Chapter 18

INTERACTIONS AMONG HUMAN BEHAVIOR, SOCIAL NETWORKS, AND SOCIETAL INFRASTRUCTURES: A CASE STUDY IN COMPUTATIONAL EPIDEMIOLOGY

CHRISTOPHER L. BARRETT

Department of Computer Science and Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA. Email: cbarrett@vbi.vt.edu

KEITH BISSET AND JIANGZHUO CHEN

Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA. Emails: kbisset@vbi.vt.edu, chenj@vbi.vt.edu

STEPHEN EUBANK

Department of Physics, Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA. Email: seubank@vbi.vt.edu

BRYAN LEWIS

Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA. Email: blewis@vbi.vt.edu

V. S. ANIL KUMAR AND MADHAV V. MARATHE

Department of Computer Science and Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA. Emails: akumar@vbi.vt.edu, mmarathe@vbi.vt.edu

S.S. Ravi, S.K. Shukla (eds.), *Fundamental Problems in Computing*,

HENNING S. MORTVEIT
Department of Mathematics and Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, 1880 Pratt Drive, Blacksburg, VA 24061, USA. Email: hmortvei@vbi.vt.edu

Abstract Human behavior, social networks, and the civil infrastructures are closely intertwined. Understanding their co-evolution is critical for designing public policies and decision support for disaster planning. For example, human behaviors and day to day activities of individuals create dense social interactions that are characteristic of modern urban societies. These dense social networks provide a perfect fabric for fast, uncontrolled disease propagation. Conversely, people's behavior in response to public policies and their perception of how the crisis is unfolding as a result of disease outbreak can dramatically alter the normally stable social interactions. Effective planning and response strategies must take these complicated interactions into account. In this chapter, we describe a computer simulation based approach to study these issues using public health and computational epidemiology as an illustrative example. We also formulate game-theoretic and stochastic optimization problems that capture many of the problems that we study empirically.

Keywords: interaction-based computing, theory of simulations, agent-based models, biological, socio-technical and information systems, urban infrastructures, discrete dynamical systems, computational complexity, combinatorial algorithms

1. Introduction

Social networks represent relationships among individual agents. Social networks are not generally static; they evolve over time. Certain aspects of this change arise from structural adaptations such as reciprocity, transitivity, etc. However, changes in social networks also occur as a result of the behavior of individual agents comprising the network. Conversely, individual characteristics and behaviors can depend on the social network to which the agent belongs. For example, it is well known that in many social situations, the behavior of individual agents mimics those of other agents with whom they interact. In other words, individual behaviors and social networks co-evolve. Examples include fashion trends in schools, market practices of firms based on strategies used by successful firms, etc. Social scientists often refer to the change in network structure as *selection* [22–24], and change in individual characteristics as *influence* [17, 23, 24]. See [1, 7, 2, 9, 8, 10, 11, 33, 18] for work done at the interface of game theory, network formation and individual behavior. We also refer the reader to the work of [12, 29, 30] for theoretical

as well as empirical research on the subject of treating selection and influence processes in a network simultaneously.

In this chapter we further motivate and study the joint evolution of selection and influence in social networks in an important application context—spread of infectious diseases. Furthermore, we also consider another component that affects this dynamic—public policy. In classical models used in computational epidemiology, individuals do not adapt their contact behavior during epidemics. For example, they do not endogenously engage in social distancing (protective sequestration) based on disease prevalence. Rather, they simply continue mixing (often uniformly) as if no epidemic were under way. Although potentially a reasonable assumption for non-lethal infections such as the common cold, it is known to fail for lethal diseases such as AIDS. People may be expected to adapt their contact patterns when they perceive a potential threat due to the onset of avian influenza. This will likely result in substantial changes in the social networks that in turn will alter epidemic dynamics. In other words, individual behaviors and the social contact networks that they generate interact and co-evolve. *For brevity we will call the problem of co-evolution of Public policy, Individual behavior and interaction Network as the* **PIN** *problem for the rest of the chapter.*

We begin by describing a computer simulation based approach to study such questions. These simulations use a detailed representation of social contact networks; such a representation is crucial for studying the questions related to co-evolution. We then describe a set of experimental results using our simulations that seeks to analyze these questions in the context of developing public policies for pandemic influenza planning. In the last section of this chapter, we formulate these questions as questions in stochastic optimization and game theory. We hope that these mathematical formulations will serve as starting points for researchers interested in algorithms, operations research and game theory in making further progress in this new and exciting research area.

2. The PIN Problem in Computational Epidemiology

Urban infrastructures have been designed for efficient functioning during normal operations. During crises, however, people's behavior can change so drastically as to render the infrastructure practically useless. Recent blackouts in the Northeast US (2003) and hurricanes such as Rita and Katrnia (2005) demonstrate this amply. In the event of an influenza pandemic, changes in the structure of the social contact network due to behavioral changes are the most important yet difficult to predict factors in determining the spread. The question of how to respond to crises most effectively is very complicated, involving public health systems, regional and urban population dynamics, economic effects, critical infrastructure availability and public policy.

It is well understood that planners must take individual behavior into account when preparing for crises. However, it is not as well appreciated that social responses to public policy can significantly impact the efficacy of public policy and disaster response. Human response, public policies and specific crisis situations are intricately intertwined with one another, making it impossible to obtain a clean simple formal model and solution. Furthermore, policy interventions can have unanticipated consequences due to complex feedback between changing conditions, individual expectations, and social connectivity.

Policy planning has been a central focus of epidemiological research over the years. In addition to empirical observations, practitioners have relied on mathematical models for understanding and comparing different public health policies and making recommendations. These models involve stochastic disease processes on social contact networks. Due to computational considerations, most work in epidemiological modeling has focused on *static* social networks. However, social networks change quite a bit during an epidemic. For instance, policies put in place by public health authorities such as school closures, quarantine, and face masks cause significant changes to the social network. Equally important though is the role of individual behavior in transforming the social network. The recent SARS epidemic (2003) served as an excellent example of how both these factors changed the social network. Thus, mathematical methods for analyzing epidemics based on models of static social contact networks are unlikely to give practical insights into the spread of diseases. We illustrate the issues by two examples.

EXAMPLE 1. First, a simple yet important decision faced by millions of people throughout the country every day during cold and flu season: should I go to work today, even though I have symptoms of a cold or flu? The immediate economic impact of absenteeism due to colds and influenza in the United States in 1980 is estimated to have been $6.5 Billion [31]. While some fraction of these infections arise from exposure outside the workplace, many and perhaps the majority occur because a co-worker decided the consequences of possibly transmitting the disease were less important than the certain consequences of staying home. Indeed, the term *presenteeism* has been coined to describe the problem.

Let us examine the factors involved in this decision more closely. Society pressures us in many ways to go to work even when we may be sick: lack of paid sick leave, need to complete tasks, fear of being seen as a malingerer, desire to be perceived as critical to an organization's success, etc. Personal interactions with co-workers can influence the decision either way. The influence co-workers exert may be tied to whether they have themselves been sick. Furthermore, when a person chooses to stay home, it affects the social network at work in at least two different ways: one is simply the removal of the sick

person as an active influence in decision-making (note that this biases the influence of the remaining people by removing precisely those who would argue for staying home); the other, more subtle, effect is a change in the probability that co-workers will be infected, and thus a possible change in their influence on the decision.

EXAMPLE 2. A second example is the individual decision whether and when to flee in the face of a crisis. As recent mass evacuations have clearly shown, we do not know the best way to clear people out of a city. Much of the uncertainty stems from poor understanding of the effects of individual decisions on the process. How is a person's decision to leave related to official evacuation orders and to decisions made by social contacts? How does it relate to the perceived congestion in the transportation system? How will a household prepare to evacuate and how long will preparations take? What additional demands will be placed on the transportation system as geographically dispersed households gather? How can we take advantage of existing mass transportation resources? How do all these choices depend on timing of an official announcement?

