DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

**Defense Threat Reduction Agency**
**8725 John J. Kingman Road, MS**
**6201 Fort Belvoir, VA 22060-6201**

**DTRA-TR-16-14**

TECHNICAL REPORT

# High Performance Computing Methods for Inference State Assessment and Course of Action Analysis in Large Socio-Technical Networks

**Distribution Statement A.** Approved for public release; distribution is unlimited.

September 2018

Grant HDTRA1-09-1-0017

Christopher Barrett et al.

Prepared by:
Virginia Bioinformatics Institute
Virginia Tech
1880 Pratt Drive, Research
Building XV
Blacksburg, VA 24060

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

| | | |
|---|---|---|
| **1. REPORT DATE** *(DD-MM-YYYY)*<br>00-09-2018 | **2. REPORT TYPE**<br>Technical | **3. DATES COVERED** *(From - To)*<br>March 11, 2009 - July 14, 2012 |

**4. TITLE AND SUBTITLE**

High Performance Computing Methods for Inference State Assessment and Course of Action Analysis in Large Socio-Technical Networks

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
HDTRA1-09-1-0017

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Christopher Barrett
Madhav Marathe
Stephen Eubank

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Network Dynamics & Simulation Science Laboratory
Virginia Bioinformatics Institute
Virginia Tech
1880 Pratt Drive, Research Building XV
Blacksburg, VA 24060

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Defense Threat Reduction Agency
8725 John J. Kingman Road, STOP 6201
Fort Belvoir, VA 22060-6201

**10. SPONSOR/MONITOR'S ACRONYM(S)**
DTRA

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
DTRA-TR-16-14

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Distribution Statement A. Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

This is the final report for this project which has a budget closing date of July 14, 2012. In this report we have included results from earlier years to present and overview of all the research conducted. Here we have highlighted results and extensions that were developed in the final reporting period. A central theme in our work under this project has been to investigate the role of details in synthesis and analysis of complex social networks.

**15. SUBJECT TERMS**

high performance computing, socio-technical networks, WMD, sub-networks

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON**<br>Paul Tandy |
|---|---|---|---|---|---|
| **a. REPORT**<br>Unclassified | **b. ABSTRACT**<br>Unclassified | **c. THIS PAGE**<br>Unclassified | SAR | 30 | **19b. TELEPHONE NUMBER** *(include area code)*<br>571-616-4097 |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

# UNIT CONVERSION TABLE

## U.S. customary units to and from international units of measurement[*]

| U.S. Customary Units | Multiply by → <br> ← Divide by[†] | International Units |
|---|---|---|
| **Length/Area/Volume** | | |
| inch (in) | $2.54 \times 10^{-2}$ | meter (m) |
| foot (ft) | $3.048 \times 10^{-1}$ | meter (m) |
| yard (yd) | $9.144 \times 10^{-1}$ | meter (m) |
| mile (mi, international) | $1.609\,344 \times 10^{3}$ | meter (m) |
| mile (nmi, nautical, U.S.) | $1.852 \times 10^{3}$ | meter (m) |
| barn (b) | $1 \times 10^{-28}$ | square meter ($m^2$) |
| gallon (gal, U.S. liquid) | $3.785\,412 \times 10^{-3}$ | cubic meter ($m^3$) |
| cubic foot ($ft^3$) | $2.831\,685 \times 10^{-2}$ | cubic meter ($m^3$) |
| **Mass/Density** | | |
| pound (lb) | $4.535\,924 \times 10^{-1}$ | kilogram (kg) |
| unified atomic mass unit (amu) | $1.660\,539 \times 10^{-27}$ | kilogram (kg) |
| pound-mass per cubic foot ($lb\ ft^{-3}$) | $1.601\,846 \times 10^{1}$ | kilogram per cubic meter ($kg\ m^{-3}$) |
| pound-force (lbf avoirdupois) | 4.448 222 | newton (N) |
| **Energy/Work/Power** | | |
| electron volt (eV) | $1.602\,177 \times 10^{-19}$ | joule (J) |
| erg | $1 \times 10^{-7}$ | joule (J) |
| kiloton (kt) (TNT equivalent) | $4.184 \times 10^{12}$ | joule (J) |
| British thermal unit (Btu) (thermochemical) | $1.054\,350 \times 10^{3}$ | joule (J) |
| foot-pound-force (ft lbf) | 1.355 818 | joule (J) |
| calorie (cal) (thermochemical) | 4.184 | joule (J) |
| **Pressure** | | |
| atmosphere (atm) | $1.013\,250 \times 10^{5}$ | pascal (Pa) |
| pound force per square inch (psi) | $6.984\,757 \times 10^{3}$ | pascal (Pa) |
| **Temperature** | | |
| degree Fahrenheit (°F) | $[T(^{o}F) - 32]/1.8$ | degree Celsius (°C) |
| degree Fahrenheit (°F) | $[T(^{o}F) + 459.67]/1.8$ | kelvin (K) |
| **Radiation** | | |
| curie (Ci) [activity of radionuclides] | $3.7 \times 10^{10}$ | per second ($s^{-1}$) [becquerel (Bq)] |
| roentgen (R) [air exposure] | $2.579\,760 \times 10^{-4}$ | coulomb per kilogram ($C\ kg^{-1}$) |
| rad [absorbed dose] | $1 \times 10^{-2}$ | joule per kilogram ($J\ kg^{-1}$) [gray (Gy)] |
| rem [equivalent and effective dose] | $1 \times 10^{-2}$ | joule per kilogram ($J\ kg^{-1}$) [sievert (Sv)] |

[*] Specific details regarding the implementation of SI units may be viewed at http://www.bipm.org/en/si/.

[†] Multiply the U.S. customary unit by the factor to get the international unit. Divide the international unit by the factor to get the U.S. customary unit.

2015-11-16

## Status:

This is the final report for this project which has a budget closing date of July 14, 2012. In this report we have included results from earlier years to present an overview of all the research conducted. Here we have highlighted results and extensions that were developed in the final reporting period.

## Accomplishments:

For this project, the main goal is to develop novel, multi-theory, agent-based, high performance computing (HPC)-based representation methods, and models coupled with analytical and algorithmic techniques for inference and state assessment of complex socio-technical networks. Sophisticated agent interaction-generated network representations of complex events will greatly extend the ability to deal with a variety of modern military missions and circumstances. For example, the range of technical, social, economic and military aspects of the use and proliferation of CBRN weapons is the kind of interwoven, heterogeneous and asymmetric, detail and individual interaction-rich situation that this approach will transform. However, simplistic approaches to agent-based representations and analysis can lead to a false sense of "surface validity". Our work has taken the first steps towards better understanding complex socio-technical networks of interest to DTRA and has provided new methods for preventing, responding to and recovering from such events as they occur in relation to WMD attacks.

A central theme in our work under this project has been to investigate the role of details in synthesis and analysis of complex social networks. Specifically, we have studied the following general questions: (i) what level of aggregation is adequate when studying dynamics over socio-technical networks, and (ii) how important are specific agent-agent interactions; meaning even when we have a model which is highly resolved (in terms of number of entities and interactions), how precise do we need to be in the representation of the interactions and the relational attributes of individual agents and the interactions among them. These questions pervade each of the topical areas proposed. The final reporting period also covers more work on sensitivity and uncertainty quantification related to these questions.

Work in past years has also focused on new generalized threshold models for social systems and spread of complex contagions. Classic threshold models capture basic phenomena, but our extensions developed throughout the project, and also this final year, have greatly extended our model variety and have led to development of theory for their behavior. The insight we have obtained on the properties of these models is of direct use in model design, verification, validation and system learning.

The major goals of the project are summarized in the following three main tasks with sub-tasks as indicated.

### Task 1: Network Inference: Structure, Attributes, Behavioral Characteristics and Co-Evolution.

1.1 Synthesis and embedding of community networks.
1.2 Inference of behavioral characteristics of networked systems: modeling spread of

malware over large communication networks.

