Computational Epidemiology as a Challenge Domain for Multiagent Systems

Samarth Swarup, Stephen G. Eubank, Madhav V. Marathe
Network Dynamics and Simulation Science Laboratory,
Virginia Bioinformatics Institute,
Virginia Tech,
Blacksburg, VA, USA.
{swarup, seubank, mmarathe}@vbi.vt.edu

ABSTRACT

This paper describes the state of the art in computational epidemiology and makes the case that the multiagent systems field can both contribute much to, and gain much from, applications in this area. We discuss some of the major problems in epidemiology that can be addressed using multiagent techniques, and the challenges therein.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics, Sociology, Psychology*

General Terms

Algorithms, Experimentation

Keywords

Computational Epidemiology; Social Modeling; Public Health; Big Data

1. INTRODUCTION

Epidemics place a huge burden upon society. For example, the annual economic burden of influenza in the United States alone has been estimated to be $87.1 billion (in 2003 USD) [18]. This includes the estimated cost of lost productivity and lost lives in addition to hospitalization costs.

Health-adverse behaviors such as tobacco use have also been shown to be contagious [12], and the "epidemic" of tobacco use places an even bigger cost upon society. The Centers for Disease Control and Prevention (CDC) have estimated that the health-related economic losses in the United States due to cigarette smoking in the years 2000-2004 were $193 billion annually [4]. Other social contagions, like alcohol abuse, drug abuse, and the spread of obesity take a similar toll.

In today's world of increasing urbanization and increasing global interconnectedness, pandemic-preparedness has taken on a growing urgency. However, epidemics are complex social phenomena, and choosing appropriate policies to minimize their spread is tricky, because policies can backfire or

have side effects that are worse than the problems they fix. This has led these problems to be termed "policy-resistant" [27].

The complex "ecology" of epidemics is illustrated in Figure 1. Beyond physical aspects of the disease such as transmission modes, incubation periods, infectious periods, etc., we have individual, social, logistical, and structural factors that affect the size of the outbreak, the rate of its progression, which populations are vulnerable, which populations are able to get treatment, and so on.

Many of the factors listed are interrelated. For example, individual behavior depends on social norms, which in turn constrain the policy-making process. Similarly demographics and socioeconomics are correlated, and are in turn related to neighborhood effects and vaccine distribution methods and policies.

In addition to the spread of the disease itself, there is typically an associated contagion of information (and misinformation), which affects individual decision-making about whether to get vaccinations, whether to alter travel plans, etc. [1, 7, 13].

Addressing these complexities requires careful and sophisticated modeling, and computational approaches are increasingly being used in this domain. We believe that multiagent techniques have a lot to offer in surmounting some of the major problems that still stand in the way. In turn, computational epidemiology has the potential to drive a lot of interesting multiagent systems research by offering a very concrete domain with complex problems and much available data.
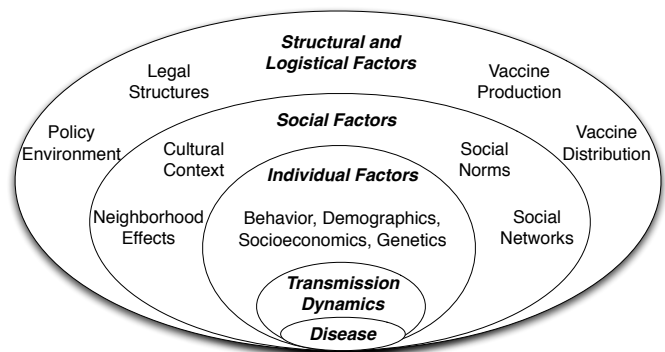

Figure 1: The complex ecology of epidemics.

2. WHAT IS COMPUTATIONAL EPIDEMIOLOGY?

Computational epidemiology consists of a collection of mathematical and computational techniques used for modeling the spread of epidemics and the effects of interventions to counter this spread.

