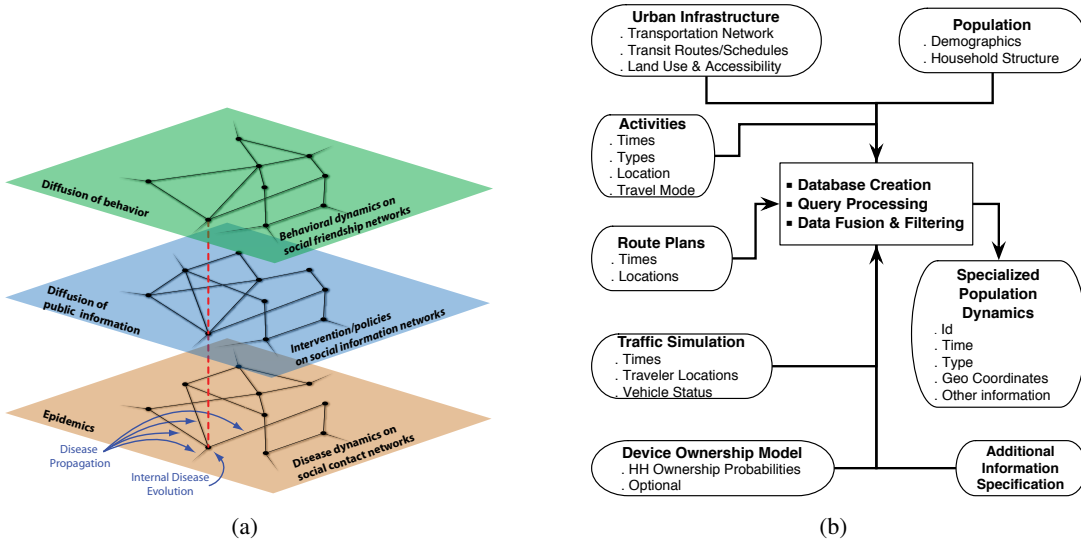
Figure 2: (a) Coupled social and information networks: the social contact graph (bottom layer), public information infrastructure network (middle layer) and friendship network (top layer). (b) Schematic diagram describing various data sources that we have integrated to produce synthetic relational networks.

each associated with demographic variables drawn from any of the demographics available in the input dataset. Each synthetic individual is created in the context of a household with other synthetic individuals and each household is located geographically.

Next, a set of activity templates for households are constructed, based on several thousand responses to an activity or time-use survey. These activity templates include the sort of activities each household member performs and the time of day they are performed. Thus, across an entire geographical region, demographic information for each person and location, together with a minute-by-minute schedule of each person's activities and the locations where these activities take place is generated using a combination of simulation and data fusion techniques. This information can be captured in the form of a *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 $p \in P$ 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 times. Note that it is *impossible* to build such a network for any region large enough to be epidemiologically significant *solely* by collecting field data, although such data can be incorporated into the synthetic population creation process. The use of generative models to build such networks is a unique feature of this work.

How do these networks compare with classical random graphs? Our results show that the resulting realistic social contact networks: (i) differ structurally from synthetic networks generated using simple random processes (Eubank et al. 2006; Barrett et al. 2009), and (ii) show interesting similarities as well as differences among urban regions (Barrett et al. 2009; Barrett et al. 2007).

## 3.2 Behavioral Adaptation and Co-evolving Networks

During a pandemic, a public health policy-maker's primary goal is to control the spread of infectious disease through the application of interventions, guided by public policy. Among other things, interventions induce behavioral changes in individuals at the same time as individuals independently change their own behavior in response to their perception of how the disease is evolving. These changes in turn produce a constantly changing social network, which thus co-evolves with individual behavioral changes and disease dynamics. In other words, individual behaviors, disease dynamics and the social contact networks that they generate interact and co-evolve as individuals try to avoid infection in the context of public health interventions.
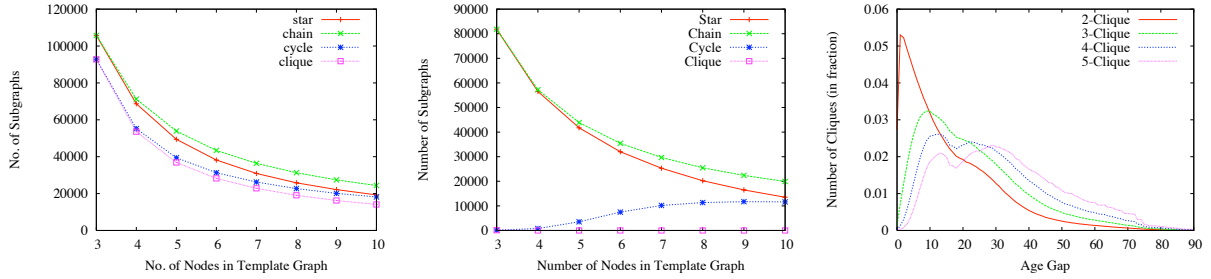
Figure 3: Counts of various kinds of subgraphs in our synthetic Portland social contact network (1.5 million nodes and 110 million edges) and a randomized version with the same degree distribution: (a) Number of non-overlapping subgraphs: star, cycle, chain, and clique, (b) Number of non-overlapping subgraphs after shuffling the edges (keeping the degree distribution invariant), and (c) number of cliques as a function of age gap – the age difference between oldest and youngest people in the clique.