1.3 Develop local and global structural measures for assessing sensitivity and robustness of networks and their evolution

**Task 2: Development of agent-based inductive models for situational awareness and consequence analysis related to acquisition, proliferation and use of CBRN weapons.**

2.1 Models and characterizations of social interactions.
2.2 Development of generalized social diffusion models for complex contagions.
2.3 Methods for accentuating and thwarting diffusion of complex contagions.
2.4 Human initiated cascading failures in inter-dependent infrastructure: modeling behavioral adaptation.
2.5 Develop model-based situational assessment using high performance simulation for planning and response for WMD

**Task 3: Statistical and analytical methods for quantifying the uncertainty in state assessment of socio-technical networks and robust stochastic optimization methods for ranking potential course of actions.**

3.1 Statistical test for required level-of-detail in model design for epidemics.
3.2 Supervised and unsupervised learning for clustering epidemic curves and categorizing real-time surveillance data.
3.3 Effects of demographic and spatial variability on epidemics – regional differences.
3.4 Effects of demographic and spatial variability on epidemics – national differences.
3.5 Structural changes to networks, measures and invariance
3.6 Network model sensitivity analysis of specific models

The work undertaken in this project is being leveraged by the DTRA CNIMS project. Conversely, the DTRA CNIMS project has provided us with important practical problems and has made us aware of the problems faced by DTRA in providing reach-back to its DoD customers. This has crucially allowed us to work on problems and develop technologies that can be transitioned to DOD in the near future. Examples of such technology include:

- Developing a formal methodology for embedding sub-networks of interest within the larger social network. The problem was directly motivated by the need at DOD for pandemic planning and response as it pertains to the US military bases. The methodology we are developing can be transitioned via the CNIMS project to develop more realistic models of US bases and embed them within the large social contact networks.
- One of DTRA's missions is to understand the impact of WMDs. This has motivated us to study the cascading failures amongst inter-dependent critical infrastructures. Our original results provide a formal methodology to represent and study such issues. The role of human-initiated cascades is an important new idea that will see useful applications in the near future. Under the DTRA CNIMS project we have developed

the SIS system (described later) to study such coupled infrastructures. The system was designed and developed during this reporting period and is continuously being extended.

- The study of various diffusion processes over socio-technical networks is motivated by DTRA's mission to understand how adversarial groups might get access to CBRN weapons, their motivation, their intent and ways of thwarting these.

Our recently awarded DTRA grant HDTRA1-11-1-0016 titled "Rigorous Approaches for Validation and Verification of Networked Systems" has also provided avenues for leverage in both directions. The work under this grant and the HDTRA1-11-1-0016 complement each other, and in some places in this report we have pointed out how this interplay has benefitted one or both projects.

## Results:

### Overview:

Our results provide new methodologies for studying the questions above. Furthermore, they provide new insights into specific dynamics we investigated that were directly motivated by DTRA mission – spread of infectious diseases and spread of complex contagion such as behaviors, fads, ideologies, and cascading infrastructure failures caused by WMD. The discussion below highlights the important findings. Additional information can be found in the listed publications.

We have written over 60 research articles; in addition, we have written a number of papers that summarize our efforts and are accessible to the broader scientific community [AK+09, BB+09, BE+09, BL+09, KMP10]. Highlights of our work under this project include:

(i) Development of embedding methodologies to incorporate detailed sub-networks within larger networks in the context of military and high school student sub-populations. We call this process "zooming" analogous to how satellite images are presented. Initial results indicate that certain details when embedding these sub-networks have a significant impact on disease dynamics. This work is central for developing general modeling principles under Topic 1. The analysis has been applied further during the final reporting year in conjunction with work under HDTRA1-11-1-0016 developing the theory of the methodology.

(ii) Construction of models for complex contagions. Complex and simple contagions are shown to be different on realistic social networks. Examples include spread of rumors, strikes, and social movements. Two main problems being studied are *control* (controlling the contagion by "immunizing" certain nodes) and *influence maximization* (maximizing the spread of contagion by choosing the right nodes to "infect"). New models have been introduced in this final year.

(iii) Development of mathematical and statistical methods for assessing the level-of-detail required in models to answer questions pertaining to epidemics and course-of-action planning based on initial outbreak data. We have conducted studies demonstrating how details matter in all aspects of modeling of networked systems. We have developed the “edge swap” methodology further this year under HDTRA1-11-1-0016 and applied it to specific cases for this grant. It is central for measuring how much and what kinds of details are important.

(iv) We have also conducted statistical sensitivity studies in the context of epidemics. The studies helped identify which parameters of the stochastic *SEIR* model are most important for assessing the disease dynamics. The methodology extends to other stochastic diffusion processes over networks and also supports the question of determining the required level of detail for given analyses.

This section gives specific results broken down by major task numbers.

**Task/Topic 1: Network Inference: Structure, Attributes, Behavioral Characteristics and Co-Evolution.**

1.1 Synthesis and embedding of community networks: Synthesis and embedding of sub-networks within existing networks is central to many studies of infrastructure robustness. Examples include identifying and adding suitable attributes to particular demographic groups such as a military sub-population in the case of force-readiness studies during an epidemic. The recent H1N1 pandemic highlighted this issue, and we worked with DTRA personnel in supporting reachback work. The methodology developed here has been leveraged by the DTRA CNIMS project. In order to embed a “zoomed” network, an existing contact network should be modified as follows: (1) identifying vertices corresponding to military personnel, and (2) by replacing this induced sub-network by a new and typically more detailed sub-network that takes into account military attributes and contacts of military nature in addition to other normal contacts. Here, not only must the military personnel sub-network be captured accurately, but its embedding within the remaining network must also be precise.

In the context of epidemiology, simulations show that schools play a critical role in disease propagation for society as a whole. As for military sub-populations, one faces the same issues when investigating how the detailed structure of student contacts affect disease dynamics: one will have to identify the student sub-population, one must derive detailed contact patterns for the students, and one must embed this new graph structure within the existing social contact graph.

The methods we developed for constructing synthetic community networks (e.g. military, schools, colleges, etc.) in the previous reporting periods have been extended and refined further [XC+12] under our DTRA grant on Verification and Validation and have been applied under this grant. Here, the detailed sub-networks that were developed have now been embedded in the larger networks, and this has been tested for synthetic networks of military bases and high schools in the US. The work for military bases was done in conjunction with the DTRA CNIMS project. For high school networks, we have synthesized detailed contact networks for three high schools in Montgomery County. These three sub-networks have now been embedded in the New River Valley synthetic population. Here our approach was based on complete, detailed, anonymized classroom assignments as well as surveys seeking to assess contact patterns among students outside of class, but also with other members of the population.

Initial studies demonstrate several interesting facts. First, when comparing the original contact network to the network with the embedded detailed sub-network, graph measures such as degree distribution are virtually identical for the two networks. However, for many choices of embedding parameters, the two networks demonstrate significantly different disease dynamics. Specifically, the network with the detailed embedded sub-network, generally seem to give larger epidemic sizes and the epidemic peak generally occurs at an earlier stage. Details are still being analyzed.

Both the examples we have used (military subpopulation with vignettes, and students with class schedules combined with publically available data) represent interesting cases of the general problem. We expect the methods that we are developing here can be generalized to other sub-populations and their embeddings. Examples include college campuses, large business campuses, and villages.

New results for this work conducted during the final reporting period includes application of new edge-swap methods to assess sensitivity with respect to the embedding [XC-2012].

1.2 Inference of behavioral characteristics of networked systems: modeling spread of malware over large communication networks: We have developed highly scalable models of malware propagation in next generation communication networks. The importance of cyber-attacks is evident from a number of recent reports, progressive use of scada and communication networks to manage critical infrastructures and the establishment of cyber command by the DoD. For this research, we constructed an enhanced EpiNet simulation framework (named **Fast-EpiNet**) with several modifications and new features that address the scalability issues of the parallel implementation of EpiNet. This framework augments the functionality of EpiNet on infection propagation to include: (*i*) Device-based detection mechanisms to detect infected devices based on behavior, (*ii*) Implementation of response policies at device- and system-levels to control the spread and (*iii*) Incorporation of ways to handle behavioral adaptations for individuals to dynamically react to the malware spread. Our work has been published in [BB+10,BC+10a,BC+10b,BC+10c,CC+09,TMB10]

The detection mechanisms modeled are derived from implementable policies that depend on malware signatures (files stored at certain locations on the device, sequence of system calls, etc.). Deviation from the normal behavior of the device is deemed as an indicator for an infection. Fast-EpiNet also provides a framework to implement and study response policies at device- and system-levels. The framework can now be used to analyze the impact of applying response mechanisms and determine the efficacy of the response. Fast-EpiNet now provides an environment in which policies can be deployed and evaluated with fine granularity. To incorporate the behavioral adaptations, we have implemented dynamic “triggers” that enable response mechanisms to be activated by malware prevalence. These triggers are not set up statically, but once enabled are dynamically activated based on certain pre-defined conditions. We have implemented both global and neighborhood based triggers that are activated when a certain number of devices become infected in the entire network or in the neighborhood of a certain susceptible device. Fast-EpiNet can:

1. Handle large networks. For example, a synthetic wireless network spanning the urban region of Miami comprising 1.2 million nodes can be simulated in about an hour. In contrast, EpiNet cannot process such large networks even with increasing processing elements (PEs).
2. Utilize more PEs to study large networks. For example, we show weak scaling of Fast-EpiNet on the Miami network with 80 PEs. Communication overheads prevent this in case of EpiNet.
3. Fast-EpiNet's computational framework allows representation of complicated static (offline) and dynamic (online) interventions.