These techniques can be divided into a few main categories. Luke and Stamatakis [16] provide a good overview.

System dynamics methods use differential equations to model the interactions between susceptible and infectious people. The population can be divided into a number of subpopulations of interest, which can lead to complex sets of equations which are then solved numerically. This approach is also referred to as the meta-population approach.

Network-based methods extend the system dynamics approach to graphs. Different classes of graphs can lead to very different epidemic behaviors. For example, the epidemic threshold goes to zero on infinite scale-free networks, which means that even diseases with very low transmissibility will result in large outbreaks on such networks. Networkbased interventions are also a very active area of research.

Agent-based models are becoming increasingly popular in this domain. They offer the benefit of being able to model dynamically changing interactions between people (as opposed to static networks), and of modeling more complex phenomena like neighborhood effects, markets, social capital, etc.

Synthetic information methods are sophisticated agentbased models which build very detailed and realistic representations of populations and interactions by combining multiple data sources [8, 23]. This approach has the benefit of providing very realistic approximations to actual populations where data on that scale cannot be gathered through surveys or other methods.

Despite the long history and increasing sophistication of all these approaches, there are several fundamental problems which remain to be solved.

3. PROBLEMS TO BE SOLVED

Here we present a selection of the conundrums pertaining to the social ecology of epidemics that might benefit from multiagent modeling.

Modeling human decision-making in social systems: Ofri, a practicing clinician, describes how many people (especially parents) vacillated between wanting and not wanting to take the vaccination (for themselves and for their children) for the H1N1 influenza in 2009 [21]. She writes that during the early stages of the pandemic, many parents would call her anxiously about whether the vaccine was available yet, though these very people would often refuse the seasonal influenza vaccination. Further, when the H1N1 vaccine did become available, they also grew reluctant to take that. There had been no new medical data or news that had appeared in the interim that would seem to have influenced their change of heart. Ofri attributes this seeming irrationality to, "a psychological contagion of myth and suspicion." More generally, the problem here is to understand how people make decisions in a social context, and how these decisions affect epidemics [9].

Forecasting: The complexity of large-scale social systems makes point prediction infeasible. It is impossible to predict exactly how many people will become sick, or exactly when the epidemic will peak. Instead the focus has to be on forecasting trends, possible-worlds reasoning, and bounding variance. This requires sophisticated statistical experiment design because the parameter space is generally too big to be covered completely. Advances at the interface of statistics (experiment design, analytics) and simulations are needed.

Modeling interventions and policy-making: The severity of an outbreak depends on how quickly and effectively we are able to respond to it. Computational modeling can be invaluable in rapid forecasting of outcomes of different policies and interventions [10]. However, policy decisions often have to be made when the information available is uncertain, and interventions have costs [15]. For example, large-scale prophylactic use of anti-virals carries a risk of rare adverse health effects, closing schools and other venues of high population mixing has an economic cost, and so on. Further, policy decisions (such as vaccination policy) are faced with other pressures (political, social, logistical). Altogether, epidemic response is a complex process involving many stakeholders. Computational modeling of this process itself can shed valuable light on epidemic outcomes.

Disease surveillance: When the symptoms of an illness are relatively mild, as in the case of influenza, people often do not go to hospitals. In such cases, data available through the healthcare system represent a biased underestimate of the true extent of the epidemic. Monitoring social epidemics, like smoking behavior, is even more difficult. Recently, the use of web and social media data has become popular for disease surveillance [3, 19, 24, 25]. However, these methods can be difficult to validate. New techniques for disease surveillance, e.g. by combining predictive models with targeted surveys, are needed to get better estimates.

Inferring causality: Identifying causes of epidemics, both biological and behavioral, has always been one of the central concerns of epidemiology. However, causes can be multilevel and feedback-driven [11]. The problem is complicated by the fact that standard methodologies like randomized controlled trials (RCTs) cannot be applied to epidemics. In such cases, computational modeling might offer new insights to the complex causality underlying epidemic processes. *In silico* RCTs, done carefully, can provide a replacement for real-world experimentation.