In health sciences, researchers have developed what one refers to as verbal social and behavioral theories and discussed the role of these theories in understanding behaviors and its relationship with diseases and maintaining a healthy life. This includes the *Health Belief Model (HBM)* proposed by Hochbaum, Rosenstock and Kegels, models of behavior as it pertains to human health, The Social Cognitive Theory (SCT) proposed by Bandura and Rotter, and the Social Ecological Model (SET) proposed by McLeroy, Bibeau, Steckler, Glanz – that emphasizes multiple levels of influence (individual, interpersonal, organizational, community, and public policy). Developing computational models of individual and collective behavior requires these *verbal theories* to be recast in an algorithmic fashion. We argue that just like biologist have recently argued the concept of *executable or procedural biology*, there is a need to develop procedural models of individual, collective and institutional behaviors. These procedural social theories can prove to be useful in multiple ways. First, as in other sciences, they make the formal specification of the underlying process explicit removing ambiguity and thus improving our ability to compare and contrast such theories. Second, they provide an information-theoretic/algorithmic representation of social systems. We have developed Indemics (see Section 3.3), a computational modeling environment for representing and analyzing complex behaviors, and its interplay with epidemics and social networks. We use a mutli-theory, multi-network representation of individual, collective and institutional behaviors, associated social networks and disease dynamics. The basic framework was outlined in Monge and Contractor in (Monge and Contractor 2003).

### 3.3 High Performance Computer Models for Disease Modeling

Representing and analyzing disease dynamics over large unstructured and time-varying social contact networks requires new work in high performance computing as well. In an effort to address the scaling problem typical in algorithms implemented for high performance computing, our group has developed three different parallel algorithms and their implementations over the last 10 years: EpiSimdemics, EpiFast, and Indemics. These differ in the tradeoff they provide between computation speed, model realism and sophistication, and ease of introducing new behavior and interventions. All three can be executed on traditional distributed memory clusters of varying sizes. EpiSimdemics, and its predecessor EpiSims (Eubank et al. 2004), is quite general and flexible, at the cost of decreased computational speed. EpiFast can model a restricted class of diseases and interventions, but more quickly. Indemics is a extension to EpiFast that integrates database computing to enable the most general set of interventions that are the easiest to create, but at the loss of computational speed. A comparison between the simulations can be found in (Bisset et al. 2011).

**EpiSimdemics** (Barrett et al. 2008) is an interaction-based, highly resolved modeling and simulation system for representing and reasoning about contagion diffusion across large ($3x10^8$ nodes) networks. It was

originally intended to model the spread of infectious disease across social contact networks, but has also been adapted for modeling the spread of worms between bluetooth equipped wireless devices (Channakeshava et al. 2011), and the movement and interaction of cells in the human immune system (Bisset et al. 2012; Wendelsdorf et al. 2011). The EpiSimdemics algorithm is based on contagion diffusion across a social network. The network is represented as a bipartite graph, with people and locations as the nodes, and edges between them representing a person's presence at a location. We refer to this as a person-location graph. This representation of the network is about 50 times sparser than a person-person contact network where each edge represents two people in contact for some time. It also allows location based interventions such as school closure to remove some contacts (e.g., contacts at a specific school) and replace them with contacts at a replacement location (e.g., home). EpiSimdemics can represent virtually all the existing models of between-host disease propagation. It supports fully dynamic social networks (nodes and edges can be added and removed in response to disease propagation). It also has the ability to represent a large collection of behavioral specifications, including the use of pharmaceuticals (e.g., vaccines or anti-virals) and change of planned activities (e.g., school closure, quarantine). Contagion and behavior are modeled as coupled **probabilistic timed transition systems** (PTTS). Written in C++, EpiSimdemics initially used MPI but has recently been modified to use the Charm++ parallel framework (Kalé and Krishnan 1993), which has led to improved scaling and execution time (Bisset et al. 2012). It can simulate the entire population of the United States, a person-location network with 300 million people and 1.5 billion edges. On a 768-core cluster, 120 days of an epidemic on this scale can be simulated in under 15 minutes.