### 1.3 Develop local and global structural measures for assessing sensitivity and robustness of networks and their evolution.

As a part of our project HDTRA1-11-1-0016 we have developed extensions of our edge-swap methodology. This work includes developing more theoretical insight into the edge-swap process and its properties. The methodology has been applied in this project, for example, to assess sensitivity with respect to the sub-network embedding described above [XC-2012].

## Task/Topic 2: Development of agent-based inductive models for situational awareness and consequence analysis related to acquisition, proliferation and use of CBRN weapons.

2.1 Models and characterizations of social interactions: We have worked with dynamic interaction data from a dormitory collected via mobile phones. These data cover approximately 80 individuals for a nine month period. Subjects were given smart phones that scanned the immediate environment for other Bluetooth devices (in particular, the phones in the study) every five minutes. There is thus very fine granularity in the proximity of individuals during this period. This data was coupled with various symptomatic information (collected via survey), as well as a wide range of other information such as mood, body-mass index, political beliefs, and eating habits. Part of the work was devoted to understanding the statistical structure within the data, with a particular focus on discerning the signal in the noise. Thus, for example, some “proximity” might involve proximity in adjacent rooms, which is quite a different phenomenon than proximity in the same room. Partially in parallel we examined the behavioral correlates of particular symptoms, with a preliminary set of findings in the Ubicomp paper [MC+10]. New results developed in final reporting period demonstrate how the meeting time proximity can be used to control attack rates. Additionally, the effect of scheduling (order of meetings/encounters) has been analyzed; see [DH+12, DO+12, DL+11, DW+12].

2.2 Development of generalized social diffusion models for complex contagions: We have designed and developed a general framework for evaluating complex contagions that has two components: (1) a transmission module and (2) a state transition module. The framework allows user-defined modules to be plugged into the system. We have implemented several models of each type, and most are agnostic in the sense that, for example, one transmission model can be used with many state transition models, for greater ranges in behaviors. Transmission models include a probabilistic phase to determine whether a contagious or affected agent transmits contagion to a susceptible agent during a pair-wise interaction. If so, then the amount of contagion transmitted can be constant, probabilistic (e.g., determined from a uniform distribution within user-defined limits), and/or a function of agent traits. The last feature, which is not yet implemented, enables customized transmission at an individual agent level. State transition models include user-specified constant thresholds (where a threshold $t = 3$, for example, can mean that an agent requires multiple interactions with contagious agents in order to change state) and a state change probability distribution where below a threshold, the probability of state transition is small, and above the threshold, the probability is greater. Other user-specified features include memory, where the effects of multiple contagious interactions over a finite timeframe are combined, and a “unique infectors” switch, such that if it is “on” then each contagious agent can contribute at most one interaction to the state transition of a susceptible agent. One implemented model, for example, includes a linear decay over a time window so interactions that are further in the past have lesser influence. These features significantly extend those of classical contagion models, and even those of generalized contagion models (e.g., [KM+10,KK+10]), and recover the earlier models as special cases.

Several other models have been developed and extended as a part of this project but most notably under the HDTRA1-11-1-0016 project during the final year. Examples include bi-threshold systems which distinguish the thresholds for the up- and down-transitions [KM+11], multi-state & multi-threshold systems developing bi-threshold systems even further [KM+12] and dynamic threshold systems capturing the case where the threshold can change with the dynamics [CC+12].

*Examples of behaviors.* The general diffusion model allows us to represent a range of social and behavioral theories, including theories of collective behavior [Gr78, DW05, CM07, CF07, N09]. The table below provides a listing of selected threshold behaviors from published sources. Not all of these have been modeled by the various sources; [Gr78] takes a step in this direction with regard to social movements. One of our goals is to be able to simulate these and other behaviors.  The table below contains published examples of threshold behaviors in social systems [Gr78, DW05, CM07, CF07].

| **Number** | **Threshold Example** |
|---|---|
| 1 | diffusion of innovations |
| 2 | rumors and disease |
| 3 | strikes |

| | |
|---|---|
| 4 | voting |
| 5 | educational attainment |
| 6 | obesity |
| 7 | social movements |

*Illustrative results.* The models and their implementations described above have been incorporated into the EpiSimdemics code, a parallel discrete event simulator. Illustrative results of the effects of threshold and deterministic and probabilistic state transition functions are given in Figure 1 for simulations of the spread of information through Miami, Florida over a 100-day period. The simulations include 2.1 million people and over 50 million daily interactions. The state transition model is a "0-to-1" (Boolean) threshold symmetric ratchet-up model, meaning that an agent either does not possess the information (state 0) or is informed (state 1), and that once an agent receives the information, she retains it and continually tries to pass it on to all of her neighbors. The plots show the running total fraction of the population that is informed over time.

A state transition occurs from 0 to 1 when the number of informed neighbors $n_t$ of an agent is $n_t \geq t$. In Figure 1 (left side), the retarding effect of increasing threshold is evident: a distinct phase transition occurs, from an almost immediate cascade for t = 2 to a delayed cascade when $t = 5$ to zero spread for $t \geq 10$. Figure 1 (right side) shows the effect of a probabilistic transition function. This model operates as follows: if $n_t < t$, then a state transition occurs with probability 0.01; if $n_t >= t$, then a state transition occurs with probability 0.9. The probabilistic transition model enables far greater diffusion, particularly for larger thresholds, but the rate of spread is a function of $t$. These results suggest the types of differences in behaviors can be realized with different models, and point to the need to characterize the features of different diffusive phenomena.

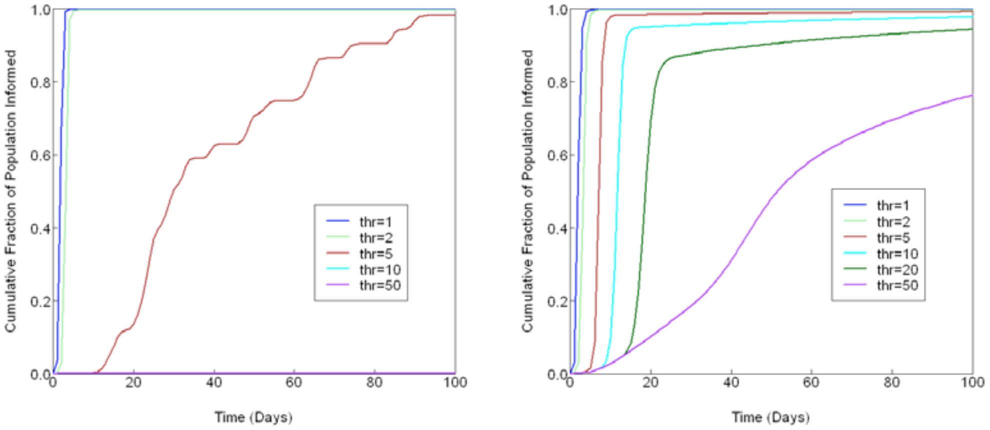

Figure 1. Cumulative fraction of the population informed as a function of time for information spread in Miami, Florida using (left side) deterministic transitions and (right side) probabilistic transitions. Here *thr* denotes the threshold.