The factors affecting decision making discussed in the above examples, namely, uncertain consequences and conflicting motivations between micro and macro levels for individuals—are at the heart of issues such as noncompliance with public policy and, more generally, breakdown of the rule of law in society. The examples, though complex, are amenable to analysis. By adding features such as public policy decisions and a co-evolving "epidemic" of panic, we can create even more realistic, though inherently more complicated, representations of decision-making with immediate applicability to crisis response and longer-term broader applicability to modeling civil order.

3. Network Based Computational Epidemiology

Computational Epidemiology is the development and use of computer models for the spatio-temporal diffusion of disease through populations. The basic goal of epidemiological modeling is to understand the dynamics of disease spread well enough to control it. Potential interventions for controlling infectious diseases include pharmaceuticals for treatment or prophylaxis, social interventions designed to change transmission rates between individuals, physical barriers to transmission, and eradication of vectors. Efficient use of these interventions requires targeting sub-populations that are on the critical path of disease spread. Computational models can be used to identify those critical sub-populations and to assess the feasibility and effectiveness of proposed interventions.

The spread of infectious diseases depends both on properties of the pathogen and the host. An important factor that greatly influences an outbreak of an infectious disease is the structure of the interaction network across which it

spreads. Descriptive models are useful for estimating properties of the disease, but the structure of the interaction network changes with time and is often affected by the presence of disease and public health interventions. Thus generative models are most often used to study the effects of public health policies on the spread and control of disease.

Aggregate or collective computational epidemiology models often assume that a population is partitioned into a few sub-populations (e.g. by age) with a regular interaction structure within and between sub-populations. The resulting model can typically be expressed as a set of coupled ordinary differential equations. Such models focus on estimating the number of infected individuals as a function of time, and have been useful in understanding population-wide interventions. For example, they can be used to determine the level of immunization required to create herd immunity.

In contrast, disaggregated or individual-based models represent each interaction between individuals, and can thus be used to study critical pathways. Disaggregated models require neither partitions of the population nor assumptions about large scale regularity of interactions; instead, they require detailed estimates of transmissibility between individuals. The resulting model is typically a stochastic finite discrete dynamical system. For more than a few individuals, the state space of possible configurations of the dynamical system is so large that they are best studied using computer simulation.

See [1, 7, 9] for work on use of game theory to study problems in epidemiology. See Kermer [21] for one of the early work on integrating behavioral and epidemiological models; the work however used traditional differential equation based mean field modeling. Recent work by Epstein et al. [13] has used individual based models to study this interaction. Excepting the work of [13], we are not aware of any other work that uses individual agent based models to study the PIN problem in epidemiology.

3.1 SimDemics

SimDemics is a tool for simulating the spread of disease on a social contact network. A brief overview of **SimDemics** is provided here. Further details can be found in [3, 14, 16, 5]. It details the demographic and geographic distributions of disease and provides decision makers with information about (1) the consequences of a biological attack or natural outbreak, (2) the resulting demand for health services, and (3) the feasibility and effectiveness of response options. See [3, 14, 15] for further details. The overall approach followed by disaggregated models consists of the following four steps.

Step 1 creates a synthetic urban population by integrating a variety of databases from commercial and public sources. It yields a set of synthetic individuals and households located geographically, each associated with demographic

variables. Synthetic populations preserve privacy and confidentiality of individuals and yet produces realistic attributes and demographics for the synthetic individuals in the following sense: a census of our synthetic population yields results that are statistically indistinguishable from the original census data, if they are both aggregated to the block group level.

Step 2 creates a synthetic social contact network. This is done by first assigning synthetic individuals a set of activity templates 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. various machine learning and data mining techniques are used for this task. By integrating, this data over all individuals, we get a minute-by-minute schedule of each person's activities and the locations where these activities take place. This information can now be used to synthesize a *time varying 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 points. Synthetic generative methods such as the ones used here are necessary to develop a realistic representation of large urban scale social contact network; such a network cannot be constructed by simply collecting field data.

Step 3 consists of detailed simulation of the epidemic process. The computational model used is called a *graphical probabilistic timed transition system*. The within hosts disease evolution is represented as a probabilistic timed transition system (PTTS). There is one transition system per individual. The state transition of a given PTTS corresponding to an individual depends on its own state, the time, a set of random bits and the state of its neighbors in the dynamic interaction network created in Step 2.

Step 4 consists of representing and analyzing various public policies and interventions using a combination of partially observable Markov decision process (POMDP) and n-way games; these formalisms allow us to capture sequential decision making processes related to interventions and individual behavioral changes in response to disease dynamics. The POMDP is specified succinctly using a co-evolving dynamical system described in the next section. It is thus exponentially larger than the problem specification and is intractable to solve optimally in general. As a result, we use efficient simulations and heuristics to solve the PIN problems. A key concept is that of *implementable policies*—policies or interventions that are implementable in the real world.

SimDemics maintains a parameterized model for the state of health of each person, and updates this continuously based on interaction with other people, and transmission of a disease through these contacts. This enables us to estimate both the geographic and demographic distribution of the disease as a

function of time. It also allows us to evaluate the impact of different intervention policies, such as vaccination and quarantine.

This is an important feature of **SimDemics**. Indeed, the success of most policies and plans depends on their ability to anticipate and adapt to all possible outcomes. However, many of the tools used to describe the range of outcomes and to quantify their relative magnitudes are based on static estimates, whereas in a crisis situation, the responses authorities make depend greatly on real-time situational awareness. **SimDemics** allows the synthetic people to change their behaviors and interactions based on their individual situation as well as characteristics of the entire population.

4. A Mathematical Model to Capture Co-Evolution

We will use a discrete dynamical system framework to capture our co-evolution between disease dynamics and individual behavior. The basic framework consists of the following components: (i) a collection of entities with state values and local rules for state transitions, (ii) an interaction graph capturing the local dependency of an entity on its neighboring entities and (iii) an update sequence or schedule such that the causality in the system is represented by the composition of local mappings.

We formalize this as follows. A **Co-evolving Graphical Discrete Dynamical System** (CGDDS) $\mathcal{S}$ over a given domain $\mathbb{D}$ of state values is a triple $(G, \mathcal{F}, W)$, whose components are as follows:

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

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

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

the vertex v_i is to be updated; $v_i(g)$ specifies that one or more incident edges will be updated.

From a modeling perspective each vertex represents an agent. Here we will assume that the states of the agent come from a finite domain $\mathbb{D}$. The maps $f_{v_i,j}$ are generally stochastic.

Computationally, each step of a CGDDS (i.e., the transition from one configuration to another), involves updating either a state associated with a vertex or modifying the set of incident edges on it. The following pseudo-code shows the computations involved in one transition.

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

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

We make several observations regarding the formal model described above.

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

2. The edge modification function as defined can modify in one step a subset of edges simultaneously. An alternate model could have been where a vertex is allowed to change exactly one edge at a time. We have chosen the former due to the specific application in mind. In all our applications,

when an agent decided to not go to a location (either due to location closure as demanded by public policy or due to the fear of contracting the disease) its edges to all other individuals in that location are simultaneously removed while adding edges to all the individuals who might be at home.

3. The model is *Markovian* in that the updates are based only on the current state of the system; it is possible to extend the model wherein updates are based on earlier state of the system.

4. We have assumed that there is exactly one function for each arity for each node. This can be relaxed easily, similarly these functions will, in general be stochastic.

4.1 Specifying PIN Problems in CE Using Co-Evolving Discrete Dynamical Systems

We briefly outline how PIN problems in Computational Epidemiology can be specified using CGDDS. In all the situations considered in this paper, we can make certain simplifying assumptions due to the specific dynamics that we consider. In **SimDemics**, we have a notion of a **day**. A day is typically 24 hours but can be smaller depending on the specific disease. We *assume* that the social contact network does not change in the course of a day. This is a realistic assumption due to the time scale of disease evolution (time it takes for a person to be infectious or symptomatic after being infected). As a result, the schedule can be specified as a sequence of days wherein we only consider disease dynamics over the entire population followed by a step in which there is a change in the social contact network.