4. APPLICABILITY OF MULTIAGENT SYSTEMS TECHNIQUES

A great deal of technology that has been developed in the MAS domain can be applied to the above (and other) problems in epidemiology. We give just a brief sampling.

Work on norms: Vaccination, and vaccine refusal, are classic examples of social norms and norm violation. Interactions between individuals can lead to large-scale patterning of behavior (emergent phenomena), e.g., due to spread of misinformation about vaccine side-effects. If even small groups of people choose not to vaccinate, it establishes a reservoir for a disease, and complexities of network dynamics often lead to outbreaks from this reservoir. In social contagions as well, how norms emerge, and under what circumstances people choose to violate them are key issues to understand in modeling epidemics. These are well-studied topics in the MAS community (e.g., [14]) and can be fruitfully applied in the public health domain.

BDI modeling: Belief-desire-intention modeling has a long history in MAS research. It has been used in modeling emotional behavior (e.g., [2]) as well as normative behavior (e.g., [17]), both of which are important in understanding the seeming irrationality of some human decision-making, for example as discussed earlier in the context of H1N1 vaccination refusal. Cognitive and behavioral agent modeling, especially in the context of decision-making under uncertainty, is a central concern for understanding epidemic dynamics.

MDP-style modeling: Since epidemics take place over a network, there is the possibility of influencing decision-making and exploiting cascading behavior to set up positive norms. Recent work in applying MDP-style modeling to influence-maximization problems [5] is very promising in this regard. This approach may also be useful for finding optimal vaccine allocation policies or for finding optimal non-pharmaceutical intervention strategies. A simple example would be where the state of the system corresponds to the vector of states of the nodes (S, E, I, R), the actions are to vaccinate some subset of nodes, and the reward is the number of people in state S at time T.

Game-theoretic modeling: Similar problems of intervening in networks have been studied using game-theoretic analysis in the MAS community, for example in the context of security [29]. Behavioral game theory (e.g., [30]) can also be applied to understanding policy-making as well as individual and organizational decision-making during epidemics. This approach could be applied to modeling the intentional spread of misinformation, e.g., against the HPV vaccine, or the controversy over the supposed link between the MMR vaccine and autism.

Learning and adaptation: The behavior of individuals and organizations also changes over time due to learning, forgetting, and the choices of other agents. Rigorous models of these adaptations can help explain and model changes in organizational and population response to epidemics [6, 28].

Simulation-based modeling: Large-scale MAS simulations offer the opportunity to tie together many of the previous approaches and to rigorously investigate the effects of intervention strategies under varying initial conditions, varying levels of vaccine efficacy, varying levels of compliance with policy directives, etc. Sophisticated simulation technology is needed to manage the scale of epidemics [20, 22].

5. CHALLENGES

While the previous section discusses how existing MAS technologies can be used, the epidemiology domain also offers a set of unique challenges that can drive MAS research.

Scalability: Epidemics are large-scale social phenomena, involving populations of hundreds of millions of people, big geographical areas, and complex networks. These "big data" scales are challenging for many computational techniques and new algorithms may prove to be necessary, including parallel methods and sampling-based methods. Simulations at this scale need high-performance computing support. In order to be useful to decision-makers, computational methods must run fast enough to provide results at a rate that can be incorporated into the decision cycle, e.g., epidemiologists often need to make policy decisions every day during an epidemic outbreak, based on new information that comes in daily from healthcare systems and other sources.

Emergent phenomena: It is important to be able to predict emergent outcomes of interventions. For example, increasing taxes on cigarettes can lead to the emergence of informal economies of single-cigarette sales [26]. While this may appear to reduce overall smoking rates, it has the adverse long-term effect of keeping adolescents and young adults addicted, so that when they get older and have more disposable income, they also start smoking more heavily.