**EpiFast** (Bisset et al. 2009) differs from EpiSimdemics in the following ways: (i) The underlying person-person social network $G$ is explicitly given, whereas EpiSimdemics builds it implicitly; (ii) EpiFast implements a simple Susceptible Exposed Infectious Recovered (SEIR) disease model for epidemic simulations; (iii) EpiFast runs in bulk-synchronous mode; (iv) Interventions in EpiFast are coded as structural changes in $G$ (e.g. non-pharmaceutical interventions) or nodes' properties (e.g. infectivities and vulnerabilities). EpiFast can handle a class of interventions of the form: *when predefined global conditions $g_1, g_2, \ldots$ and predefined local conditions $\ell_1, \ell_2, \ldots$ are satisfied for node v, change v's properties and/or labels of a predefined subset of edges incident on v.*

EpiFast is implemented in C++, using MPI for communication. It is similar to EpiSimdemics in that it can run on any distributed memory system, so long as the total of all available memory can hold the whole network and modeling parameters. The computing architecture comprises of one master processing element (PE) and multiple worker PEs. The system runs in a bulk synchronous parallel mode. The master PE passes messages and data and for coordinating synchronizations. The network is partitioned online and stored on worker PE such that each node has a unique "owner" PE and all edges incident on this node "reside" together on the same PE. A worker PE can distinguish "local" nodes and "foreign" nodes and the master PE knows the "owner" PE of every node. This way, all data about foreign nodes are relayed by the master PE. For the problems we have considered to date, the master PE does not appear to be a bottleneck.

In each time step of a simulation, computations consist of either applying a set of interventions or moving time forward and computing the disease dynamics. In case of interventions, the master PE and worker PE collaborate to evaluate global conditions. Then each worker PE evaluates local conditions for each node. Corresponding changes are applied to nodes that satisfy all conditions and to edges incident on them. In order to compute disease dynamics, each worker PE identifies its infectious nodes. For each such node, a PE computes adjacent nodes that it infects and sends messages about infected foreign nodes to their owner PE via the master PE. This information is communicated to the master PE that updates the state of each node based on the information received from the worker PEs. This information is then relayed back to the appropriate worker PEs completing one global step.

**Indemics** (Interactive Epidemic Simulation) is a data intensive, interactive, high performance modeling environment for pandemic planning, situation assessment, and course-of-action analysis. Indemics supports online interaction between users and the computational model. The user can stop the simulation at a specific time point or specify a stopping criteria. When the simulation stops, a user can assess the epidemic state

by appropriate queries. A typical query captures a set of individuals with specified demographic attributes and disease state within a given spatial region. Based on this, the user can then formulate intervention strategies and continue to move the simulation forward. Simulations can also be rewinded back in time to a specified time point. The start-stop and rewind features for HPC-simulations require that we hold appropriate information about the system state in memory. Another unique feature of Indemics is the use of databases and relational query languages to specify interventions and state assessment. In Indemics the disease propagation is separated from state assessment and intervention computation. This separation facilitates scaling, use of databases to actively guide the disease dynamics and improves human productivity by allowing analysts to use powerful query languages to specify interventions.

## 4    ANALYTICS AND SCALABLE MIDDLEWARE

The epidemiological modeling tools described above are capable of providing very detailed information on spatio-temporal disease dynamics. The size and scale of the data and the expertise required to use the simulations demand a user friendly environment that provides an easy way to set up experiments and analyze the results.