We can make this a bit more precise as follows: We denote the functional modules for mobility, disease propagation and activity generation by M, D and A, respectively; these are described in Appendix. Each individual is assigned a set of initial activities based on their preferences, demographics, and infrastructure constraints in the activity assignment module A. The module M assigns locations to all entities based on the current set of activities which in turn induces the current contact graph, or social network. Using the contact graph, the module D computes the next stage which is disease dynamics. This corresponds to updating the disease state of every individual in the network over one day. The activity generator A uses the current disease state to update the current activities. Models of individual behavior or policy that affects individual behavior constitutes this module.

In general, the dynamics is time dependent and is generated by iteration of the composed map F given by

$$F = D \circ M \circ A.$$

This is illustrated on the right in Fig. 18.1. Notice that this is already a substantial simplification over all possible choices for the string W.

Interventions and behavioral changes can be broadly categorized based when they occur:

1. *Non-Adaptive*: Non-adaptive interventions and behavioral changes occur before the start of the simulations. The non-adaptive interventions unrealistically assume the population does not change during the course of the epidemic and is limited to studying treatments that have a permanent effect, like vaccination.

   Letting the initial state of the system be $\mathbf{x}_0$, the final state of the system can be written as $\mathbf{x}(t) = F^t(\mathbf{x}(0))$ as $(D^t \circ (M \circ A))(\mathbf{x}(0))$, illustrated on the left in Fig. 18.1.

2. *Adaptive*: The adaptive strategies on the other hand, incorporate changes in the movement of the people, treatments that have only temporary effects (antiviral medications are only effective when being taken), and wholesale changes to the interactions within the population (like school closure). This is represented most generally as $\mathbf{x}(t) = F^t(\mathbf{x}(0))$ as $(D \circ M \circ A)^t(\mathbf{x}(0))$. We can now differentiate various strategies by how frequently M and A are applied as compared to D. In other words, we view the dynamics as the following composition: $(D^{t/r} \circ M \circ A)^r(\mathbf{x}(0))$, where the exponents reflect the different time scales. This can be viewed as degree of adaptation. Policy based change in the social network is usually caused by changing the behavior of a set of individuals in some uniform way. Furthermore, it is natural to expect that these changes do not occur often. Individual behavior based changes on the other hand can occur every day—individuals can change their behavior and thus their probability of contracting a disease on a daily basis. A simulation is computationally most efficient when t is small, since it amounts to fewer updates to the social network and individual behavior. On the other hand, making t small makes the simulation less realistic since the

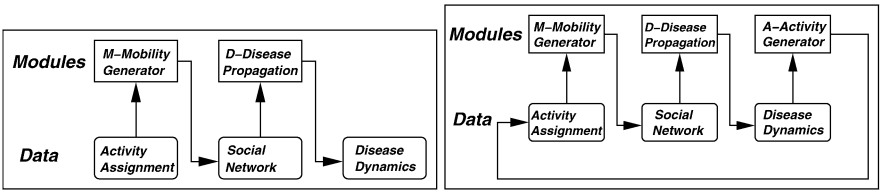

Figure 18.1. The left diagram shows the data flow for disease dynamics without feedback, that is, where e.g. interventions do not alter activities. The diagram on the right shows data flow with feedback from the disease dynamics to the activities

interaction between individual behavior and disease dynamics is not well represented.

5. Computational Experiments

This section illustrates how complicated PIN problems can be specified and studied using computational models such as **SimDemics**. See [6] for additional details. We will compare the effectiveness of both adaptive and non-adaptive interventions on the same population with the same contact network, using disease models of the same disease (pandemic strain of influenza). Non-adaptive interventions are done before the epidemic starts—in this setting, we (unrealistically) assume that the activities of all people are unchanged during the experiment. Adaptive interventions, on the other hand, are done based on the information available about the epidemic, and can change as the epidemic proceeds. The interventions we will consider include medical (such as administration of vaccines and anti-virals), governmental (such as school closures), and societal (such as social distancing)—some of these interventions are external, and some are endogenous, i.e., people themselves implement them.

These computational experiments show the following:

- They illustrate the qualitative differences between adaptive and non-adaptive strategies and highlight the need for more realistic dynamic modeling.

- They illustrate the power of **SimDemics** modeling system in terms of (i) its ability to handle various kinds of adaptive and non-adaptive interventions, (ii) handle large instances.

5.1 Basic Experimental Setup

The contact network we study models a population of about 8.86 million people in Chicago. The network is constructed by synthesizing information from a number of different sources [4]. We model pandemic influenza with all the characteristics of normal influenza, with a much higher transmissibility. Influenza has a short incubation period, can be infectious even in the absence of symptoms, and is transmitted through the air or by certain kinds of contact.

The *heterogeneous* symptomatic and incubation periods are drawn from a distribution, and are fixed for every person initially. The transmissibility, or the probability of infection on a contact, per minute is chosen to be 0.000048 and 0.0003. The number of initial infections is 4. The disease model in Experiment 2 differs from the one in Experiment 1 in by incorporation of additional states needed to capture the effects of the antiviral treatment, but the gross features are still the same.

Parameter	Values
Social network	Chicago, 8.86M individuals
Transmissibility (τ)	0.000048 and 0.0003
Age groups	0–5 (group 1), 6–15 (2), 16–20 (3), 21–60 (4), >60 (5)
Number of people intervened	50K, 100K, 150K, 200K, 250K, 300K, 400K, 500K
Number of initial infections	4
Number of iterations	50, 2 initial infection sets, 25 iterations per set
Policies	random, high degree, high vulnerability, household with specific activity types, specific age groups

Table 18.1. Summary of parameters used in experimental studies

We describe below the specific experiments we perform and the various experimental parameters.

- We choose two values for the transmissibility parameter τ, namely $\tau = 0.000048$ and $\tau = 0.0003$.
- We choose 25 different sets of initial infections and run 2 random iterations for each of them, for a total of 50 iterations. We then compute an *average run*, where the number of new infections on each day is the average of the new infection number on this day in the 50 iterations, and report the measures based on the average runs.
- For each vaccination policy, we consider the following sizes (where K means thousand): 50K, 100K, 150K, 200K, 250K, 300K, 400K, 500K.
- For random people, we choose a subset of given size from the population uniformly at random. This trivial vaccination scheme can be viewed as a benchmark for evaluating effectiveness of other vaccination schemes.
- An individual is *active* if his/her activities belong to many types, or s/he lives in the same household with an active individual. The list of active people is determined from the given contact network and has about 500K people.
- We have five age groups: 0–5 years in age group 1; 6–15 years in age group 2; 16–20 years in age group 3; 21–60 years in age group 4; older than 60 years in age group 5. We are especially interested in age groups 2 and 5, i.e., school kids and seniors.

5.2 Experiment 1

Non-Adaptive Interventions: Study the effect of pre-vaccination of specific sub-populations assuming no changes in behavior throughout the course of the

epidemic. Here we will compare vaccination policies targeting the following sub-populations:

- randomly chosen people
- people of high degree
- people of high vulnerability
- active people
- people of a specific age groups

While many other policies and groups can be explored, even in this static case, these groups are chosen to illustrate a sample of the types of policies that can be represented in this modeling environment. We measure the effectiveness of these policies in terms of the percentage decrease in the epidemic size as compared to the unmitigated case as well as the unit efficiency. We will need some notation in order to define these measures formally. For subset $A \subseteq V$ of people, we let $\mathcal{I}_A(G)$ denote the set of infected people, when the people in A are immunized, subject to some specific starting conditions, and disease model in the contact network $G(V, E)$. Mathematically, vaccinating a person is equivalent to either removing a node from the network, or reducing its incident infection probabilities. Note that $A = \emptyset$ means no vaccination, i.e., *base case*. The two measures we use to compare different policies are:

- the *percentage decrease in epidemic size*, defined as:

$$\text{DES} = \frac{|\mathcal{I}_\emptyset| - |\mathcal{I}_A|}{|\mathcal{I}_\emptyset|}$$

- the *unit efficiency of vaccination*, defined as:

$$\text{UE} = \frac{|\mathcal{I}_\emptyset| - |\mathcal{I}_A|}{|A|}$$

Results and Analysis: The most basic question is which policy is the most effective for a given disease. We are also interested in finding a policy that is easy to implement from a public health point of view. These policies would be compared empirically in the sections that follow.

