Insights from Modeling Epidemics of Infectious Diseases – A Literature Review

Elisa Canzani
Universität der Bundeswehr München
Elisa.canzani@unibw.de

Ulrike Lechner
Universität der Bundeswehr München
Ulrike.lechner@unibw.de

ABSTRACT

The relevance of modeling epidemics' spread goes beyond the academic. The mathematical understanding of infectious diseases has become an important tool in policy making. Our research interest is modeling of dynamics in crisis situations. This paper explores the extant body of literature of mathematical models in epidemiology, with particular emphasis on theories and methodologies used beyond them. Our goal is to identify core building blocks of models and research patterns to model the dynamics of crisis situations such as epidemics. The wide range of applications of epidemic models to many other disciplines that show biological analogies, makes this paper helpful for many modelers and mathematicians within the broader field of Crisis Management.

Keywords

Epidemics modeling, infectious diseases, dynamics, crisis modeling.

INTRODUCTION

Through the years, the "persistent threat" of epidemics received attention of researchers. Epidemiology as a discipline has emerged in the last century. In 1911, Ross (Ross, 1911) proposed a model for the transmission of malaria. Later a couple of seminal works established the basis of mathematical epidemiology. Among them, the first more general and rigorous study was made by Kermack and McKendrick (Kermark & McKendrick, 1927). The etiological agent of AIDS (i.e. HIV) emerged in 1981 and motivated new models for studying dynamics and social complexity of the spread of this disease. Plague, cholera and hemorrhagic fevers such as Ebola contribute to the needs for better tools to understand epidemics and the underlying dynamics. According to Hethcote (Hethcote, 1994), understanding transmission characteristics of infectious diseases can lead to better approaches to decreasing their transmission in communities, regions and countries. Accordingly decision-making tools need underlying knowledge of epidemic dynamics and how to capture such nonlinear behaviors. This nonlinear dynamics is of interest in understanding crisis situations in general and motivates this study.

There are a number of reviews of epidemiology models. They tend to focus on different modeling approaches to understand the dynamics of a particular infectious agent. E.g., a comprehensive survey of mathematical models (mainly differential equation based) for the transmission of Malaria is Mandal et al. (Mandal, Sarkar, & Sinha, 2011). A more mathematical monograph on modeling

patterns for infectious diseases is by Diekmann and Heesterbeek (Diekmann & Heesterbeek, 2000). They argue that the theory of dynamical systems is the tool where real experiments are impossible or unethical. Our research interest is modeling of crises in general and this literature study aims at exploring the field of epidemiology with its main mathematical models of linear and non-linear dynamics.

METHODOLOGY

Attempting to identify core building blocks of models and research patterns to model epidemic dynamics, a literature review was performed. Guidelines followed to structure the review refer to Randolph (Randolph, 2009) and Webster (Webster & Watson, 2002).

The evident focus on research methods and the wide range of existing epidemic models led to a qualitative literature review which look for pivotal articles in the field. In line with Hethcote (Hethcote, 1994), we believe inappropriate to analyze the many different possible models one by one in published papers. Rather, only the analyses of models which break new ground or show important new aspects are of significant interest. Thus, a concept-centric approach was adopted to address main research questions such as: which are the relevant models developed for studying epidemics' spread? Which are theories and methodologies used beyond them?

Papers from journals and high-quality conferences were collected through an electronic search of academic databases on the Internet (e.g. Science Direct, Elsevier, Google Scholar, ACM digital library, Springer link etc.). The data collection process was done on the basis of specific key words, such as "crisis dynamics", "crisis modeling", "epidemic models", "mathematical epidemiology", "modeling infectious diseases", and combinations of them. Then, seminal reviews and most cited articles were analyzed in a snowball approach to identify relevant articles until a well-rounded collection of articles has been identified.

DETERMINISTIC AND STOCHASTIC MODELS

Epidemic modeling dates back to 1760 when Daniel Bernoulli modeled Smallpox to show that the life expectancy could be increased by a universal inoculation against Smallpox (Bernoulli, 1760). In the early 1900, Ross used mathematical functions to study the transmission of Malaria (Ross, 1911). Modern modeling began with McKendrick and Kermack and their deterministic model of outbreaks validated with existing records of epidemics (Kermark & McKendrick, 1927).