The focus to this point has been on developing and implementing features of generalized contagions. Timing studies have shown that execution times can increase, over those of the simplest model in EpiSimdemics, by some 25% to 40%. Compared to the execution times of EpiSimdemics with the pre-existing simple-contagion model, the execution durations for complex contagions will increase. Our general scheme uses polymorphism, which means resolution of transmission and transition models must be done at runtime and hence incurs overhead. Also, some of these more detailed models are significantly more complicated and are composed of many more lines of code than required for the simple contagion model. Nonetheless, we believe we can bring down these execution times. Results on various simple and complex contagions and behavioral adaptations appear in [SG09,SG10,KL+10,KM+10,FS+09,CMM10]

2.3 Methods for accentuating and thwarting diffusion of complex contagions: We have examined two related problems. The first case analyzes how to increase the likelihood of diffusion of complex contagions spreading among agents in a network. The solution to this first problem involves computing *k*-cores of the graph. Seeding higher cores results in greater likelihood of wide-spread diffusion. We also determined an upper bound on the maximum spread size. This is an example of using a static property (*k*-cores of a graph) to compute a dynamical property: the extent of diffusion spread. This work is detailed [KK+11].

The second problem of thwarting diffusion is shown to be NP-hard for complex contagions. Thus, heuristics were devised and tested against three networks from the literature. The methods were shown to be effective in inhibiting diffusion. Results are contained in [KK+10].

2.4 Human initiated cascading failures in inter-dependent infrastructure: modeling behavioral adaptation: We have developed an integrated modeling environment to study *human-initiated* cascading failures in coupled transportation, social and cellular infrastructure systems. The research makes the following broad contributions: (i) it develops an individual based modeling framework for representing the social, transportation and cellular systems, and (ii) conducts illustrative case studies that demonstrate this modeling environment – the study is chosen to highlight the individual behavioral adaptation in the event of a no-notice crisis and its emergent effect on multiple infrastructures and the feedback that results from these interactions. The modeling environment provides policy makers and analysts a way to compare various response strategies and what-if scenarios. Specific contributions include:

We constructed a first principles based approach for synthetic wireless cellular network traffic during an evacuation. Here, we developed a model for wireless cellular network traffic during an evacuation in the downtown region of Portland, Oregon. This involves the following technical extensions: (i) construction of an instance of the activity based mobility model with activities for all individuals in the evacuation region changed because of the evacuation, and (ii) modification of the calling patterns for individuals during the evacuation period. Note that virtual data sets (at least in the open literature) exist for activity and call pattern changes during evacuations and disasters. In this study, we used simple parameterized models for the two behavioral adaptations; the system itself can handle more complicated behavioral adaptation. Note that the two behavioral adaptations are coupled in that individuals use their devices while evacuation is underway. This induces time-varying loads on both these infrastructures; furthermore, the time series of the loads cannot be decoupled due to feedbacks.

We found the load characteristics show high variation at different base stations, depending on where they are located, and if there are significant call drops around the evacuation region, enabling us to identify critical towers. We studied a simple iterative strategy to balance the load in a hybrid mesh and cellular network, in which the excess load on a base station is routed to its neighboring cells has a significant impact on the call drops. The work appears in [BB+10].

See [CB+10, FS+10, GK+10, FS+09, MM+09] for related work we have done on behavioral modeling, adaptation and co-evolution with network structure.

2.5 Develop model-based situational assessment using high performance simulation for planning and response for WMD. As a part of the NPS-1 study developed under the DTRA CNIMS grant in the current reporting period, we have developed a comprehensive simulation system built directly on the formalism of graph dynamical systems [BM+12]. For the prototype study, we analyzed a hypothetical scenario with a small human-initiated WMD attack in Washington DC (national planning scenario 1). In this study and in on-going work, we have analyzed the effects of having full/partial communication and also of human behavioral factors in the time after the incidence.

The system (abbreviated SIS) is an implementation of highly detailed, coupled, networked socio-technical systems, and includes models and analysis for:

- communication via wireless and other devices
- individual behavioral properties and group dynamics
- health evolution
- transportation and routing
- infrastructure restoration

The simulation models, which are illustrated in Figure 1, are applied in the given sequence for each iteration of the full simulation model. In this system, the data flow is via databases. As a mathematical abstraction, the global simulation model is therefore a coupled, co-evolving sequential dynamical system [MM+09].

Having the simulation model firmly rooted in a precise mathematical framework ensures clarity of design as facilitates verification of implementation. Our earlier work on the foundations of these systems and the work conducted under the DTRA grant HDTRA1-11-1-0016 have assisted in all these aspects for the SIS model. With a formally specified model that is amenable to analysis, we also have more avenues for optimization ensuring efficient and provably correct computation.

For the purpose of HPC simulation models and validation, SIS offers a great opportunity for conducting large-scale sensitivity analysis for realistic networked systems. Network and agent behavioral sensitivities can be analyzed in isolation as well as in a comprehensive manner by appropriately configuring or adding models within SIS. In this manner, both practical systems as well as general theory can be advanced. The work conducted under this grant has supported the underlying theoretical foundations of SIS.

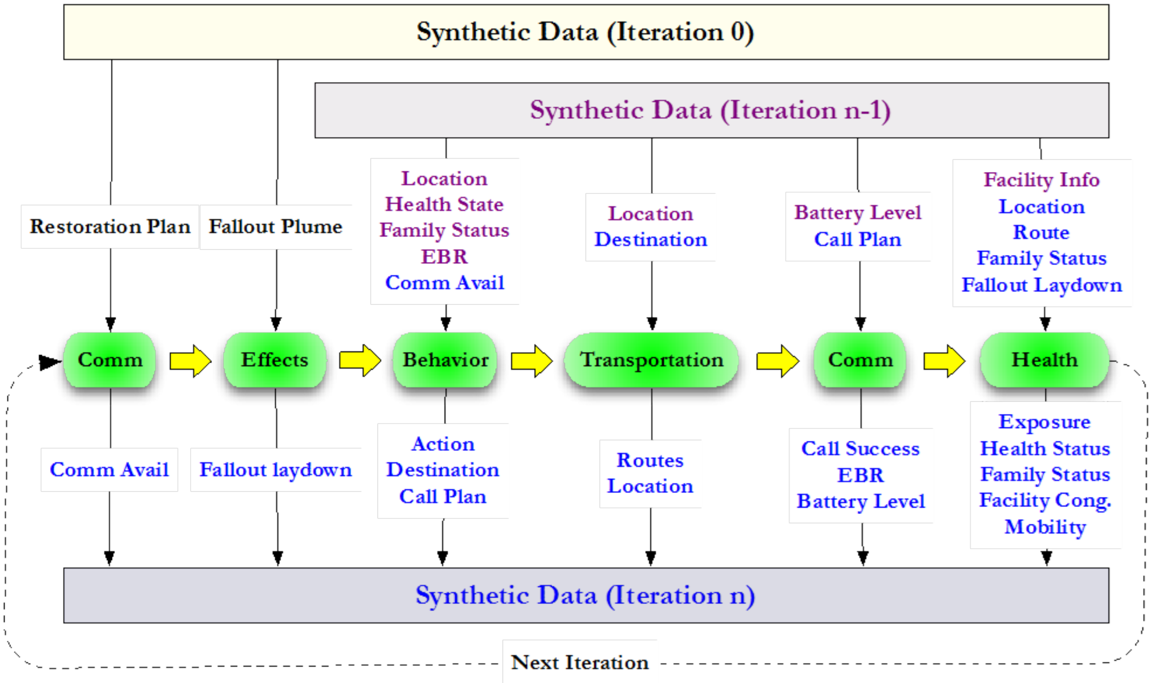

**Figure 1:** Modules and their organization in the SIS model.

### Development of web-enabled graph dynamical system analysis tool.

In accord with our plans before the final project year, we began the development of a web-enable computational tool to explore network dynamics on graphs (GDS calculator). Being mainly targeted at research and education, this web-enabled tool allows one to investigate phase spaces and trajectories of several classes of graph dynamical systems. Additionally, it will determine and explore all functional and cycle equivalence classes of sequential dynamical systems. This information allows one to assess stability and sensitivity aspects across the update scheme that is used. To bridge the gap to more realistic systems and their analysis, we have begun the development of the InterSim framework [KK+11b]. This framework scales to networks of size $10^6$ and is focused primarily on determining trajectories of systems and their long-term behaviors. The work will be continued under other on-going funded projects.

## Task/Topic 3: Statistical and analytical methods for quantifying the uncertainty in state assessment of socio-technical networks and robust stochastic optimization methods for ranking potential course of actions.