We have built a tool called the *Interface to Synthetic Information Systems (ISIS)* that achieves the first goal. ISIS allows a user to set up detailed factorial experiments. Using a simple interface to an underlying digital library, a user can choose from among many pre-constructed instances: (i) a social contact

We have built a tool called the *Interface to Synthetic Information Systems (ISIS)* that achieves the first goal. ISIS allows a user to set up detailed factorial experiments. Using a simple interface to an underlying digital library, a user can choose from among many preconstructed instances: (i) a social contact network; (ii) a within-host disease progression model; and (iii) a set of interventions. Each intervention requires additional details such as compliance level, sub-populations to which the interventions are applied and intervention triggers. An experiment consists of sweeping one or more parameters across a user-specified range of values. After setting up the experiment, the user is provided access to the results of the simulations. A set of basic analyses are performed automatically and the results



Figure 4: ISIS User interface including Indemics.

are displayed. See Figures 4 and 5 for an illustration of ISIS. The standard plots and epidemic curves provide very detailed information about the epidemic. Additional information such as the spatio-temporal dynamics and disease dendrogram (how the disease moved over the social network) is also available. A key aspect of ISIS is its simplicity – we can train public health analysts to make effective use of the system in about three hours. Paradoxically, by hiding the computational models from the end user, ISIS makes them much more accessible.

Additional analysis is also possible. Our colleagues at the University of Utah have partnered with us to develop an elegant visual analytics tool for epidemiological simulation. *Epinome*, shown in Figure 6, is an integrated dynamic visual analytic workbench for epidemic investigation in which a user can explore, compare, and analyze geo-temporal data from simulated infectious disease outbreaks (Livnat et al. 2010;
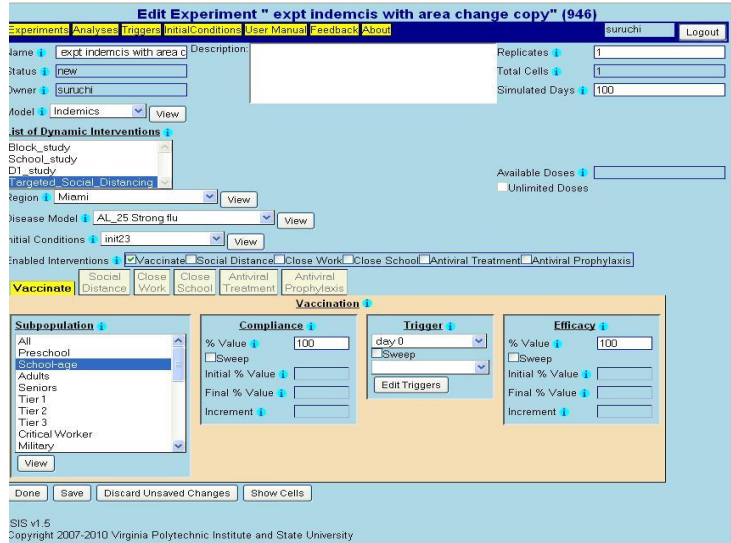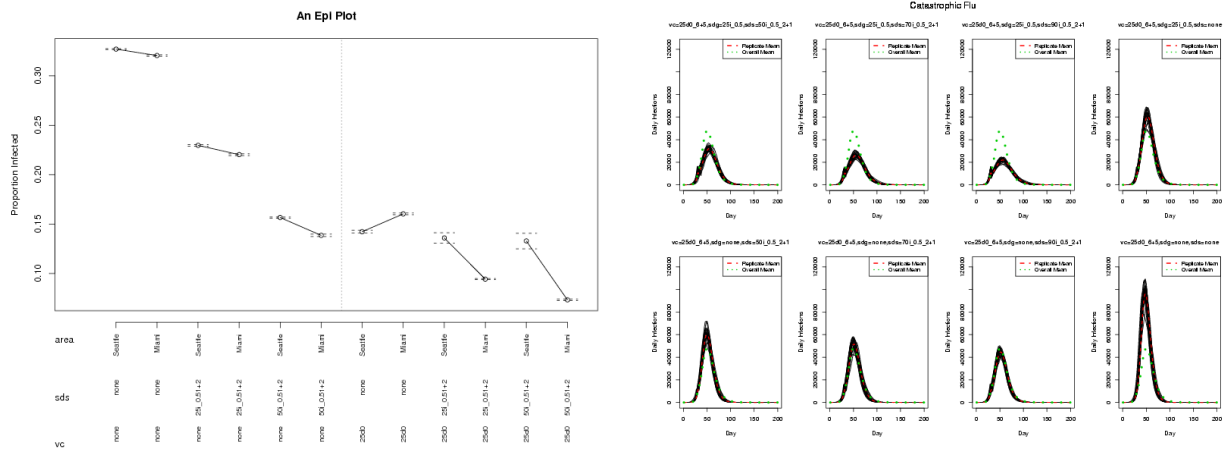
Figure 5: Example standard plot (left) and simulated epidemic curves (right). The standard plots can capture interaction effects as well display ranking of interventions.