Their model assumes that only susceptible individuals get infected and, after having been infectious (infectious period), an individual dies or recovers and becomes completely immune. Real situations are far away from this scenario, but it led to a milestone in epidemiology: the McKendrick-Kermack threshold condition. It states that an infection outbreaks if and only if $R_0 > 1$, where the basic reproduction ratio R_0 is defined as the average number of secondary infections produced when one infected individual is introduced into a fully susceptible host population. The threshold quantity R_0 is studied under more realistic conditions (See e.g. (Hethcote, 2000)).

In general, deterministic epidemiology modeling often refers to 'compartments models'. Compartments represent epidemiological categories of communities of individuals. A standard compartment model MSEIR is depicted in Fig 1.
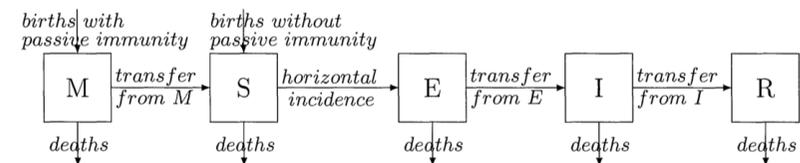

Figure 1. MSEIR model (Hethcote, 2000).

Class M contains the infants born with passive immunity, in S there are susceptible individuals that can move in the class of exposed E if they are in the latent period after they get infected. Individuals enter in the class I of infectives when the latent period ends, and R is the class containing those who recover from the infection and have acquired a permanent immunity.

Mathematically, compartmental models are formulated as a set of differential equations. The SIR model of McKendrick and Kermack is based on:

1. Given a *close* and *large* population of size N, $S(t) + I(t) + R(t) = N$ at time t;
2. Let β the constant contact rate, an average member of the population makes contact sufficient to randomly transmit infection with βN other individuals per unit time;
3. The quantity $1/\gamma$ is the average infectious period so that infectives leave the class I with a rate of γI per unit time.

The SIR model is defined as an initial value problem as follows:

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \qquad S(0) = S_0 \geq 0,$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \qquad I(0) = I_0 \geq 0, \qquad \text{(a)}$$

$$\frac{dR}{dt} = \gamma I, \qquad R(0) = R_0 \geq 0.$$

Formulating models as differential equations it is assumed so far that the epidemic process is deterministic. This means that the behavior of a population is completely determined by its history.

Stochastic epidemic model are a popular generalization of deterministic models formulated as Markov chains, and they enable to consider standard errors in estimating parameters from real data of a disease spread. Stochastic models account for *probability* of major outbreaks. (Notable examples cf (Britton, 2010)).

Models that take into account population demography are another extension that allows to study 'dynamic' communities with a varying population size. The SIR model is extended to include also the births which change according to the variation in age distribution of the total population over time. (cf (Hethcote, 2000)).

While early mathematical contributions to epidemiology generally modeled the contact process by the law of mass action, which asserts that the infection is proportional to the term $I(t)S(t)$ in (a), later epidemic models consider that in real life the contact rate depends on the disease being studied as well as on social and behavioral factors or on the age structure of a population. Hence, age-dependent mixing models include both time t and age a as independent variables (Hethcote, 2000) (Anderson, 1991).

Newman argues that the research on traditional epidemics modeling seems to be exhausted and the field of epidemiology has been revolutionized by the science of networks (M. E. J. Newman, 2002). The following section focuses on modeling efforts that highlight the interplay between network theory and dynamics of epidemics' spread.

NETWORKS AND SPATIAL MODELS

A network topology in an epidemic model facilitates an accurate representation of real contact patterns, which can depend on spatial and geographic components. This is important for airborne and vector-borne diseases, host-pathogen biology and sexual transmitted diseases (M. E. J. Newman, 2002). Although many modifications to the basic SIR framework have been made by modeling different mixing rates between population subgroups (metapopulation models), Keeling and Eames (Keeling & Eames, 2005) point out that the approximation of random mixing remains at least between individuals within each pair of subgroups.