<u>3.1 Statistical test for required level-of-detail in model design for epidemics:</u> Network models of infectious disease epidemiology can potentially provide insight into how to tailor control strategies for specific regions, but only if the network adequately reflects the structure of the region's contact network. Typically, models that incorporate details

about human interactions produce the network. Each detail added renders the models more complicated and more difficult to calibrate, but also more faithful to the actual contact network structure. We have developed a statistical test to determine when sufficient detail has been added to the models, and have demonstrated its application to the models used to create a synthetic population and contact network for the USA. The work was published in [EV+10].

3.2 Supervised and unsupervised learning for clustering epidemic curves and categorizing real-time surveillance data: Classification and clustering techniques are widely used for identifying underlying groupings within datasets. We have applied unsupervised clustering techniques such as principal components clustering with *k*-medoids, and nonparametric Bayesian clustering with Dirichlet process priors to find groupings within epidemic curves generated from flu models with varying parameters. We have also evaluated the comparative performance of supervised classification techniques such as support vector machines, linear discriminant analysis, random forests and nearest neighbor in categorizing real-time surveillance data similar to simulated outbreaks from our disease models. In addition, we have performed nonparametric Bayesian clustering with the Dirichlet process for categorizing partial epidemic curves from real-time surveillance data for new disease outbreaks.

The unsupervised clustering methods correctly identified epidemic curve types with minimal misclassification error rate. The comparative performance of the supervised classification methods varied significantly depending on the length assessed into the outbreaks, the disease type and the distinctiveness between disease outbreaks. Classification techniques, which are easier to implement and consist of fewer assumptions regarding the structure of the outbreak, perform better than more complicated techniques. In addition, severe influenza outbreaks are easier to categorize since the epidemic curve structures are usually distinct from other epidemic curves. A preliminary version of our work has been presented at [NB+10].

3.3 Effects of demographic and spatial variability on epidemics – regional differences (Miami/Seattle): The goal of this research was to determine if the same influenza vaccination strategies would have the same level of effectiveness when applied to two different US metropolitan areas, Miami and Seattle. The composition of the population differs significantly in age distribution and household size distribution. We used an individual-based network modeling approach in which every pair of individuals connected in the social network is represented. Factorial design experiments were performed to estimate the impact of age-targeted vaccination strategies to control the transmission of a ‘flu-like’ virus. The findings showed that: (1) age composition of the city matters in determining the effectiveness of a vaccination strategy; and (2) vaccinating school children outperforms every other strategy. The most significant policy implication of this research is that there may not be a universal vaccination strategy, which works across all cities with the same level of effectiveness. Secondly, given the important role of school children in the transmission of influenza, the US Government should consider the vaccination of school children a top priority. The work appears in [TMB10].

3.4 Effects of demographic and spatial variability on epidemics – national differences (Beijing, Los Angeles, New Delhi): The structure of social contact networks influences the spread of infectious diseases in an urban region. An important goal of network science is the development of structure to function theory – identifying structural features of the social network that yield insights into the disease propagation. Effective pharmaceutical as well as non-pharmaceutical interventions can be identified by analyzing the social contact networks. In this work, motivated by the recent pandemic caused by H1N1, we focused on influenza-like diseases and corresponding people-people social contact networks and studied such social infrastructures in Beijing, China; Los Angeles, US; and New Delhi, India. We compared each infrastructure in terms of static structural properties, as well as disease dynamics and intervention efficacy.

We have generated synthetic populations and social contact networks for Los Angeles, Beijing and New Delhi using different methodologies. For Los Angeles, we used the US census data, Dun & Bradstreet location data, and activity survey data. For Beijing and New Delhi, we had only the Landscan population density data and limited census data. We developed a generic methodology that takes into account variable data availability and granularity across different regions of the world. This model, based on Landscan data, can be applied to generate a synthetic population, including individual demographics, home locations, and daily activities, for any area in the world. The Los Angeles, New Delhi and Beijing contact networks are not equal due to different construction methodologies and the fact that the two populations inherently have very different demographic structures and activity patterns. To construct the contact networks, we explicitly generated activity sequence of each individual in the population taking into account the variability in demographic and activity patterns for each city individually. Other works in current literature either ignore the detail activities of the individual persons or use the same generic model for every city/country in the world. This loses the crucial demographic and spatial variability, which has significant effect on network structure as well as disease dynamics.

For comparison, we have first computed major structural measures for the two networks, including degree distribution, clustering coefficient distribution, connectivity, and vulnerability distribution. We then use a high performance agent based simulation (EpiFast) to compare the efficacy of widely accepted public health interventions, which have different impacts on the epidemic progress in the two different populations. We are especially interested in non-pharmaceutical interventions (NPI), such as school closure, that play an important role while containing pandemics caused by emerging infectious diseases. Our results highlight the importance of spatial and demographic structure of the social contact network when designing effective interventions. For example, the distribution of school-aged children varies widely between the three cities – this difference affects the efficacy of NPIs such as school closures. Structural analysis of the networks provides important cues in this regard. The results have an important implication, namely, guidelines developed by global health organizations such as WHO should be evaluated and adapted by each country based on the specific demographic and spatial characteristics. Preliminary work appears in [CH+10].

3.5 Structural changes to networks, measures and invariance: The structure and dynamical properties of networked systems are often characterized by the degree distribution of the underlying graph. For instance, the Erdös-Renyi model of random graphs, where sampling is done by including each edge at random with independent probability *p,* is adequate to infer a number of structural properties of the generated graph. However, this is not the case for real world networks since the presence of edges is generally correlated with the presence of other edges. As a step towards making more realistic models, scientists have studied degree distributions of many real world networks which have often been claimed to obey power laws. A general assumption in this research is that the local properties of a network (often, even parameters such as the average degree and clustering coefficient) are adequate to determine its global structural and dynamical properties. However, the variability among the space of graphs having a specified degree distribution is not usually taken into account. Researchers prove general theorems that say something about the dynamical process on such power law graphs. These theorems are probabilistic statements – they usually say that certain properties hold either in expectation or some cases a slightly stronger statement is possible that says that this holds with high probability. In proving such theorems, the proofs usually average over all graphs with a given degree distribution. This approach, although promising, is simply inadequate when dealing with real world networks for the simple reason that real world networks constitute a small fraction of the space of all power law degree bounded networks. Moreover, (*i*) real world networks need not have power-law degree distributions, (*ii*) if they have there are many reasons why such a distribution could have come about, (*iii*) and just because the graph has power law degree distribution does not say much about other structural properties of the network. We have demonstrated through explicit constructions that local properties such as the degree distribution or assortativity are insufficient to characterize the global dynamics of reaction-diffusion processes. We used “edge flipping” to generate a Markov chain of random graphs with fixed degree distribution and rewired edges while keeping the degree distribution unchanged. We instantiated several such chains, starting from a highly structured, realistic representation of a social contact network and study the global dynamics of a canonical reaction-diffusion process: epidemic spread. Our main observations are:

1. **Local properties such as degree distributions and assortativity do not adequately characterize social contact networks:** we found huge variability in the space of graphs with the same degree distribution as our synthetic social contact networks. Specifically, the graphs generated by the swap chain (which have the same degree distribution) exhibit varying structural and epidemic properties. Furthermore, the difference in these properties increases as the Markov chain gets further from its starting point. We also consider variants of the swap chain which preserve assortativity constrains, e.g., based on age or degree, and find that the disease dynamics are still quite different.

2. **Social contact networks are a small and not easily accessible subset of the space of graphs with a given degree distribution:** we found that the “convergence time” with respect to epidemic properties (as quantified by the number of swaps taken for the Markov chain to reach a distribution of graphs with roughly fixed epidemic properties) is much larger for synthetic social contact networks than for random graphs with similar degree distributions. This suggests that the space of synthetic social contact networks form a much smaller subset of the whole space of graphs with the same degree distribution.

Our work provides a cautionary note on the use of local random graph models to infer global network and dynamical properties.

3.6 Statistical Sensitivity Analysis for Epidemic Models: In large-scale realistic models of socio-technical systems it can be challenging to identify which parameters affect outcomes the most. Key factors that complicate such an analysis include the size of the network, the many sources of data required for calibration and the large number of parameters involved. In [NB+12], we conducted a sensitivity analysis of an individual-based SEIR model for influenza-like epidemics. It should be noted that although the analysis was specific to disease propagation, the general principles extend to a broad range of diffusion phenomena over complex networked systems.