Spatio-temporal complexity: More broadly, an epidemic is a very complex spatial and temporal process. Nearby regions can have very different outcomes, and these can vary considerably over time. Modeling and understanding these variations requires analytic tools that can make sense of the outcomes of complex multiagent simulations. The problem of inferring causality applies equally to simulation models.

Fidelity, precision, adequacy: A general problem faced by models (especially simulation models) consists of understanding what information needs to be included in the model in order to be able to answer a particular query. In this context, we use the term *fidelity* to refer to the set of properties represented: for example, do we need to represent age and gender, or age, gender, and race, or some other properties? *Precision* refers to the accuracy with which these properties correspond to the actual population. *Adequacy* refers to the set of outcomes of the model which are reflective of real-world outcomes. For example, a model with a given fidelity and precision may be adequate for modeling baseline epidemic progression, but not modeling particular interventions like school closures. Rigorous methods are needed to be able to quantify these properties.

Incorporating real-time data into models/simulations: Epidemics are temporally extended phenomena, and as the state of the epidemic changes in responses to interventions and contingencies, it is necessary to be able to incorporate new information into models on the fly. How to incorporate streaming data, e.g., from social media, into simulations and other models represents an exciting direction for research. New methods for gathering data relevant to interactions within a population, activity patterns, and mobility patterns are also needed. The use of digital data sources has led to a stream of work known as digital epidemiology [24].

6. DATA AVAILABILITY

To spur research in this domain in the MAS community, we are making some synthetic population data available. Populations, activities, contact graphs, and sample dendrograms for simulated epidemic outbreaks are available online. A contact network for Montgomery County in Virginia is also available. The URL for the data is
`http://ndssl.vbi.vt.edu/synthetic-data/`

The populations contain demographics and home locations. Combining these with activity sequences for each person and activity locations leads to a social contact graph, which is a network of people coming into contact with each other over the course of a typical day. This is essentially the network over which an epidemic spread happens. These social contact graphs can be used for network-based analyses and simulations. The data are freely available for research.

7. CONCLUSION

We have tried to make the case that computational epidemiology offers a rich set of problems that can drive mul-

tiagent systems research, and that MAS techniques in turn can make a big contribution to solving some of the most challenging problems in epidemiology and public health.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] C. T. Bauch and A. P. Galvani. Social factors in epidemiology. *Science*, 342:47–49, Oct 2013.

[2] C. Becker-Asano and I. Wachsmuth. Affective computing with primary and secondary emotions in a virtual human. *Auton Agent Multi-Agent Syst*, 20:32–49, 2010.

[3] J. S. Brownstein, C. C. Freifeld, B. Y. Reis, and K. D. Mandl. Surveillance sans frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Medicine*, 5(7):e151, 2008.

[4] Centers for Disease Control and Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses–United States, 2000-2004. *Morbidity and Mortality Weekly Report*, 57(45):1226–1228, 2008.

[5] G. C. Chasparis and J. S. Shamma. Control of preferences in social networks. In *Proc. IEEE Conference on Decision and Control*, Atlanta, GA, USA, Dec 15–17 2010.

[6] Q. Duong, M. P. Wellman, S. Singh, and M. Kearns. Learning and predicting dynamic networked behavior with graphical multiagent models. In *Proc. AAMAS*, 2012.

[7] J. M. Epstein, J. Parker, D. Cummings, and R. A. Hammond. Coupled contagion dynamics of fear and disease: Mathematical and computational explorations. *PLoS ONE*, 3(12):e3955, 2008.

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

[9] N. Ferguson. Capturing human behavior. *Nature*, 446:733, 2007.

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

[11] S. Galea, M. Riddle, and G. A. Kaplan. Causal thinking and complex systems approaches for computational epidemiology. *International Journal of Epidemiology*, 39:97–106, 2010.