Mathematical modeling of non-linear diseases' dynamics with networks traces back to the mid-1980s with the need to understand AIDS\HIV. (A pioneering study is (Klovdahl, 1985)).

In general, the study of networks has its grounding in social sciences and graph theory. While social literature speaks about 'actors' and 'relations', graph theory uses the terms 'nodes' and 'edges', epidemiology deals with 'individuals' and 'contacts' (Keeling & Eames, 2005). There are three main techniques to build real networks: *Infection tracing*, *Contact tracing* and *Diary-based tracing*. Currently, POLYMOD is the most comprehensive diary based study (Mossong et al., 2008).

In general, the tracing method applied depends on the purpose for which the data is retrieved and on the resource available. An alternative source of information comes from recorded movements of individuals as, e.g., airline transportation network, the movement of individuals to and from work, of dollar bills, and of

livestocks (e.g., the Cattle Tracing Scheme records all the movements of cattle between farms in Great Britain to understand Mad Cow Disease (Robinson, Everett, & Christley, 2007)).

Simulated networks are developed to match available data on observed social characteristics (e.g. Smallpox network model by (Halloran, Jr, Nizam, & Yang, 2002)). Anderson (Anderson, 1991) provides examples on how social behavior observation and mathematical tools converge in generating contact networks.

Population mixing is modeled by a matrix P with mixing probabilities $p(i,j)$, having a series of constraints as follows. For all i,j combinations,

$$0 \leq p(i,j) \leq 1 \qquad (b.1)$$
$$\sum_i p(i,j) = 1 \qquad (b.2)$$
$$p(i,j) = p(j,i) \qquad (b.3)$$

Each element $p(i,j)$ describes the proportion of contacts of an individual i with an individual j. The third propriety (symmetry) states an infection can pass either ways across a contact, as in most mixing networks. Shifting such idea to the concept of sexual activity classes, simple forms for P are obtained with *restricted, proportional or preferred mixing*. (Mathematical details in (Anderson, 1991).

Mathematical theory underlying such studies refers to Complex Networks. A graph (or network) is defined by nodes (individuals or classes of individuals), and edges between nodes (contacts). Network epidemic models refer to the set of contacts of an individual as his 'neighborhood', and the size of such neighborhood is the individual's degree k. The 'degree distribution' is a set of probabilities $P(k)$ that a node chosen at random will have degree k. $P(k)$ captures heterogeneity in individuals' potential to become infected and cause further infections. Therefore, $P(k)$ strongly influences the recovery of epidemic threshold, breaking down the Kermack-McKendrick result (Pastor-Satorras & Vespignani, 2001).

Other measures relevant for epidemiology are the distance between two nodes, betweenness centrality, and clustering coefficient (i.e. measurement of local density). Thinking a network as collection of subgraphs, communities can be modeled as subgraphs as formalized by Newman and Givran (M. Newman &

Girvan, 2004) in terms of modularity measure. (See for details (Danon et al., 2011)).

The roles of social networks for the transmission of diseases is important and following the seminal review work conducted by Keeling and Eames (Keeling & Eames, 2005), five distinct types of networks are used (Fig. 2(a)-(e)).

- **Lattice network,** where individuals are positioned on a regular grid and contacts are localized in space. They are homogeneous at individual level and highly clustered because of the localized nature of connections.

- **Small-World network** was introduced by Milgram (Milgram, 1967) and modeled by Watts and Strogatz (Watts & Strogatz, 1998) to describe topological characteristics of social relationships and communities.

- **Random network** characterized by a random distribution of connections, a lack of clustering and short path lengths (while lattices have long path length due to the high clustering). Iconic examples are Erdös-Rényi random graphs with $P(k) \approx e^{-k}$.

- **Scale-free network** provide a means to achieve high levels of heterogeneity and can be modeled according to the Barabási-Albert algorithm. A common characteristic of such networks is $P(k) \approx k^{-3}$.

- **Spatial network** in which, starting from a set of locations, individuals are connected with a probability given by a connection kernel that usually decays with the distance. Spatial networks have an approximately Poisson degree distribution, with a reasonably high degree of heterogeneity.