This study is motivated by the need to improve methods for real-time modeling and predicting during a pandemic [LF+11]. The usefulness of real-time modeling was illustrated during the 2009 influenza pandemic using both compartmental models [OC+10] and individual-based models [CM+11]. The three parameters used in this study were transmissibility, incubation period distribution, and infectious period distribution. Here, the transmissibility is the diffusion intensity of a disease through a population. The transmissibility is usually measured using the reproductive number, the number of secondary cases for each primary case. The incubation period is the interval during which infected individuals cannot spread the disease and usually lasts between 1-4 days for seasonal influenza [CD+10]. The infectious period duration is the period during which infected individuals can transmit the disease to susceptible individuals. The epidemics were analyzed using ANOVA, with Bonferroni correction for multiple comparisons where applicable, and clustering with principal components to identify groupings of epidemic curves. Key results of the analysis include:

(1) The transmissibility and mean infectious period duration significantly affected the time to peak, peak attack rate and total attack rate. In contrast, an increase in the mean incubation duration did not significantly affect the total attack rate, but slightly influenced the time to peak and peak attack rate.

(2) The mean of the incubation period distribution appeared to be the sole determinant of its effects on the epidemics. In contrast, the mean and variance of the infectious period distribution were needed to determine its influence on epidemic dynamics.

(3) Compared to the other parameters, the infectious period distribution exerted the strongest influence on the total attack rate and structure of the epidemic curves.

(4) The model sensitivity was consistent across social networks with demographic and rural-urban differences.

(5) School-age children had the highest age-specific attack rates irrespective of mean infectious period and susceptibility of the other age groups.

Although the study is specific to epidemics, we note that the methodology developed here allows us to identify critical parameters in a broad range of complex, networked, agent-based models and also identifies particular demographic groups and components that influence the dynamics the most. Moreover, the particular values observed for transmissibility in an actual epidemic can be used as a measure to predict the severity of that epidemic. This construction of this early-warning observable can also be generalized to other complex systems.

### Opportunities for training and professional development:

The project has supported graduate students that have completed master theses and doctoral thesis as well as students in pursuit of such degrees.

### Dissemination of results to communities of interest:

Results have been presented at conferences and scientific meetings as well as in peer-reviewed journals. Details are given in the references listed below in this report.

### Plans for next reporting period:

This is the final report of the project.

## Products:

### Journal publications, books (reporting period):

[CC+12] L. Chang, J. Cochran, H. S. Mortveit, S. Raval and M. Schroeder (2012). “Complex contagions and threshold dynamical system”. Status: in preparation, August 2012.

[KK+11] Kuhlman, Chris J., V. S. Anil Kumar, Madhav V. Marathe, Samarth Swarup, Gaurav Tuli, S. S. Ravi, and Daniel J. Rosenkrantz (2011). “Inhibiting the Diffusion of Contagions in Bi-Threshold Systems: Analytical and Experimental Results,” Proceedings of the AAAI Fall 2011 Symposium on Complex Adaptive Systems (CAS-AAAI 2011), pp. 91-100, Arlington, VA, 4-6 November 2011.

[KK+11b] Kuhlman, Chris J., V. S. Anil Kumar, Madhav V. Marathe, Henning S. Mortveit, Samarth Swarup, Gaurav Tuli, S. S. Ravi, and Daniel J. Rosenkrantz (2011). “A General-Purpose Graph Dynamical System Modeling Framework,” Proceedings of

the 2011 Winter Simulation Conference, 2011 (WSC 2011), Phoenix, AZ, 11-14 December 2011.

[KM+11a] Kuhlman, C., Mortveit, H. S., Murrugarra, D. and Kumar, V. S. A. (2012). "Bifurcations in Boolean Networks". In proceedings of Automata 2011: appeared in DMTCS proc AP, 2012, 29-46.

[KM+12] Kuhlman C, Mortveit H, (2012) Title: Multi-state, multi-threshold systems. In preparation, August 2012.

[BM+12] Barrett C, Bisset K, Chen J., Eubank S, Feng A, Kumar VS Anil, Marathe M, Marathe A, Mortveit H, Stretz P, Swarup S, Xie D (2012). DTRA CNIMS. *DTRA Report and Presentation*. National Capital Region, April 5, 2012.

[DH+12] Dong W, Heller K, Pentland A (2012). Social-interaction coupled hidden Markov model for tracking the spread of infection. In submission.

[DO+12] Dong W., Olguin-Olguin D., Waber B., Kim T., Pentland A. (2012). Modeling Organizational Dynamics from Sensor Network Data. In *Proceedings of 9th International Conference on Wearable and Implantable Body Sensor Networks*. London, U.K., 2012.

[DL+11] Dong W, Lepri B, Pentland A (2011). Modeling the Co-evolution of Behaviors and Social Relationships Using Mobile Phone Data. In Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia. Beijing, China, 2011:134-143. Invited and extended work in Tsinghua Science and Technology 2012, 17(2) 136-151.

[DW+12] Dong W., Heller K., Pentland A. (2012). Modeling Infection with Multi-agent Dynamics. In *Proceedings of the 2012 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction*. D.C., USA, 2012. **Best paper award**.

[NB+12] Elaine O. Nsoesie, Richard J. Beckman, Madhav V. Marathe (2012). Sensitivity Analysis of an Individual-Based Model for Simulation of Influenza Epidemics. Plos One. Under review.

[PD+12] Pan W., Dong W., Cebrian M., Kim T., Fowler J., Pentland A. (2012). Modeling Dynamical Influence in Human Interaction: Using Data to Make Better Inferences. IEEE Signal Processing Magazine, 2012, 29(2): 77-86.

[XC+12] Xia H, Chen J, Marathe M, Mortveit H (2012). Sensitivity analysis of network epidemic dynamics to the detailed structure in subnetworks. In preparation, August 2012.

**Presentations at conferences (reporting period):**

Barrett C, Bisset K, Chen J., Eubank S, Feng A, Kumar VS Anil, Marathe M, Marathe A, Mortveit H, Stretz P, Xie D (2012). DTRA CNIMS. *DTRA Presentation*. National Capital Region, April 5, 2012.

Dong W., Martino M., Lazer D. (2012). Why Temporal Resolution Matters in Epidemic Modeling? Contributed presentation at NetSci. Evanston, IL, USA, 2012.

Kuhlman, Chris J., Kumar, V. S. Anil, Marathe, Madhav V., Mortveit, Henning S., Ravi, S. S. and Rosenkrantz, Daniel J. (2012). "Dependence of Diffusion Dynamics on Network Construction Methods," The International Sunbelt Social Network Conference, Redondo Beach, CA, 12-18 March 2012.

Korkmaz, Gizem, Kuhlman, Chris J., Marathe, Achla, Marathe, Madhav V. and Vega-Redondo, Fernando (2012). "The Role of Communication via Online Social Networks in the Dynamics of Collective Action," The International Sunbelt Social Network Conference, Redondo Beach, CA, 12-18 March 2012.

Marathe M and Barrett C. (2012), Planning and Responding to Human Initiated Crisis: Role of Data Intensive Computing and Computational Socio-Technical Sciences, International Conference on Networks in Biology, Social Science and Engineering, Indian Institute of Sciences, July 2012.

Marathe M. (2012), Control and Optimization Problems in Co-evolving networks International Conference on Networks in Biology, Social Science and Engineering, Indian Institute of Sciences, July 2012.

Marathe M. (2012), Science and Engineering of Co-evolving Networks: Population Dynamics and Epidemics, Network Science in Electrical Engineering and Computer Science, Special year on Network Science, Indian Institute of Sciences, January 2012.

Marathe M. (2011), Network Science of Socially Coupled Systems: A Computational Viewpoint, ASSYST Workshop – Mathematics in Network Science: Implications to Socially Coupled Systems, November 2011.

Marathe M. (2011) Pervasive High Performance Computing Meets Network-Epidemiology, U. Warwick, Dept. of Mathematics.

Marathe M. (2011), Pervasive Informatics to Support Science and Engineering of Co-evolving Networks, Center for Human Computer Interaction, Virginia Tech, November 2011.