The two measures (DES and UE) are plotted against vaccination size in Figs. 18.2 and 18.4 for the case $\tau = 0.000048$ and in Figs. 18.3 and 18.5 for the case $\tau = 0.0003$.

The effectiveness of vaccination is highly dependent on who is selected for vaccination and what the transmissibility of the disease is. When the disease has high transmissibility ($\tau = 0.0003$) vaccination policies have little effect

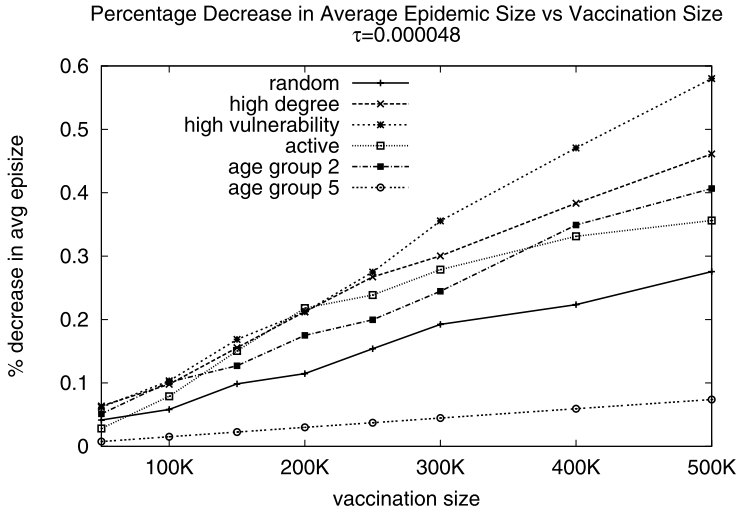

Figure 18.2. Percentage of decrease in average epidemic size ($\tau = 0.000048$)

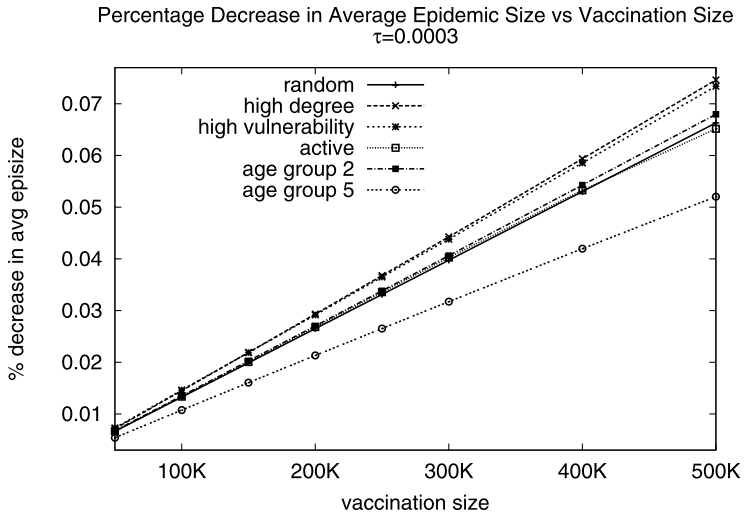

Figure 18.3. Percentage of decrease in average epidemic size ($\tau = 0.0003$)

(Fig. 18.3): even if half a million vaccinations are given (5.6% of the population) there is only a 7% decrease in epidemic size (6.3% of the population). If these vaccines were not randomly assigned, but specifically given to people older than 60 then they are even less effective, only decreasing the epidemic size by 5.1% which is even lower than the vaccination percentage

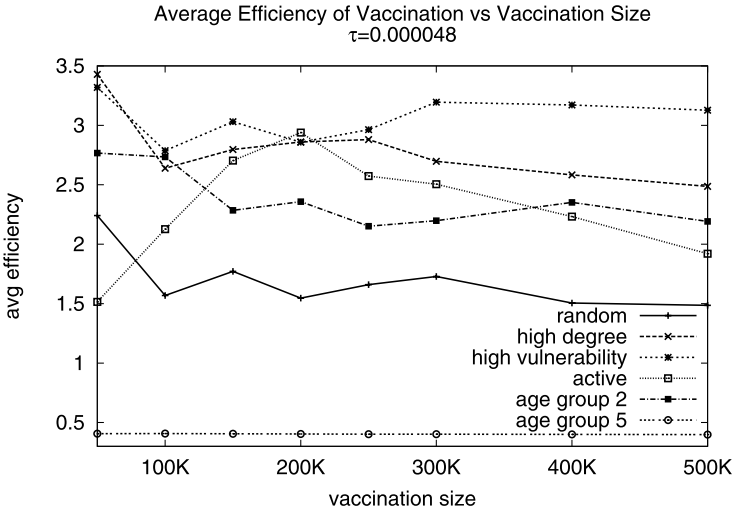

Figure 18.4. Average unit efficiency of vaccination ($\tau = 0.000048$)

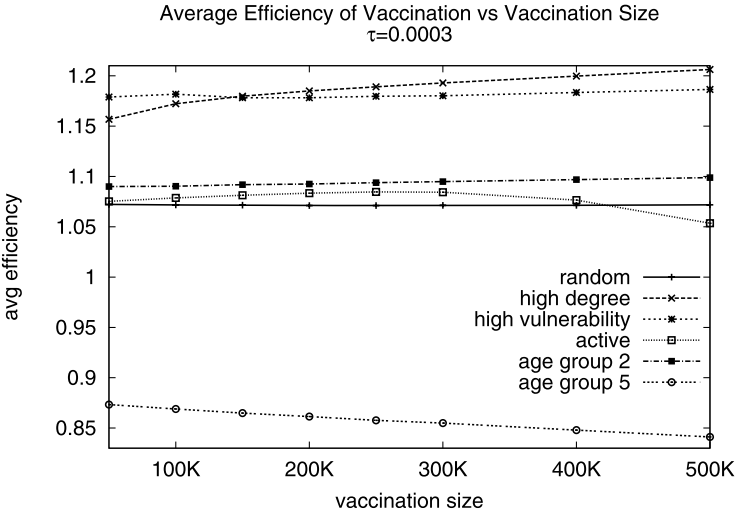

Figure 18.5. Average unit efficiency of vaccination ($\tau = 0.0003$)

(5.6%). The effect is only slightly greater for less transmissible diseases: with $\tau = 0.000048$ the epidemic is decreased by 8%. The limited effect of vaccinating those over 60 is a result of the low connectivity of this population. They are more susceptible to severe effects of the disease, however, so vaccination ensures lower mortality, which was not considered in these simulations.

To further illustrate the impact of who is vaccinated, note that schemes where "high degree" individuals are vaccinated result in significant decrease of epidemic size. For instance, in Fig. 18.2 ($\tau = 0.000048$) when half a million "high degree" individuals are vaccinated the epidemic size is decreased by more than 45%. Furthermore, if the "high vulnerability" individuals are vaccinated the epidemic decreases by almost 60%. Identifying these individuals requires complete knowledge of the contact network and in the case of vulnerability, requires previous simulations and analysis. It is more realistic to identify people that are in high risk age groups or have behaviors that might put them at higher risk. Figure 18.4, shows that vaccinating school children aged 6–15 is much more efficient than simply vaccinating random individuals. Additionally, *active* individuals, who engage in many types of activities (i.e. school, work, and/or college) or live with an individual with these activities, may be relatively easy to identify. Vaccinating these individuals is also shown to be more efficient than random vaccination.

The effectiveness of vaccinating "high degree" and "high vulnerability" people stands to reason given the significant reduction of overall degree that their removal would produce. However, these strategies require perfect knowledge of the contact network. Additionally, note that the effectiveness of the strategies is based on the assumption that the social network does not change. Individuals that are high degree before the arrival of an epidemic disease, may not have high degree under epidemic conditions. The evolution of the contact network under these conditions can also cause individuals that might not be obviously high degree to become more highly connected (for instance health care workers).