In Fig. 2 each network type containing 100 individuals is represented on the left. The right side shows a graph plotting infectives' percentage over time for standard SIR epidemics solved on the correspondent network type. Considering a population of 10000 individuals with an average number of contacts per individual that is approximately 4 for all five networks, 100 epidemic curves (in grey) are shown together with the average curve (in black) for all major epidemics.
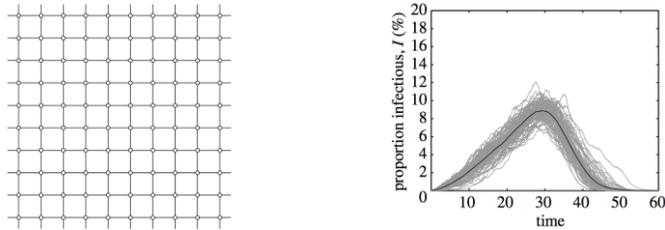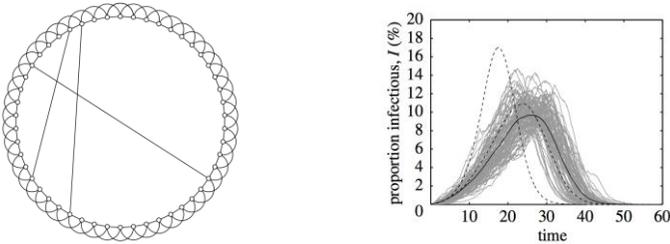
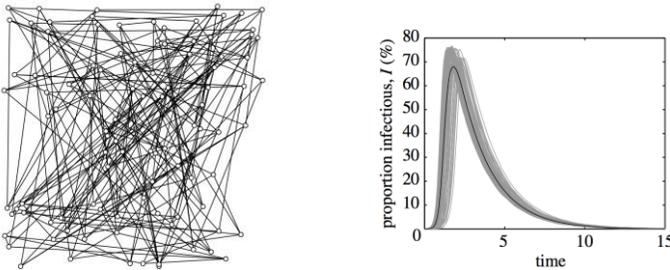**Figure 2(a). Lattice network**.

Figure 2 (b) Small-world network.

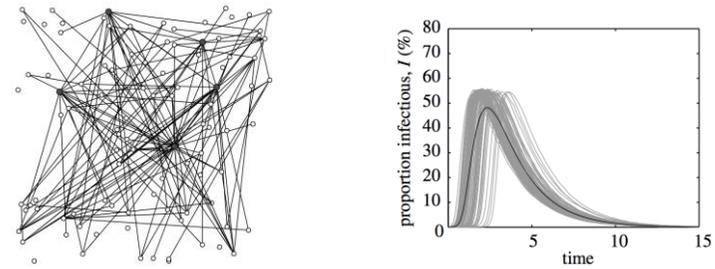**Figure 2 (c) Random network**.
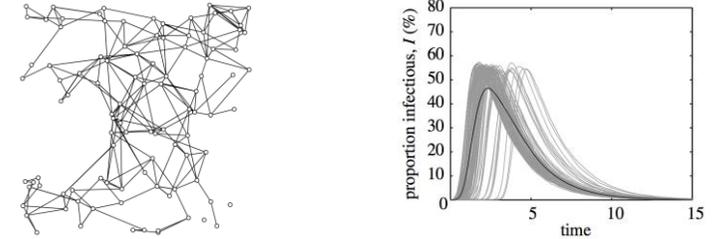

Figure (d) Scale-free network.

Figure 2(e). Spatial network.

Spatial epidemiology dates back to 1930s, when the parasitologist Pavlovsky developed concept of 'landscape epidemiology' to gather three simple observations: diseases are geographically limited, spatial variation lies on physical and biological changes of condition, and future risks should be predictable delimiting such conditions in a map. A review of the major approaches used for mapping spatiotemporal dynamics is (Ostfeld, Glass, & Keesing, 2005).

Classes of dynamic networks used in epidemiology modeling are *exponential random graphs* (Strauss & Frank, 1986) and c*o-evolutionary or adaptive networks* (Risau-Gusman & Zanette, 2009).

Current research in epidemiology moves from analytical methods to computer simulation techniques and other tools to track complex dynamics. This research is referred to as Computational Epidemiology.