Livnat et al. 2012). The design of this tool focuses on the interaction between public health practitioners and the surveillance system and on the human factors affecting information comprehension and decision-making. Epinome features a dynamic display in which various data views can be laid out dynamically to form a coordinated multiple views environment (Livnat et al. 2012). As tasks and focus change, a user can drag additional tools into the workspace, rearrange them interactively, or remove them all together. Parallel workspaces are also supported, each with its own collection of views centering on different subsets of the data. The use of multiple workspaces allows users to explore multiple simulation runs in parallel. Filtering is achieved through an intuitive filter-by-example paradigm in which a user directly drags a data value (either text or graphics) from one view to another to create and apply a data filter on the receiving view.

Simfrastructure is a service oriented middleware system (see Figure 1) that serves as a multiplexer for four types of components in CIEPI: user interfaces, digital libraries, models and simulations, and high performance computing resources. It is designed to allow any component to easily access the services provided by any other component without needing to know about the existence of any particular service provider, through the use of an asynchronous *Message Board*. The Simfrastructure *Resource Broker* matches service requests to appropriate service providers and computing resources, taking into account required timeliness and accuracy, as well as security, cost and other factors.

## 5    CASE STUDIES AND APPLICATIONS TO PUBLIC POLICY

*Using Simdemics for Supporting Federal Pandemic Influenza Planning.* In 2006, at the request of federal agencies, as members of the NIH Models of Infectious Disease Agent Study (MIDAS) project, we participated in a study to support federal pandemic plans. Three teams in the MIDAS program analyzed combinations of strategies for responding to influenza. These strategies included both pharmaceutical and non-pharmaceutical interventions (NPI) targeted at those parts of the population where they might most effectively control the spread of disease. Examples of NPI include: closing schools or reducing contacts at work and in the community. The goal of an NPI is to change human social behaviors so as to mitigate an outbreak. The MIDAS groups modeled the effectiveness of various combinations of these measures, implemented at various times during the outbreak, and with varying compliance rates.

Results of the MIDAS analyses were reviewed in a Letter Report by the Institute of Medicine, Modeling Community Containment for Pandemic Influenza (Committee on Modeling Community Containment for Pandemic Influenza and Institute of Medicine 2006) and have been subsequently published in (Halloran et al.

2008). The report concludes, among other things, that the evidence suggests a role for community restrictions in reducing pandemic influenza virus transmission. The evidence does not allow for differentiating possible effects of specific types of community restrictions, nor does it allow differentiation between voluntary versus mandatory community restrictions. The committee further identified "intervention fatigue as a factor that could undermine the potential effectiveness of any of these community restriction measures." Note that community restrictions (voluntary and mandatory) involve a behavioral change that in turn implies structural changes in the underlying social contact network. Network-based epidemiological modeling played an important role in the case study.
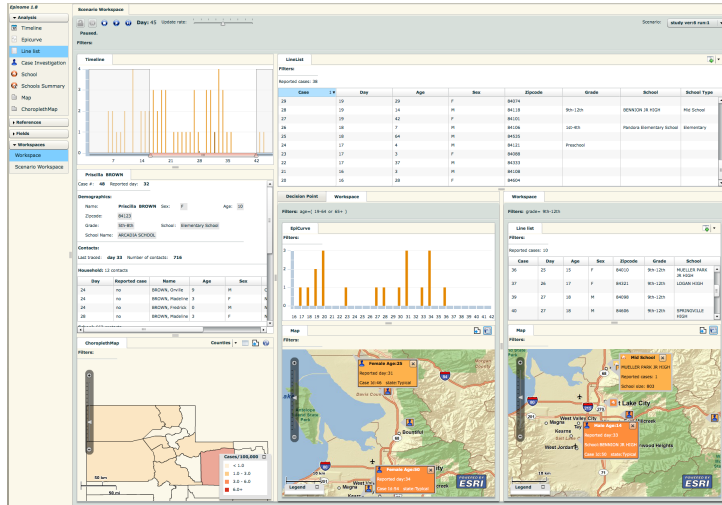


Figure 6: Screen shot of the Epinome tool.