Nevertheless, even assuming a fixed contact graph, significant insights into the effectiveness of various vaccination schemes can still be made. For instance, the effectiveness of vaccinating high degree individuals suggests that it might be useful to identify individuals in a specific age groups or individuals carrying out specific trade (e.g. emergency care workers) as potential targets.

5.3 Experiment 2

Adaptive Interventions: Study the effects of dynamic changes to the social network, treatments with antivirals, and changes in individual behaviors throughout the course of an influenza epidemic. An effective vaccine for pandemic influenza is not likely to be available until the pandemic is well established. Currently available antiviral medicines used for treatment of influenza have limited efficacy in preventing infection and are likely to be in short supply. Without these tools, control of an influenza pandemic must be attempted through more general infection control measures. This experiment studies the effectiveness of a collection of interventions both together and in isolation as

well as the sensitivity of when they are implemented. The interventions are dynamically triggered at different points in the epidemic and the timing of these triggers is also studied. The interventions are designed to reduce the opportunities for infections by removing infectious people from circulation, reducing their infectivity through treatment, and keeping potentially infectious people from transmitting disease before they develop symptoms. These interventions drastically alter the daily activities of many of the people in the simulation, and these dynamic changes can effectively control the epidemic.

Experimental Setup: As mentioned earlier, the same population with the same contact network are exposed to the same disease modeled on a highly infectious influenza, as was done in Experiment 1. However, the interventions modeled are very different. They are derived from interventions recommended in federal pandemic planning documents[1] and require that they be dynamically applied under conditions specific to the individual. The modeling environment is designed to accommodate these kinds of interventions, and thus allow the simulation to closely represent what might actually occur in reality.

The specific interventions we will consider are:

1. Case isolation: once an individual experiences symptoms of the disease, they remain home through the duration of their illness.

2. Case treatment and household quarantine: if a case is diagnosed, they are administered anti-viral medications (reduces their infectivity and duration of illness) and all household members are given prophylactic anti-viral medications (reduces their chance of infection) and are quarantined at home until no one in the household is sick.

3. School closure: all schools are closed, some children remain at home while the remaining substitute other activities during normal school hours. An adult in the household of a young child (less than 15) must stay home to supervise them.

4. General social distancing: 50% of people eliminate all non-essential activities (shopping, visiting, recreation).

5. Workplace social distancing: to reduce workplace exposure, workers in large offices interact with 50% co-workers.

These interventions were studied across different levels of adherence to the interventions (30%, 60%, and 90%) and were implemented at different points in the progress of the epidemic (from 0.0001% of the population to 10% of the

[1] See http://www.whitehouse.gov/homeland/pandemic-influenza.html.

Prevalence trigger	Cumulative proportion ill
Never	44.7%
10%	20.3%
1%	3.9%
0.10%	2.0%
0.01%	1.7%
0.001%	1.7%
0.0001%	1.7%

Table 18.2. Epidemic size decreases when the interventions are implemented at lower prevalence thresholds

Compliance	Early threshold (0.01%)	Later threshold (0.1%)
30%	1.7%	2.0%
60%	0.1%	1.3%
90%	0.1%	1.2%

Table 18.3. Epidemic size decreases when societal compliance with interventions increases

population infected, or 9 cases to 886,000 cases). All permutations were not studied due to limits on computational resources.

Results and Analysis: The modeled disease epidemic can be completely controlled by the adaptive interventions. The overall magnitude is significantly curtailed when the interventions are triggered at a lower level of disease prevalence. Similarly, when societal compliance increases the size of the epidemic decreases.

The size of the epidemic is very sensitive to when the interventions are instituted (Table 18.2). An uncontrolled epidemic, i.e., when the interventions are never implemented, leads to nearly half the population becoming ill. Even if the interventions are not applied until after the epidemic has made 10% of the population ill, the interventions are able to prevent half of these infections. Interestingly, there is a limit to how effective the interventions can be, even if implemented at levels of infection in the population that would be impossible to detect (0.01% to 0.0001%) they cannot completely prevent the epidemic. While the overall attack rate may be the same, note the difference in the timing and shape of the epidemic (see tables below). The epidemic that follows the interventions triggered at 0.01% peaks nearly three weeks earlier but has the same area under the curve, which could translate into other changes in the population were there further adaptive measures in place.

The levels of compliance with the interventions also have an effect on the size of the epidemic, though less so than the timing of the intervention (Table 18.3, Fig. 18.6). Similarly, at the extremes of the control (both 60% and 90%) the overall attack rates are limited to the same level, but shape of the

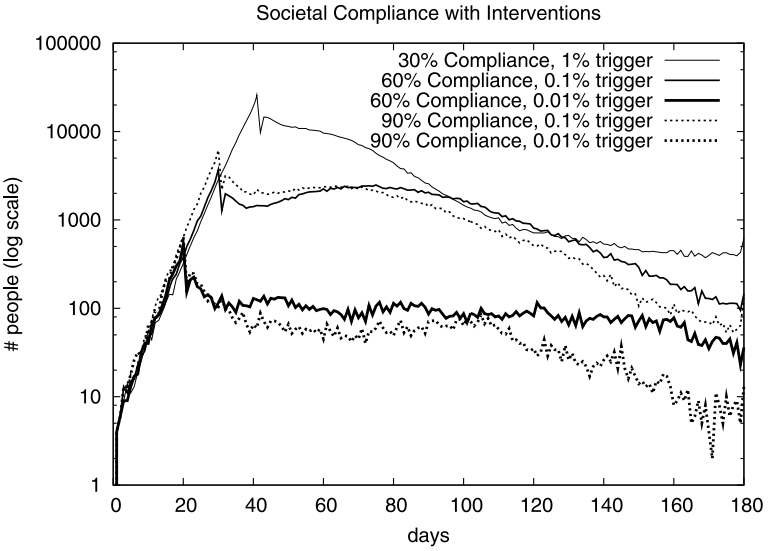

Figure 18.6. Epidemic curves by levels of compliance and time of intervention: Scenario 2—30% compliance and trigger at 1% prevalence; Scenario 3—60% compliance and trigger at 0.1% prevalence; Scenario 4—60% compliance and trigger at 0.01% prevalence; Scenario 5—90% compliance and trigger at 0.1% prevalence; Scenario 6—90% compliance and trigger at 0.01% prevalence

epidemic curves are different, which in turn could have an effect on additional adaptive measures.

5.4 Comparing Adaptive and Non-Adaptive Strategies

The two case studies above tell different stories to Public Health policy makers, with varying levels of refinement. The non-adaptive strategies studied on the static network can be useful for informing permanent modifications to the potential disease transmission network, such as vaccination. These approaches could determine which groups are best suited for vaccination when supply is limited, or could be used for planning how many vaccines are needed to control an epidemic. However, they can not answer questions about behavior modifications based on an individual's state. The adaptive strategies studied using the **SimDemics** modeling environment are designed to handle these exact types of dynamic changes to the social network. This more flexible architecture allows the exploration of a wider range of public health policy options, and can reproduce behaviors in the system that may not be obvious. The transparency in the representation of the framework also allows for a more direct interpretation of the results, which allows for greater understanding across a wider audience of policy makers. The framework still requires some coarse adjustments

based more on time and average behavior of the model, and full situational awareness. Further refinement of these adaptive strategies is needed so that the implementation of all these strategies is more fluid and evolving.

6. A Mathematical Formulation

We have seen how one can use computer simulations to study the effects of various adaptive and non-adaptive interventions to control the spread of avian flu through a social network. In this section, we will try to formulate many of these questions as combinatorial questions in stochastic optimization, game theory, dynamical systems and algorithms. This serves to expose the reader to various mathematical formalisms, each capturing a different facet of the underlying problem. Nevertheless, our primary goal is algorithmic here—we concentrate on the algorithmic issues arising in these formalisms. Often the questions are based on a simplified mathematical abstraction of the realistic situation; nevertheless, we believe that this allows us to formulate questions that might be tractable in the sense of obtaining rigorous mathematical proofs. Progress on these questions will help us understand and guide simulation based experimental results.