COMPUTATIONAL EPIDEMIOLOGY

Computational Epidemiology uses methods and tools from computer science to study epidemiology in multidisciplinary approaches. The growing recognition that multi level factors and interrelations among these factors are often characterized by dynamic feedback and changes over time, resulted in calls for new technologies and simulation tools. Arguments for this methodological shift to complex dynamic systems can be found in (Galea, Riddle, & Kaplan, 2010) and on the complexity of the field in (Swarup, Eubank, & Marathe, 2014) as well as the need for usable tools in (Patlolla, Gunupudi, Mikler, & Jacob, 2006). Major modeling paradigms are System Dynamics, Network Analysis and Agent-based Modeling. Luke and Stamatakis (Luke & Stamatakis, 2012) present three case studies where these methods have been used in the areas of epidemiology.

While System Dynamics (SD) methodology centers on the fact that the complex behaviors of systems are the result of ongoing accumulations (of people, information, money, etc.) that change within complex feedback structures, Agent Based models the real world as a set of behaviors of agents, in which the overall dynamics of infection is the result of different events that interest the single agents. Comparisons in epidemics modeling are in (Bagni, Berchi, & Cariello, 2002) .

Sterman (Sterman, 2000) presents the classical SIR model developed using System Dynamics. Further references on background and opportunities of SD modeling for Public Health can be found in (Homer & Hirsch, 2006).

Patlolla et al. (Patlolla et al., 2006) conduct a survey of the state-of-the-art in agent based simulation tools used for computational epidemiology.

In particular, synthetic information methods are sophisticated agent based models that provide realistic approximations by combining multiple data sources. For modeling see (Marathe & Ramakrishnan, 2013).

CONCLUDING REMARKS

The dynamics of crisis situations is our research interest. This study on the extant body of literature in epidemiology illustrates the variety of approaches used for modeling dynamics. The complexity of epidemics modeling increased from compartmental SIR models that consider only biological pathogens to the various social network models. Understanding relations and modeling social architecture is crucial for understanding the dynamics of epidemics. Computational epidemiology goes beyond these traditional mathematical models to capture non-linear dynamics, complex behavior patterns in heterogeneous contexts by multi-method approaches.

Traditional models and pioneering results, such as the Kermack–McKendrick threshold condition, have been tremendously successful in informing public health policy. More work needs to be done to inform the general public and develop corresponding criteria and thresholds for crisis situations. A major aspect being explored today lies on the wide range of applications of epidemic models to many other disciplines that show biological analogies, e.g. innovation diffusion and growth of the market for new products (Sterman, 2000). In particular, this review represents a preliminary work to explore high dynamics of crisis situations such as the spread of computer viruses in the cyber world (Kephart, White, & Chess, 1993) and of the information diffusion on the IT infrastructure of a firm.

ACKNOWLEDGMENTS

REFERENCES

1. Anderson, R. M. (1991). Discussion: the Kermack-McKendrick epidemic threshold theorem. *Bulletin of Mathematical Biology*, *53*(1951), 3–32.

2. Bagni, R., Berchi, R., & Cariello, P. (2002). A comparison of simulation models applied to epidemics. *Journal of Artificial Societies and Social Simulation*, *5*(3).

3. Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad. Roy. Sci.(Paris) Avec Mém. Des Math. et Phys. and Mém*, 1–45.

4. Britton, T. (2010). Stochastic epidemic models: a survey. *Mathematical Biosciences*, *225*(1), 24–35. Retrieved from http://arxiv.org/abs/0910.4443v1

5. Danon, L., Ford, A. P., House, T., Jewell, C. P., Keeling, M. J., Roberts, G. O., … Vernon, M. C. (2011). Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, *2011*, 284909. doi:10.1155/2011/284909

6. Diekmann, O., & Heesterbeek, J. (2000). *Mathematical epidemiology of infectious diseases*. Wiley, Chichester. Retrieved from http://oai.cwi.nl/oai/asset/13562/13562A.pdf

7. Galea, S., Riddle, M., & Kaplan, G. a. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*, *39*(1), 97–106. doi:10.1093/ije/dyp296

8. Halloran, M. E., Jr, I. M. L., Nizam, A., & Yang, Y. (2002). Containing Bioterrorist Smallpox. *Science*, *298*(5597), 1428–1432.