[12] B. R. Hoffman, S. Sussman, J. B. Unger, and T. W. Valente. Peer influences on adolescent cigarette smoking: A theoretical review of the literature. *Substance Use and Misuse*, 41:103–155, 2006.

[13] D. M. Kahan. A risky science communication environment for vaccines. *Science*, 342:53–54, 2013.

[14] M. Knobbout and M. Dastani. Reasoning under compliance assumptions in normative multiagent systems. In *Proc. AAMAS*, 2012.

[15] M. Lipsitch, S. Riley, S. Cauchemez, A. C. Ghani, and N. M. Ferguson. Managing and reducing uncertainty in an emerging influenza pandemic. *N Engl J Med*, 361(2):112–115, 2009.

[16] D. A. Luke and K. A. Stamatakis. Systems science methods in public health: Dynamics, networks, and agents. *Annu. Rev. Public Health*, 33:25.1–25.21, 2012.

[17] F. Meneguzzi and M. Luck. Norm-based behavior modification in BDI agents. In *Proc. AAMAS*, 2009.

[18] N.-A. M. Molinari, I. R. Ortega-Sanchez, M. L. Messonnier, W. W. Thompson, P. M. Wortley, E. Weintraub, and C. B. Bridges. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine*, 25:5086–5096, 2007.

[19] M. Myslin, S.-H. Zhu, W. Chapman, and M. Conway. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of Medical Internet Research*, 15(8):e174, 2013.

[20] L. Navarro, F. Flacher, and V. Corruble. Dynamic level of detail for large scale agent-based urban simulations. In *Proc. AAMAS*, 2011.

[21] D. Ofri. The emotional epidemiology of H1N1 influenza vaccination. *New England Journal of Medicine*, 361:2594–2595, Dec 2009.

[22] N. Parikh, S. Swarup, P. E. Stretz, C. M. Rivers, B. L. Lewis, M. V. Marathe, S. G. Eubank, C. L. Barrett, K. Lum, and Y. Chungbaek. Modeling human behavior in the aftermath of a hypothetical improvised nuclear detonation. In *Proc. AAMAS*, 2013.

[23] N. Parikh, M. Youssef, S. Swarup, and S. Eubank. Modeling the effect of transient populations on epidemics in Washington DC. *Scientific Reports*, 3:Art 3152, Nov 2013.

[24] M. Salathé, L. Bengtsson, T. J. Bodnar, D. D. Brewer, J. S. Brownstein, C. Buckee, E. M. Campbell, C. Cattuto, S. Khandelwal, P. L. Mabry, and A. Vespignani. Digital epidemiology. *PLoS Comp. Biol.*, 8(7):e1002616, 2012.

[25] A. Signorini, A. M. Segre, and P. M. Polgreen. The use of Twitter to track levels of disease activity and public concern in the U.S. during the Influenza A H1N1 pandemic. *PLoS ONE*, 6(5):e19467, 2011.

[26] K. C. Smith, F. Stillman, L. Bone, N. Yancey, E. Price, P. Belin, and E. E. Kromm. Buying and selling "loosies" in Baltimore: the informal exchange of cigarettes in the community context. *Journal of Urban Health*, 84(4):494–507, 2007.

[27] J. D. Sterman. Learning from evidence in a complex world. *Am. J. Public Health*, 96(3):505–514, 2006.

[28] B. Subagdja, W. Wang, A.-H. Tan, Y.-S. Tan, and L.-N. Teow. Memory formation, consolidation, and forgetting in learning agents. In *Proc. AAMAS*, 2012.

[29] J. Tsai, T. H. Nguyen, and M. Tambe. Security games for controlling contagion. In *Proceedings of AAAI*, Toronto, CA, July 22-26 2012.

[30] J. R. Wright and K. Leyton-Brown. Behavioral game-theoretic analysis: A Bayesian framework for parameter analysis. In *Proc. AAMAS*, 2012.