6.1 Preliminaries: A Simplified Model

Let V denote a population. We refer to individuals in V as nodes. Let $G(V, E)$ denote a contact graph on this population—each edge $e = (u, v) \in E$ denotes that the individuals u and v come into contact and can infect each other. The spread of infection is assumed to be a stochastic process. For each edge $e = (u, v) \in E$ let $r(e)$ (also, sometimes denoted by $r(u, v)$) denote the probability of the infection spreading from u to v per unit time—this is sometimes referred to as the infection rate. Let $\tau(u)$ denote the time that node u remains infected. Note that the infection rates need not be symmetric, i.e., $r(u, v)$ and $r(v, u)$ need not be the same. We will assume that $r(u, v)$ does not vary with time, though this happens in reality. Most disease models have additional states. For instance, there is an *incubation period*, which is the period right after the infection, in which the individual is infected, but not yet contagious. Let $I(u)$ denote the incubation period for node u. We let $\bar{r}$, $\bar{\tau}$ and $\bar{I}$ denote the vectors specifying the above quantities for all nodes and edges. We will use $\bar{x}$ to denote the initial conditions: $x(v)$ denotes the probability that v is infected initially.

We will be considering discrete time models for epidemics, where the probability that node v does not get infected by node u in t time steps after u got infected is given by:

$$\Pr[\text{node } v \text{ not infected}] = (1 - r(u, v))^t \tag{6.1}$$

In an epidemic model such as SIR, each node u recovers and becomes immune $\tau(u)$ time steps after becoming infected. In endemic models such as SIS, node u returns to the susceptible state after this time. A crucial assumption made in almost all epidemic models is that of *independence*: we assume that the spread of infection from a node u to node v is completely independent of the infection from a node u' to node v. Similarly, an infected node u spreads the infection to each neighbor v, independent of the other neighbors of u. This is a central assumption in almost all the epidemic models and the analytical results based on percolation. However, there exist other epidemic models, such as the *Descending Cascade Model* [20], in which this independence assumption does not hold.

6.2 Policy Planning Problems

We begin by formulating one of the policy planning problems studied earlier empirically—determining whom to vaccinate—as a stochastic optimization problems. The optimization issue arises because of limited resources, e.g., of vaccines—this raises the question of whom to vaccinate so that the "public good" is maximized. However, public good can be defined in a number of ways, and therefore, there is no unique solution. In this section, we will take an easy route by just attempting to determine a policy that minimizes the epidemic size. This gives us the following problem, which we call the Vaccination Problem, following our earlier results in [16], which we denote by $\mathsf{VP}(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x}, k)$:

- **Given:** Contact graph $G(V, E)$, which is directed, an SIR disease model, as described in Sect. 6.1, which is specified by the vectors $\bar{r}$, $\bar{\tau}$, $\bar{I}$, and a parameter k, and a vector $\bar{x} \in [0, 1]^n$, which describes the initial conditions—$x(v)$ denotes the probability that node v is infected initially. The most common starting conditions are: (i) there is a single node v such that $x(v) = 1$ and $x(w) = 0$ for all $w \neq v$, or (ii) $x(v) = 1/n$ for each v.
- **Objective:** Choose $S \subseteq V$, $|S| \leq k$ so that the number of nodes infected when the disease is run on $G[V \backslash S]$ is minimized. In the initial conditions where some specific nodes are infected, none of them should be in the set S.

The SIR model leads to several simplifications in the formulation of the above problem, and relates it to percolation. First, the incubation period $I(u)$ of node u plays no role in the expected epidemic size. Also, the above formulation does not care for the temporal aspects, and so it suffices to simply consider the effective infection probability on edge $e = (u, v)$ as $r'(e) = 1 - (1 - r(e))^{\tau(u)}$. Let $G(r')$ denote a random subgraph of G in which each

edge e is retained with probability $r'(e)$. Also, consider a simple initial condition $\bar{x}$ in which there is a single node s with $x(s) = 1$ and $x(v) = 0$ for all $v \neq s$. Thus, the $\mathsf{VP}(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x}, k)$ problem can be restated as:

- Choose a subset $S \subseteq V$ with $|S| \leq k$ such that:
- The expected number of nodes reachable from s in the (random) subgraph $G(r')$ is minimized—the expectation here is over the random subgraphs $G(r')$. If the initial condition $\bar{x}$ is different, the expectation above would also be over different choices of initial sets, by sampling from $\bar{x}$.

The above formulation is the simplest possible one, but is already non-trivial. It remains non-trivial even if we consider the simplest possible disease model in which $r(e) = 1$ for each edge (modeling a "highly infectious disease"), as the following result from [16] shows:

THEOREM 18.1 [16]. *$\mathsf{VP}(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x}, k)$ is NP complete if $r(e) = 1$ for each e, and there is a node s such that $x(s) = 1$ and $x(v) = 0$ for all $v \neq s$. For any $\epsilon > 0$, there is a polynomial time bi-criteria approximation algorithm that deletes a set S of $O((1+\epsilon)k)$ nodes, so that the number of nodes reachable from s in $G[V \setminus S]$ is $O((1 + 1/\epsilon)OPT)$, where OPT denotes the optimum solution to this problem.*

The complexity of the $\mathsf{VP}(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x}, k)$ for more realistic disease models (i.e., when $r(e) < 1$) is likely to be $\#\mathsf{P}$-hard, and determining this remains an open problem.

Adaptive Policies: The VP problem described above corresponds to a *non-adaptive* vaccination policy. Using the stochastic optimization framework developed by [28, 32, 19], we can formulate an *adaptive* version of this problem, which we call Adaptive Vaccination Problem (AVP). In this formulation, the nodes to be vaccinated, or deleted do not have to be chosen in one shot. Instead, a feasible solution corresponds to choosing set S_i at the start of the ith time step. As in [28], we assume that there is an inflation factor σ_i in step i, so that the cost of choosing set S_i in step i is $\Pi_{j \leq i} \sigma_j |S_i|$; following [28], we also assume that $\sigma_i \geq 1$. The $\mathsf{AVP}(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x}, k, \bar{\sigma})$ problem is defined in the following manner:

- The quantities G, $\bar{r}$, $\bar{\tau}$, $\bar{I}$ and $\bar{x}$ are defined as before. The parameter k denotes the total cost that feasible solution must have, and $\bar{\sigma}$ specifies the inflation factor.
- **Feasible Solution:** This is a sequence of disjoint sets $S_1, S_2, \ldots, S_\ell$. The set S_i denotes the set of nodes to be vaccinated on the ith timestep.

The set S_i can be chosen after observing the state of the epidemic in the ith step.

- **Objective:** Choose a feasible solution $S_1, \ldots, S_\ell$ such that the inequality $\sum_i (\Pi_{j \leq i} \sigma_j)|S_i| \leq k$ holds and the expected number of infected nodes is minimized.

In reality, only *partial* information is known about the epidemic reliably at each step, and the AVP problem above can be easily modified to incorporate this aspect.

The 2-person Vaccination Policy Game: We now consider a variant of the AVP problem as a 2-person game. One player is the policy maker who has to choose the vaccination policy, and the second player is "nature", which decides on the spread of the epidemic, following the framework of *Games against Nature* (*GAN*) [27].

We denote this game as VPG$(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x}, k, \ell, M)$. Let P denote the single player, and let N denote nature. The game runs in rounds with P and N playing alternately. N plays first, and infects nodes according to the starting condition $\bar{x}$, i.e., each node v is infected with probability $x(v)$, independently of other nodes. Then, P plays, and it can decide to vaccinate (or delete) up to ℓ nodes. In the next round, N plays, and spreads the infection to the unvaccinated neighbors of the infected nodes, according to the disease model specified by $\bar{r}$, $\bar{\tau}$ and $\bar{I}$. Let S_i denote the set of nodes chosen to be vaccinated by P; we must have $\sum_i |S_i| \leq k$. The goal is to decide whether there is a vaccination strategy for P, specified by the sequence of sets $S_1, S_2, \ldots$, such that $\sum_i |S_i| \leq k$ and the total number of infected nodes is at most M. Is this problem PSPACE complete, as some of the other GAN problems are?