9. Hethcote, H. (1994). A thousand and one epidemic models. *Frontiers in Mathematical Biology*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-50124-1_29

10. Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, *42*(4), 599–653. Retrieved from http://epubs.siam.org/doi/abs/10.1137/S0036144500371907

11. Homer, J. B., & Hirsch, G. B. (2006). System dynamics modeling for public health: background and opportunities. *American Journal of Public Health*, *96*(3), 452–8. doi:10.2105/AJPH.2005.062059

12. Keeling, M. J., & Eames, K. T. D. (2005). Networks and epidemic models. *Journal of the Royal Society, Interface / the Royal Society*, *2*(4), 295–307. doi:10.1098/rsif.2005.0051

13. Kephart, J., White, S., & Chess, D. (1993). Computers and epidemiology. *Spectrum, IEEE*, (May). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=275061

14. Kermark, M., & McKendrick, A. G. (1927). Contributions to the mathematical theory of epidemics. Part I. *In Proc. Roy. Soc. London Ser A*, *115*(5), 700–721.

15. Klovdahl, A. S. (1985). Social networks and the spread of infectious diseases: the AIDS example. *Social Science & Medicine*, *21*(11), 1203–1216.

16. Luke, D. A., & Stamatakis, K. A. (2012). System science methods in public health: dynamics, networks and agents. *Annual Review of Public Health*, *33*, 357–376. doi:10.1146/annurev-publhealth-031210-101222.Systems

17. Mandal, S., Sarkar, R. R., & Sinha, S. (2011). Mathematical models of malaria - a review. *Malaria Journal*, *10*(1), 202. doi:10.1186/1475-2875-10-202

18. Marathe, M. V, & Ramakrishnan, N. (2013). Recent Advances in Computational Epidemiology. *IEEE Computer Society*, *13*, 96–101. doi:10.1109/MIS.2013.114

19. Milgram, S. (1967). The small world problem. *Psychology Today*, *2*(1), 61–66. doi:10.1007/BF02717530

20. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, *5*(3).

21. Newman, M. E. J. (2002). The spread of epidemic disease on networks. *Physical Review E*, *66*(1). Statistical Mechanics; Disordered Systems and Neural Networks; Quantitative Biology. doi:10.1103/PhysRevE.66.016128

22. Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 1–16. doi:10.1103/PhysRevE.69.026113

23. Ostfeld, R. S., Glass, G. E., & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, *20*(6), 328–36. doi:10.1016/j.tree.2005.03.009

24. Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, *86*(14). Statistical Mechanics. doi:10.1103/PhysRevLett.86.3200

25. Patlolla, P., Gunupudi, V., Mikler, A. R., & Jacob, R. T. (2006). Agent-based simulation tools in computational Epidemiology. In *Innovative Internet Community Systems* (pp. 212–223). Springer Berlin Heidelberg. doi:10.1007/11553762_21

26. Randolph, J. J. (2009). A Guide to Writing the Dissertation Literature Review. *Practical Assessment, Research & Evaluation*, *14*(13).

27. Risau-Gusman, S., & Zanette, D. H. (2009). Contact switching as a control strategy for epidemic outbreaks. *Journal of Theoretical Biology*, *257*(1), 52–60. doi:10.1016/j.jtbi.2008.10.027

28. Robinson, S. E., Everett, M. G., & Christley, R. M. (2007). Recent network evolution increases the potential for large epidemics in the British cattle population. *Journal of the Royal Society Interface*, *4*(15), 669–674.

29. Ross, R. (1911). *The Prevention of Malaria* (2nd ed.). London: Murray.

30. Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin/McGraw-Hill.

31. Strauss, D., & Frank, O. (1986). Markov Graphs. *Ournal of the American Statistical Association*, *81*(395), 832–842.

32. Swarup, S., Eubank, S., & Marathe, M. (2014). Computational epidemiology as a challenge domain for multiagent systems. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, 1173–1176. Retrieved from http://dl.acm.org/citation.cfm?id=2617434

33. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442.

34. Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MISQ*, *26*(2), 13–23.