*Using Simdemics for Pandemic Planning: Optimal Resource Allocation.* In a recent NIH sponsored study during the summer of 2008, we modeled different anti-viral drug distribution strategies to mitigate pandemic flu. Several strategies were considered, including prophylactic home stockpiling, public distribution, market based distribution, and income based distribution. The results showed that the market based allocation plan was less effective in reducing incidence and the epidemic duration than the public allocation (treatment based) strategy. This difference was explained by the fact that the market (i) under-provides public goods that have positive social externalities; (ii) often leads to inequitable deployment; (iii) favors the affluent and reinforces distrust among the non-affluent; and (iv) leads, via income stratification and assortative mixing, to spatial clustering of the antiviral recipients. It was clear that market based allocation would provide the medkits to those who do not necessarily need them and hoarding would cause the price to rise enough to make it unaffordable to the really sick and poor. Defensive behavior from the affluent would concentrate the medkits in small spatial clusters which in turn would allow the epidemic to spread in less affluent areas. More details of this study can be found in (Chen et al. 2010).

*Federal Interagency Support during the Emergence of H1N1 Influenza during April and May, 2009.* This effort was in direct response to the initial reports of the emergence of the H1N1 Influenza virus that eventually caused a global pandemic. At first, infections were confirmed in Mexico, California, and Texas and, a few days later, New York. The rapid spread combined with initial overestimates of its mortality rate raised serious concerns of a repeat of the 1918 influenza pandemic.

Initial reports about the disease characteristics were unreliable, with wide variations placed on important parameters like proportion of symptomatic individuals and duration of the infectious period. Having developed a web-based front-end to our CIEPI, called DIDACTIC, for just this purpose, we were able to quickly run a series of studies to learn the likely impact of variations in these parameters on the US population. A report was quickly drafted about the impact of disease characteristics on the size and shape of the expected epidemic curve. Several variants of disease models were added to the DIDACTIC tool.

As H1N1 Influenza continued to spread in the US, the Department of Health and Human Services (DHHS) teamed up with the Defense Threat Reduction Agency (DTRA) to place the DIDACTIC tool in the hands of US government analysts to provide day-to-day modeling results. At a daily interagency meeting at DHHS, led by the Office of the Assistant Secretary for Preparedness and Response (ASPR) and

the Biomedical Advanced Research and Development Authority (BARDA), analysts and policy-makers discussed overnight projections of the pandemic and possible interventions. These projections were generated using DIDACTIC as one of just two tools that were available and capable of providing detailed forecasts and analyses at these timescales. This integration inside the 24-hour decision cycle running the federal government's response to this emerging crisis would not have been possible without the development of CIEPI. The analysts at DHHS were able to perform course-of-action analyses to estimate the impact of closing schools and shutting down workplaces.

DIDACTIC, in conjunction with other modeling tools, was used to study a number of important policy questions, including: (i) the role of anti-virals as prophylactic drugs to reduce the severity of the pandemic and delay its peak; (ii) the effect of school closures in slowing the spread; and (iii) the role of travel restrictions, especially international travel restrictions. Officials were able to use modeling tools such as DIDACTIC themselves to evaluate these scenarios. This experience demonstrates the importance and feasibility of placing sophisticated modeling tools in the hands of public health decision makers and highlights the role that highly detailed modeling can play during a response to an emerging crisis.

## 6 CONCLUDING REMARKS

We have described CIEPI, an integrated cyberinfrastructure that allows public health analysts unprecedented access to high performance computational models and analytics to support policy making. CIEPI has been used for planning as well as during the response phase of recent epidemic outbreaks. A number of issues were not discussed due to space limitations; we mention a few important ones here. First, the modeling methodology described here can in fact be used to support event forecasting. Second, we did not discuss a number of interesting computational issues surrounding control and optimization problems arising in networked epidemiology; see (Eubank et al. 2006). Finally, modeling environments such as these are applicable to understanding zoonotic diseases (Alexander et al. 2013) as well as vector borne diseases such as Cholera and Dengue.

## 7 ACKNOWLEDGMENTS

## REFERENCES

Alexander, K., B. Lewis, M. Marathe, S. Eubank, and J. Blackburn. 2013. "Modeling of Wildlife Associated Zoonoses: Applications and Caveats". *to appear in Vector-Borne and Zoonotic Diseases.*

Bailey, N. 1975. *The Mathematical Theory of Infectious Diseases and Its Applications*. Hafner Press, New York.

Barrett, C. L., R. J. Beckman, K. P. Berkbigler, K. R. Bisset, B. W. Bush, K. Campbell, S. Eubank, K. M. Henson, J. M. Hur-ford, D. A. Kubicek, M. V. Marathe, P. R. Romero, J. P. Smith, L. L. Smith, P. L. Speckman, P. E. Stretz, G. L. Thayer, E. V. Eeckhout, , and M. Williams. 2001. "TRANSIMS: Transportation Analysis Simulation System.". Unclassified Technical Report No. Los Alamos National Laboratory LA-UR-00-1725, Los Alamos National Laboratory.