6.3 Individual Behavior Problems: A Game Theoretic/ Dynamical Systems Viewpoint

A common problem with implementation of policies is compliance. This is especially true in the case of vaccinations, which may have side effects and involve additional costs, and in the case of directives to "stay home", might simply be infeasible. Incentives are needed to make people comply. An interesting way to give an incentive could be to enter all the people who get vaccinated into a lottery—such schemes have also been studied in other settings, such as voting. This scenario immediately leads to interesting game theoretic questions, since each individual now has a set of conflicting costs and rewards, and has to make a choice that would optimize his or her perceived utility. There are several papers that study game theoretic questions [1, 7, 9] related to epidemics on networks. However, these results either assume that the graph is very simple, or that the disease model is very simple. The approaches in [7, 9] use differ-

ential equations and mean field approximations to formulate realistic disease models on complete mixing networks (cliques). The paper by Aspnes, Chang and Yampolskiy [1] is much closer to our models, in the sense that the network is general, but the disease model is simple and assumes a "highly contagious disease". Extensions of this game have been studied in [26, 25]. There are, admittedly, several difficulties with these non-cooperative formulations, e.g., it is hard for nodes to compute their utility functions, and there is no persuasive reason for equilibria to exist. However, the structure of these games may give useful insights into their dynamics. We also give equivalent dynamical system formulations of these games.

The Vaccination Game (VG) This game is denoted by $\mathsf{VG}(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x})$, and is defined in the following manner. Each node corresponds to a player, and a strategy for player v is denoted by a quantity $a_v \in [0, 1]$, which is the probability that node v decides to get vaccinated; vaccinating a node is equivalent to lowering the infection probabilities on all edges incident on v. The disease model is specified by $\bar{I}$, $\bar{r}$ and $\bar{\tau}$, and $\bar{x}$ gives the initial conditions, as discussed earlier. We formulate the utility function U_v for node v as

$$U_v = a_v C + \Pr[v \text{ gets infected}]L,$$

where C denotes the cost (or reward) of getting vaccinated, and L denotes the cost of getting infected. The probability that node v gets infected is defined over the initial condition $\bar{x}$ and the strategy $\bar{a}$.

One of the main problems of interest is to study the structure of equilibria, if they exist, and compare their cost to that of a social optimum. Aspnes et al. [1] consider a simple disease model, in which a node gets infected if there is a path to it from an infected node, and the disease can start initially at any node, i.e., $x(v) = 1/n$ for each v. For illustration, consider a pure strategy $\bar{a}$. Suppose $a_v = 0$ for some node v. Then $\Pr[v \text{ gets infected}]$ is proportional to the size of the component containing v, after all the nodes w with $a_w = 1$ are deleted. This is illustrated in Fig. 18.7. For this model, Aspnes et al. [1] show that pure Nash equilibria always exist, and can be completely characterized in terms of the quantity $t = Cn/L$—a strategy profile $\vec{a}$ is a Nash equilibrium provided:

1. every component in $G_{\vec{a}}$ has size at most t, and
2. flipping the strategy of a node v from 1 to 0 gives a component of size strictly greater than t.

They also show that computing Nash equilibria that have minimum total cost is NP-complete, but a simple switching strategy always converges to an equilibrium. Finally, the cost of the worst Nash equilibrium can be $\Theta(n)$ times

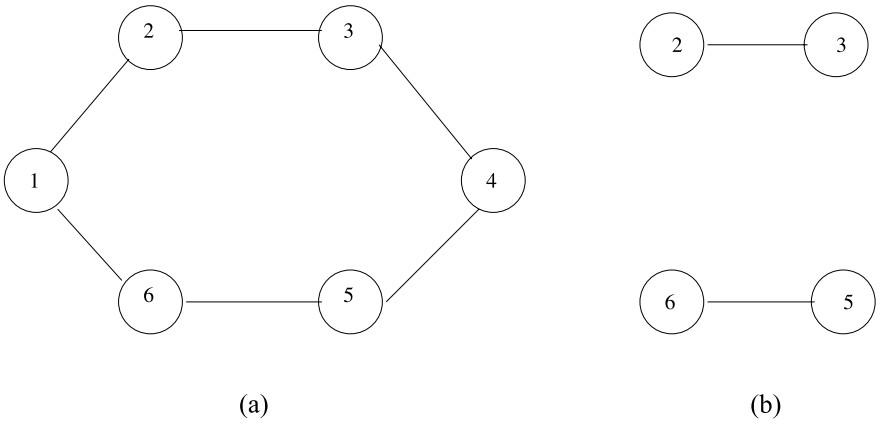

Figure 18.7. (a) A sample contact graph. (b) The components resulting from the strategy $\bar{a}$ with $a_1 = a_4 = 1$ and the rest being 0. The probability that any of the nodes $2, 3, 5, 6$ gets infected is $1/2$

the social optimum. Extending these results to more general disease models remains open questions.

The Multi Stage Policy Game (MSPG): While the above questions are mathematically interesting, an inherent difficulty with the above model is that it is hard for individuals to estimate their costs. In light of this, we will consider the following multi-stage version of this problem. We call this the Multi Stage Policy Game (MSPG), and it is denoted by MSPG$(G, \bar{r}, \bar{\tau}, \bar{I}, \bar{x})$. In this game, the strategy a_v of player v is actually a vector, and $a_v(i)$ denotes the probability that v stays home—people find it much easier to decide whether to stay home or not, when an outbreak has started, than deciding the utility of getting vaccinated. Node v can choose $a_v(i)$ depending on how many nodes in its neighborhood are infected. The main objective would be to study this as a dynamical system and explore its limit distribution, and the parameters that influence these distributions.

Preliminary empirical results related to this problem can be found in [13]. A simpler variation of the above based on differential equations was formulated and studied in [21]. The model proposed here is more general and network based which makes the problem substantially harder. The MSPG and VPG games are instances of anti-coordination games [8]—it is in the interest of a player to get vaccinated or stay home if a lot of people around her are not doing so. The realism in the contact network and disease model make this a very rich problem area. In addition to the limiting distributions of these games, the computational complexity of these problems is an interesting problem.

6.4 Discussion of the Different Formulations

The different theoretical problems formulated above deal with specific aspects of epidemic processes and policy planning—the different variants highlight the richness of this area, and the limitation of any single theoretical model to capture all of its complexity. The computational complexity of these problems is in general an open question.

As an example, we consider the VP and AVP problems and their computational variants. The complexity of computing the expected number of infected nodes, $N_{\inf}$, for initial conditions $\bar{x}$, when set S_i of nodes is vaccinated at step i, is not exactly known. As mentioned in [20], it is not known how to compute this quantity, even when all the nodes to be vaccinated are chosen in step 1 itself, though it is a reasonable conjecture that this problem is #P-hard. An (ϵ, δ) approximation to $N_{\inf}$ can be computed by a simple sampling scheme:

---

1. **For** $i = 1$ to t **do**
   (a) Generate a random instance of the bond percolation process, by retaining each edge with probability $r'(e)$, as defined earlier in the discussion of the AVP problem.
   (b) Let Z_i denote the number of nodes reachable from the initial infected nodes, specified by $\bar{x}$.
2. **Output** $Z = \frac{Z_1 + \cdots + Z_t}{t}$

LEMMA 18.2. *For $t \geq n^2/\epsilon^2\delta$ we get an (ϵ, δ) approximation to $E[N_{\inf}]$, i.e.,*

$$\Pr[|Z - E[N_{\inf}]| > \epsilon E[N_{\inf}]] \leq \delta$$

Proof. Clearly, $E[Z] = E[Z_i] = E[N_{\inf}]$ for each $i = 1, \ldots, t$. Therefore, by Chebyshev's inequality, we have

$$\begin{aligned}\Pr[|Z - E[Z]| \geq \epsilon E[Z]] &\leq \frac{var(Z)}{\epsilon^2 E[Z]^2} \leq \frac{var(Z_1)}{t\epsilon^2 E[Z_1]^2} \\ &\leq \frac{E[Z_1^2]}{t\epsilon^2 E[Z_1]^2} \leq \frac{n^2}{t\epsilon^2} \leq \delta\end{aligned}$$

for $t \geq n^2/(\epsilon^2\delta)$. □

The above sampling works only because $N_{\inf}$ takes integral values in the range $\{1, \ldots, n\}$. It would not immediately work for other problems, such as determining the probability that a node v gets infected.