Marathe M. (2011), Policy Informatics for Co-evolving Socio-technical Networks: Issues in Believability and Usefulness, Isaac Newton Institute for Mathematical Sciences Accelerating Industrial Productivity via Deterministic Computer Experiments and Stochastic Simulation Experiments, Sept 2011.

Marathe M. (2011), Policy Informatics for Co-evolving Socio-technical Networks: Modeling Population Dynamics and Infectious Disease, Indian Institute of Science Education and Research, Pune, Sept 2011

Marathe M. (2011), From Desktops to Clouds: Informatics for Public Health and Systems Biology, BioCrats, Pune Sept 2011.

Marathe M. (2011), Pervasive Informatics to support Science and Engineering of Co-evolving Networks, Los Alamos National Laboratory, August 2011.

Mortveit. H. (2011). "The Dynamics Group of Asynchronous Systems". In *DISCO 2011*. Valparaiso, Chile, 24-26 November, 2011.

Mortveit, H., Murrugarra, D., Kuhlman, C. and Kumar V.S. Anil (2011). "Bifurcations in Boolean Networks". Automata 2011. Santiago Chile, November 2011.

**Website(s) or other Internet site(s):**

Nothing to report

**Technologies and techniques:**

Nothing to report

**Inventions, patent applications, and/or licenses:**

Nothing to report

**Other products:**

Nothing to report

## <u>Participants and collaborating organizations:</u>

**Individuals on project:**

**NDSSL Virginia Tech:**

| | |
|---|---|
| Barrett, Christopher[1] (PI) | Faculty |
| Beckman, Richard (Co-PI) | Research Faculty |
| Marathe, Madhav[1] (Co-PI) | Faculty |
| Mortveit, Henning[1] (Co-PI) | Faculty |
| | |
| Lewis, Bryan[1] | Research Faculty |
| | |
| Adiga, Abhijin[1] | Post-Doc |
| Pan, Zhengzheng[1] | Post-Doc |
| Swarup, Samarth[1] | Post-Doc |
| | |
| Channakeshava, Karthik | Graduate Student |
| Dutta, Tridib | Graduate Student |
| Tuli, Gaurav[1] | Graduate Student |
| Korkmaz, Gizem[1] | Graduate Student |
| Kuhlman, Christopher[1] | Graduate Student |
| Leidig, Jonathan | Graduate Student |
| Liao, Shaojuan[1] | Graduate Student |
| Nagaraj, Kalyani S. [1] | Graduate Student |
| Nsoesie, Elaine O. [1] | Graduate Student |
| Parikh, Nidhi [1] | Graduate Student |
| Rivers, Caitlin[1] | Graduate Student |
| Saha, Sudip | Graduate Student |
| Venugopal, Vivek | Graduate Student |
| Xia, Huadong[1] | Graduate Student |
| Zhao, Zhao[1] | Graduate Student |

**Northeastern University:**

| | |
|---|---|
| Lazer, David[1] (Co-PI) | Faculty, Kennedy School of Government & Northeastern University |

**Note:** Superscript [1] indicates personnel that charged to project during the final reporting period.)

**Organizations involved as partners:**

NDSSL Virginia Tech, Kennedy School of Government and Northeastern University

**Other collaborators or contacts.**

None

**<u>Impact:</u>**

1. Impact on principal disciplines:
   Results developed for threshold systems and generalized contagion models have advanced the state of the art for the mathematical and computational theory of graphical dynamical systems. New methods for inferring and analyzing relational networks have advanced the field of network science.

2. Impact on other disciplines:
   The frameworks and algorithms that are being developed constitute an important step towards achieving the goals of this project and, along with the mathematical findings regarding threshold systems, provide a basis for theoretical understanding of computational methods as well as modeling of many social systems.

3. Impact on development of human resources:
   Nothing to report.

4. Impact on physical, institutional and information resource that form infrastructure:
   Nothing to report.

5. Impact on technology transfer:
   Nothing to report.

6. Impact on society beyond science and technology.
   Nothing to report.

7. Dollar amount of budget spent in foreign countries: $0.00

## Changes/Problems:

No changes were made to the objectives or approach the study. Also, no problems or delays were encountered, and the budget/expenditures remained unaltered. The project did not involve human or animal subjects nor biohazards. The location of the project remained unchanged.

## References (developed in full/parts under project)

[AK+09] Apolloni A, Kumar VSA, Marathe M, Swarup S. (2009) Computational epidemiology in a connected world. *IEEE Computer Magazine*, **42**(12): 83-86.

[BA+11] Bisset K, Ashwin A, Marathe M, Feng W-C (2011) High performance biocomputing for simulating the spread of contagion over large contact networks. In *Proceedings of 1st IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. Orlando, Florida, February 3-5, 2011.

[BB+09] Barrett C, Bisset K, Leidig J, Marathe A, Marathe M (2009) Estimating the economic impact of public strategies for controlling an epidemic: A multi-agent approach. In *Proceedings of 21st Innovative Applications of Artificial Intelligence Conference,* Pasadena, Califorinia.

[BB+10] Barrett C, Bisset K, Leidig J, Marathe A, Marathe M (2010) An integrated modeling environment to study the co-evolution of networks, individual behavior and epidemics. *AI Magazine, Spring 2010;* **31** (1): 75-87.

[BB+10] Barrett C, Beckman R, Channakeshava K, Huang F, Kumar VSA, Marathe A, Marathe M, Pei G (2010). Cascading failures in multiple infrastructures: From transportation to communication network. In *Proceedings of Interacting Critical Infrastructures for the 21st Century,* Beijing, China, September 20-22, 2010.

[BC+09a] Bisset K, Chen J, Feng X, Kumar VSA, Marathe M (2009) 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),* pp 430-439, Yorktown Heights, New York.

[BC+10b] Barrett C, Channakeshava K, Eubank S, Kumar VSA, Marathe A, Marathe M (2010) From biological and social network metaphors to coupled bio-social wireless networks. *International Journal of Autonomous Adaptive Communications,* 4(2):122-144.

[BC+10c] Beckman R, Channakeshava K, Huang F, Marathe A, Marathe M, Pei G, Vullikanti A (2010) Synthesis and analysis of spatio-temporal spectrum demand patterns: A first principles approach. In *Proceedings of IEEE International Dynamic Spectrum Access Networks*, Singapore.

[BC+10] Beckman R, Channakeshava K, Huang F, Marathe A, Marathe M, Pei G, Kumar VSA (2010) Implications of dynamic spectrum access on the efficiency of primary wireless market. In *Proceedings of IEEE International Dynamic Spectrum Access Networks (DySpan),* Singapore.

[BH+11] Barrett C, Hunt HBI, Marathe M, Ravi S, Rosenkrantz D, Stearns R (2011) Modeling and analyzing social network dynamics using stochastic discrete graphical dynamical systems. *Theoretical Computer Science, Special Issue*. 412(30):3932-3946

[BC+10] Bisset K, Chen J, Feng X, Ma Y, Marathe M (2010) Indemics: An interactive data intensive framework for high performance epidemic simulation. In *Proceedings of 24th International Conference on Supercomputing,* pp. 233-242, Tsukuba, Japan**.**

[CB+11] Channakeshava K, Bisset K, Marathe M, Kumar VSA, Yardi S (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*. Anchorage, Alaska, May 16-20, 2011

[BE+09] Barrett C, Eubank S, Marathe M (2009). Policy informatics for co-evolving socio-technical networks. In *Proceedings of 4th International Conference on Critical Infrastructures,* Linkoping, Sweden.

[BL+09] Barrett C, Lewis B, Chen J, Kumar VSA, Eubank S, Marathe M, Mortveit, H (2009) Interactions among human behavior, social networks, and societal infrastructures: A case study in computational epidemiology. In *Fundamental Problems in Computing: Essays in Honor of Professor Daniel J. Rosenkrantz,* Ravi S, Shukla S (eds), XXII. Springer Verlag, pp 477-507.

[BM09] Bisset K, Marathe M (2009) A cyber environment to support pandemic planning and response. *DOE SCIDAC Magazine*, pp 36-47.

[CB+10] Channakeshava K, Bisset K, Kumar VSA, Marathe M, Yardi S (2010) Fast-EpiNet: An efficient and scalable framework for simulating wireless epidemics on multi-core clusters. *NDSSL Technical Report 10-037*.