Barrett, C. L., R. J. Beckman, M. Khan, V. A. Kumar, M. V. Marathe, P. E. Stretz, T. Dutta, and B. Lewis. 2009, December. "Generation and Analysis of Large Synthetic Social Contact Networks". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin,

and R. G. Ingalls, 1003–1014. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barrett, C. L., K. Bisset, S. Eubank, M. V. Marathe, V. A. Kumar, and H. Mortveit. 2007. *Modeling and Simulation of Biological Networks*, Chapter Modeling and Simulation of Large Biological, Information and Socio-Technical Systems: An Interaction Based Approach, 101–147. AMS.

Barrett, C. L., K. R. Bisset, S. Eubank, X. Feng, and M. V. Marathe. 2008. "EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks". In *Proceedings of the ACM/IEEE Conference on High Performance Computing (SC)*, 37. IEEE Press.

Barrett, C. L., K. R. Bisset, J. Leidig, A. Marathe, and M. V. Marathe. 2010. "An Integrated Modeling Environment to Study the Co-evolution of Networks, Individual Behavior and Epidemics". *AI Magazine* 31 (1): 75–87.

Barrett, C. L., S. Eubank, and M. V. Marathe. 2006. *Interactive Computation: The New Paradigm*, Chapter An Interaction Based Approach to Modeling and Simulation of Large Biological, Information and Socio-Technical Systems, 353–392. Springer.

Beckman, R., K. Baggerly, and M. McKay.. 1996. "Creating base-line populations.". *Transportation Research Part A- Policy and Practice* 30:415–429.

Bisset, K., J. Chen, X. Feng, A. Vullikanti, and M. Marathe. 2009, June. "EpiFast: A fast algorithm for large scale realistic epidemic simulations on distributed memory systems". In *Proceedings of 23rd ACM International Conference on Supercomputing (ICS'09)*. ACM Press. NYC, NY.

Bisset, K., J. Chen, C. J. Kuhlman, V. S. A. Kumar, and M. V. Marathe. 2011. "Interaction-based HPC modeling of social, biological, and economic contagions over large networks". 2933–2947: IEEE.

Bisset, K. R., A. M. Aji, E. Bohm, L. V. Kale, T. Kamal, M. V. Marathe, and J.-S. Yeom. 2012, May 21. "Simulating the Spread of Infectious Disease over Large Realistic Social Networks using Charm++". In *17TH International Workshop on High-Level Parallel Programming Models and Supportive Environments*.

Bisset, K. R., M. M. Alam, J. Bassaganya-Riera, A. Carbo, S. Eubank, R. Hontecillas, S. Hoops, Y. Mei, K. Wendelsdorf, D. Xie, J.-S. Yeom, and M. Marathe. 2012, May 21–25. "High-Performance Interaction-Based Simulation of Gut Immunopathologies with ENteric Immunity SImulator (ENISI)". In *26th IEEE International Parallel & Distributed Processing Symposium (IPDPS)*. To Appear.

Carley, K. M., D. B. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, B. Kaminsky, and D. Nave. 2006. "BioWar: Scalable Agent-Based Model of Bioattacks". *IEEE Transactions on Systems, Man, and Cybernetics, Part A* 36 (2): 252–265.

Channakeshava, K., K. Bisset, M. Marathe, A. Vullikanti, and S. Yardi. 2011. "High performance scalable and expressive modeling environment to study mobile malware in large dynamic networks". In *Proceedings of 25th IEEE International Parallel & Distributed Processing Symposium*.

Chen, J., A. Marathe, and M. V. Marathe. 2010. "Coevolution of Epidemics, Social Networks, and Individual Behavior: A Case Study". In *SBP*, edited by S.-K. Chai, J. J. Salerno, and P. L. Mabry, Volume 6007 of *Lecture Notes in Computer Science*, 218–227: Springer.

Committee on Modeling Community Containment for Pandemic Influenza and Institute of Medicine 2006. *Modeling Community Containment for Pandemic Influenza: A Letter Report*. Washington D.C.: The National Academies Press.

Eubank, S., H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. "Modeling Disease Outbreaks in Realistic Urban Social Networks". *Nature* 429:180–184.