An interesting question is whether the AVP and VPG problems are PSPACE complete. In light of the above discussion of an (ϵ, δ) sampling for $N_{\inf}$, is it possible that reasonable polynomial time approximation algorithms exist for these problems?

7. Concluding Remarks

We have described an agent based modeling approach to study the interaction between public policy, individual behavior and spread of infectious disease in an urban region. Our experimental results demonstrated how realistic modeling considerations can impact the disease dynamics; the modeling framework is general enough and yet efficient to undertake such studies. Further development of the modeling framework is necessary for modelers to study this interaction. We also described formal mathematical questions that arise when studying these complicated interactions. Most of the computational complexity as well as the algorithmic questions arising in this context are open problems and represent interesting directions for future research.

8. Thank You Dan

The group members of Network Dynamics and Simulation Science Laboratory want to wish Professor Daniel Rosenkrantz a happy retirement from active academics. He has been a collaborator for over eight years now; the computational theory of discrete dynamical systems to understand computer simulations of socio-technical systems was developed jointly with him. Dan's contributions and insights to the development of this theory have been invaluable, and his continued collaboration with us is a source of new ideas and inspiration. Madhav Marathe would like to express a special note of thanks and gratitude to Dan for being his teacher, mentor, colleague and a friend over the last 19 years.

Acknowledgments

We thank the members of Network Dynamics and Simulation Science Laboratory; the work presented here is based on work done by the entire group over the last 10 years. This work has been partially supported NSF Grants Nets CNS-062694, HSD SES-0729441, and NECO CNS 0831633, CDC Center of Excellence in Public Health Informatics Grant 2506055-01, NIH-NIGMS MIDAS project5 U01 GM070694-05, and DTRA CNIMS Grant HDTRA1-07-C-0113. Computational support for the work was provided in part by the National Science Foundation through TeraGrid resources provided by NCSA, TACC and PSC.

References

[1] J. Aspnes, S. Chang, and Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.*, 72(6):1077–1093, 2006.

[2] V. Bala and S. Goyal. A non-cooperative model of network formation. *Econometrica*, 68 (5):1181–1231, 2000.

[3] C. Barrett, J. Smith, and S. Eubank. Modern epidemiology modeling. *Scientific American*, 292(3):54–61, 2005.

[4] C. L. Barrett, R. J. Beckman, K. P. Berkbigler, K. R. Bisset, B. W. Bush, K. Campbell, S. Eubank, K. M. Henson, J. M. Hurford, 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. D. Williams. Transims: Transportation analysis simulation system. Technical Report LA-UR-00-1725, Los Alamos National Laboratory Unclassified Report, 2001.

[5] C. L. Barrett, K. Bisset, S. Eubank, V. S. A. Kumar, M. V. Marathe, and H. S. Mortveit. Modeling and simulation of large biological and information and socio-technical systems: An interaction-based approach. In *Proc. Short Course on Modeling and Simulation of Biological Networks, AMS Lecture Notes, Series: PSAPM*, 2007.

[6] C. Barrett, K. Bisset, J. Chen, B. Lewis, S. Eubank, V. S. A. Kumar, M. Marathe, and H. Mortveit. Effect of public policies and individual behavior on the co-evolution of social networks and infectious disease dynamics. In *Proc. DIMACS DyDAn Workshop on Computational Methods for Dynamic Interaction Networks*, 2007.

[7] C. Bauch and D. Earn. Vaccination and the theory of games. *Proc. Natl. Acad. Sci.*, 101(36):13391–13394, 2004.

[8] Y. Bramoulle, D. Lopez-Pintad, S. Goyal, and F. Vega-Redondo. Social interaction in anti-coordination games. *International Journal of Game Theory*, 33(1):1–19, 2004.

[9] R. Breban, R. Vardavas, and S. Blower. Inductive reasoning games as influenza vaccination models: Mean field analysis. In arXriv: q-bio.PE/0608016, 2006.

[10] N. Durlauf and P. Young. *Social Dynamics*. MIT Press, Cambridge, 2001.

[11] G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61:1047–1071, 1993.

[12] M. Emirbayer and J. Goodwin. Network analysis, culture and the problem of agency. *American Journal of Sociology*, 99:1411–1454, 1994.

[13] J. Epstein, J. Parker, and D. Cummings. Coupled contagion dynamics of fear and disease: A behavioral basis for the 1918 epidemic waves: Mathematical and computational explorations. Technical Report, Brookings Institute, 2006. Presentation made at the MIDAS meeting.

[14] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modeling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184, 2004.

[15] S. Eubank, V. S. A. Kumar, M. Marathe, A. Srinivasan, and N. Wang. Structural and algorithmic aspects of large social networks. In *Proc. 15th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 711–720, 2004.

[16] S. Eubank, V. S. A. Kumar, M. Marathe, A. Srinivasan, and N. Wang. Structure of social contact networks and their impact on epidemics. In *AMS-DIMACS Special Volume on Epidemiology*, 2005.

[17] N. Fredkin. *A Structural Theory of Social Influence*. Cambridge University Press, Cambridge, 1998.

[18] S. Goyal and F. Vega-Redondo. Learning, network formation, and coordination. *Games and Economic Behavior*, 50(2):178–207, 2005.

[19] N. Immorlica, D. Karger, M. Minkoff, and V. S. Mirrokni. On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. In *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 684–693, 2004.

[20] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Proc. International Colloquium on Automata Programming and Languages (ICALP)*, pages 1127–1138, 2005.

[21] M. Kermer. Integrating behavioral choice into epidemiological models of the aids epidemic. *The Quarterly Journal Of Economics*, CXI:549–573, 1996.

[22] P. Lazarsfeld and R. Merton. Friendship as social process. In T. Abel and C. Page, editors, *Freedom and Control in Modern Society*, Van Nostrand, New York, 1957.

[23] R. Leenders. Models for network dynamics. *J. Mathematical Sociology*, 20:1–21, 1995.

[24] R. Leenders. Structure and influence, statistical models for the dynamics of actor attributes, network structure and their independence. PhD Thesis, Amsterdam, 1995.

[25] M. Mavronicolas, V. Papadopoulou, A. Philippou, and P. Spirakis. A network game with attacker and protector entities. In *Proceedings of the 16th Annual International Symposium on Algorithms and Computation (ISAAC 2005)*, volume 3827, pages 288–297, 2005.

[26] T. Moscibroda and R. Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. In *25th Annual Symposium on Principles of Distributed Computing (PODC)*, pages 35–44, 2006.

[27] C. Papadimitriou. Games against nature. *Journal of Computer and System Sciences*, 31:288–301, 1985.

[28] R. Ravi and A. Sinha. Hedging uncertainty: Approximation algorithms for stochastic optimization problems. *Math. Program.*, 108(1):97–114, 2006.

[29] T. Snijders, C. Steglich, and M. Schweinberger. Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud and A. Satorra, editors, *Longitudinal Models in the Behavioral and Related Sciences*. Routledge/Taylor & Francis, New York, 2006.

[30] C. Steglich, T. Snijders, and M. Pearson. Dynamic networks and behavior: Separating selection from influence. Technical Report, University of Groningen, The Netherlands, 2007. Available at http://stat.gamma.rug.nl/snijders/.

[31] W. Stewart, J. Ricci, E. Chee, and D. Morganstein. Lost productive work time costs from health conditions in the United States: Results from the American productivity audit. *Journal of Occupational & Environmental Medicine*, 45(12):1234–1246, 2003.

[32] C. Swamy and D. Shmoys. Approximation algorithms for 2-stage stochastic optimization problems. *ACM SIGACT News*, 37(1):33–46, 2006.

[33] P. Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton, 1998.