[CC+09] Channakeshava K, Chafekar D, Bisset K, Kumar VSA, Marathe M (2009) EpiNet: A simulation framework to study the spread of malware in wireless networks. In *Proceeding of 2nd International Conference on Simulation Tools and Techniques (SIMUTools '09),* Rome, Italy.

[CH+10] Chen J, Huang F, Khan M, Marathe M, Stretz P, Xia H (2010) The effect of demographic and spatial variability on epidemics: A comparison between Beijing, Delhi and Los Angeles. In *Proceedings of Interacting Critical Infrastructures for the 21st Century*, Beijing, China, September 20-22, 2010.

[CMM10] Chen J, Marathe A, Marathe M (2010) Coevolution of epidemics, social networks and individual behaviors: A case study. In *Proceeding of 2010 International Conference on Social Computing, Behavioral Modeling & Prediction,* Lecture Notes in Computer Science, **6007**: 218-227

[EB+10] Eubank S, Barrett C, Beckman R, Bisset K, Durbeck L, Kuhlman C, Lewis B, Marathe A, Marathe M, Stretz P (2010) Detail in network models of epidemiology: Are we there yet? *Journal of Biological Dynamics*, Doi:10.1080/17513751003778687.

[EV+10] Eubank S, Vullikanti A, Khan M, Marathe M, Barrett C (2010) Beyond degree distributions: Local to global structure of social contact graphs. In *Proceedings of the 3rd International Conference on Social Computing, Behavioral Modeling and Prediction (SBP10)*, NIH, Bethesda, Maryland, March, 2010.

[FS+09] Fagyal Z, Swarup S, Escobar A, Gasser L, Lakkaraju K (2009) Centers, peripheries, and popularity: The emergence of norms in simulated networks of linguistic influence. *Penn Working Papers in Linguistics,* **15**(2)**:** Selected Papers from NWAV 27.

[FS+10] Fagyal Z, Swarup S, Escobar A, Gasser L, Kiran L (2010) Centers and peripheries: Network roles in language change. *Lingua*, **120**(8): 2061-2079.

[GK+10] Goldstein E, Apolloni A, Lewis B, Miller J, Macauley M, Eubank S, Lipsitch M, Wallinga J (2010) Distribution of vaccine/antivirals and the "Least Spread Line" in a stratified population. *Journal of the Royal Society Interface*, **7**(46): 755-764.

[HH+10] Han B, Hui P, Kumar VSA, Marathe M, Pei G, Srinivasan A (2010) Cellular traffic offloading through opportunistic communications: A case study. In *Proceedings of the 5th ACM Workshop on Challenged Networks (CHANTS)***:** 31-38. Chicago, Ilinois, September 2010.

[KK+09] Khan M, Kumar VSA, Marathe M, Pandurangan G, Ravi S (2009) Bi-Criteria approximation algorithms for power-efficient and low-interference topology control in unreliable ad hoc networks. In *Proceedings of the 28th Conference on Computer Communications,* Rio De Janeiro, Brazil.

[KK+10] Kuhlman C, Kumar VSA, Marathe M, Ravi S, Rosenkrantz D (2010) Finding critical nodes for inhibiting diffusion of complex contagions in social networks. In *Proceeding of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, Barcelona, Spain, September 20-24, 2010, **to appear.**

[KL+10] Kuhlman C, Lewis B, Beckman R, Eubank S, Dutta T (2010) Clustering method incorporating network topology and dynamics. In *Proceedings of Society for Modeling & Simulation International*, Orlando, Florida.

[KM+09] Kumar VSA, Marathe M, Parthasarathy S, Srinivasan A (2009) Scheduling on unrelated machines under tree-like precedence constraints. *Algorithmica*, **55**(1)**:** 205-206.

[KM+10] Kuhlman C, Marathe M, Ravi SS, Rosenkrantz D (2010) Exploiting network structure in enhancing diffusion of complex contagions. In *Proceedings of Analysis of Complex Networks (ACNE) Workshop of the ECML PKDD 2010 Conference*. Barcelona, Spain, September 20-24, 2010.

[KMP10] Kumar VSA, Marathe M, Parthasarathy S (2010) Cross-layer capacity estimation and throughput maximazation in wireless networks. In *Algorithms for Next Generation Networks*, Cormode G, Thottan M (eds), pp 67-98, Springer London.

[LF+11a] Leidig J, Fox E, Hall K, Marathe M, Mortveit H (2011). SimDL: A model ontology driven digital library for simulation systems. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries*; Ottawa, Canada, June 13-17, 2011.

[LF+11b] Leidig J, Fox EA, Hall K, Marathe M, Mortveit H (2011). Improving simulation management systems with model ontologies. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries*; Ottawa, Canada, June 13-17, 2011.

[MC+10] Madan M, Cebrian M, Lazer D, Pentland A (2010) Social sensing for epidemiological behavior change. In *Proceedings of 12th ACM International Conference on Ubiquitous Computing*, Copenhagen, Denmark (10% acceptance rate, nominated for best paper).

[MM+09] Macauley M, Mortveit H (2009) Cycle equivalence of graph dynamical systems. *Nonlinearity*, **2**: 421-436.

[MM+11] Macauley M, Mortveit H (2010) Update sequence stability in graph dynamical systems. *Discrete and Continuous Dynamical Systems S*, 4(6): 1533-1541.

[NB+10] Nsoesie E, Beckman R, Marathe M (2010) Estimation of an epidemic curve: A classification approach. To appear in *Proceedings of Joint Statistical Meetings 2010*

[Pa10] Pan Z (2010) *Trust, influence, and convergence of behavior in social networks*. *Mathematical Social Sciences*, **60**(1): 69-78.

[PK+11] Pei G, Kumar VSA, Parthasarathy S, Srinivasan A (2011). Approximation algorithms for throughput maximization in wireless networks with delay constraints. In *Proceedings of IEEE INFOCOM*; Shanghai, China, April 10-15, 2011.

[SG09] Swarup S, Gasser L (2009) The iterated classification game: A new model of the cultural transmission of language. *Adapative Behavior*, **17**(3)**:** 213-235.

[SG10] Swarup S, Gasser L (2010) The classification game: Combining supervised learning and language evolution. *Connection Science*, **22**(1): 1-24.

[TMB09] Taylor C, Marathe A, Beckman R (2009) Role of demographics in the performance of influenza intervention strategies. In *Proceedings of 4th International Conference on Social Science Research,* Nashville, Tennessee.

[TMB10] Taylor C, Marathe A, Beckman R (2010) Same influenza vaccination strategies but different outcome across US cities. *International Journal of Infectious Diseases*, 14(9):e792-5.

[XC+11] Huadong Xia, Jiangzhuo Chen, Madhav V. Marathe, Henning S. Mortveit, "Synthesis and Refinement of Detailed Subnetworks in a Social Contact Network for Epidemic Simulations," 2011 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP11)

[ZK+10] Zhao Z, Khan M, Kumar VSA and Marathe M (2010) Subgraph enumeration in large social contact networks using parallel color coding and streaming. In *Proceedings of 39th International Conference on Parallel Processing (ICPP),* San Diego, California.

## General References:

[CF07] Christakis N, Fowler J. (2007) The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357: 370-379.

[CM07] Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *American Journal of Sociology*, **113**(3): 702-734.

[CM+11] Chao DL, Matrajt L, Basta NE, Sugimoto JD, Dean B, et al. (2011) Planning for the control of pandemic influenza A (H1N1) in Los Angeles county and the united states. American Journal of Epidemiology 173: 1121-1130.

[DW05] Dodds P, Watts D (2005) A generalized contagion model of social and biological contagion. *Journal of Theoretical Biology*, **232**:587-604.

[CDC-2010] CDC (2010). Centers for disease control and prevention: Weekly flu reports. URL http://www.cdc.gov/h1n1flu/estimates/April%5F March %5F13.htm.

[Gr78] Granovetter M (1978) Threshold models of collective behavior. *American Journal of Sociology*, **83**(6): 1420-1443.

[LF+11] Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd SC (2011) Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. Biosecurity and bioterrorism biodefense strategy practice and science 9: 89-115.

[N09] Nicholls W (2009) Place, networks, space: theorising the geographies of social movements. *Transactions of the Institute of British Geographers*, **34**: 78-93.

[OC+10] Ong JBS, Chen MIC, Cook AR, Lee HC, Lee VJ, et al. (2010) Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. PLoS ONE 5: e10036.