Eubank, S., V. S. A. Kumar, M. Marathe, A. Srinivasan, and N. Wang.. 2006. "Structure of social networks and their impact on epidemics.". In *Discrete Methods in Epidemiology*, edited by G. Abello J, Cormode, Volume 70 of *DIMACS Series in Discrete Mathe-matics and Theoretical Computer Science*, 179–185.

Ferguson, N. M., D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai1, S. Iamsirithaworn, and D. S. Burke. 2005. "Strategies for containing an emerging influenza pandemic in Southeast Asia". *Nature* 437:209–214.

Ferguson, N. M., D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke. 2006. "Strategies for mitigating an influenza pandemic". *Nature* 442:448–452.

Ferguson, N. M., C. Fraser, C. A. Donnelly, A. C. Ghani, and R. M. Anderson. 2004. "Public Health Risk from the Avian H5N1 Influenza Epidemic". *Science* 304 (5673): 968–969.

Germann, T. C., K. Kadau, I. M. Longini, and C. A. Macken. 2006. "Mitigation strategies for pandemic influenza in the United States". *Proceedings of the National Academy of Sciences* 103 (15): 5935–5940.

Halloran, H., I. Longini Jr., A. Nizam, and Y. Yang. 2002. "Possible Containment of Bio-terrorist Smallpox". *Science* 298:1428–1432.

Halloran, M., N. Ferguson, S. Eubank, I. Longini, D. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. Germann, D. Wagener, R. Beckman, K. Kadau, C. Barrett, C. Macken, D. Burke, and P. Cooley. 2008, March 10. "Modeling targeted layered containment of an influenza pandemic in the United States". *Proceedings of the National Academy of Sciences (PNAS)* 105 (12): 4639.

Hethcote, H. W. 2000. "The mathematics of infectious diseases". *SIAM Review* 42:599–653.

Kalé, L., and S. Krishnan. 1993, September. "CHARM++: A Portable Concurrent Object Oriented System Based on C++". In *Proceedings of OOPSLA'93*, edited by A. Paepcke, 91–108: ACM Press.

Kaplan, E., D. Craft, and L. Wein. 2002. "Emergency response to a smallpox attack: the case for mass vaccinatio". *Proc. Natl. Acad. Sci. USA* 99:10935–10940.

Livnat, Y., P. Gesteland, J. Benuzillo, W. Pettey, D. Bolton, F. Drews, H. Kramer, and M. Samore. 2010. "Epinome - A Novel Workbench for Epidemic Investigation and Analysis of Search Strategies in Public Health Practice". *Proceedings of the Annual American Medical Informatics Association Symposium*:647–651.

Livnat, Y., P. Gesteland, and M. Samore. 2012. "Loosely Coordinated Multiple Views". *IEEE Visualization, submitted*.

Longini, I. M., A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. T. Cummings, and M. E. Halloran. 2005. "Containing Pandemic Influenza at the Source". *Science* 309:1083–1087.

Meyers, L. A., M. E. J. Newman, M. Martin, and S. Schrag. 2003. "Applying network theory to epidemics: Control measures for outbreaks of Mycoplasma pneumoniae". *Emerging Infectious Diseases*:204–210.

Monge, P. R., and N. S. Contractor. 2003. *Theories of Communication Networks*. Oxford University Press.

Newman, M. 2003. "The structure and function of complex networks". *SIAM Review* 45.

Wendelsdorf, K., J. Bassaganya-Riera, K. Bisset, S. Eubank, R. Hontecillas, and M. Marathe. 2011. "ENteric Immunity Simulator: A tool for in silico study of gut immunopathologies". *Proceedings of the IEEE International Conference Bioinformatics and Biomedicine*.

## AUTHOR BIOGRAPHIES

**KEITH R. BISSET** is senior research associate in the Network Dynamics and Simulation Science Laboratory at Virginia Tech. He received his PhD degree in Computer Science from New Mexico State University. His research interests include high performance computing, parallel and distributed simulation, and modeling complex networks. He is a member of the IEEE Computer Society and ACM. His email address is kbisset@vbi.vt.edu.

**STEPHEN EUBANK** is a Professor and Deputy Director of Network Dynamics and Simulation Science Laboratory, Virginia Bioinofrmatics Institute at Virginia Tech. His email address is seubank@vbi.vt.edu.

**MADHAV M. MARATHE** is a Professor of Computer Science, and Deputy Director of Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute at Virginia Tech. His email address is mmarathe@vbi.vt.edu.