



Ana Pastore y Piontti
Nicola Perra
Luca Rossi
Nicole Samay
Alessandro Vespignani

CHARTING THE NEXT PANDEMIC

Modeling Infectious
Disease Spreading in the
Data Science Age

With Contributions by
Corrado Gioannini
Marcelo F. C. Gomes
Bruno Gonçalves

 Springer

Charting the Next Pandemic

Ana Pastore y Piontti • Nicola Perra • Luca Rossi
Nicole Samay • Alessandro Vespignani

Charting the Next Pandemic

Modeling Infectious Disease Spreading
in the Data Science Age

With contributions by Corrado Gioannini, Marcelo F. C. Gomes,
and Bruno Gonçalves

Authors

Ana Pastore y Piontti
Northeastern University
Boston, MA
USA

Luca Rossi
Institute for Scientific Interchange
Torino
Italy

Alessandro Vespignani
Northeastern University
Boston, MA
USA

Nicola Perra
University of Greenwich
London
UK

Nicole Samay
Northeastern University
Boston, MA
USA

Collaborators

Corrado Gioannini
Institute of Scientific Interchange
Torino
Italy

Bruno Gonçalves
Center for Data Science
New York, NY
USA

Marcelo F. C. Gomes
Oswaldo Cruz Foundation
Rio de Janeiro
Brazil

ISBN 978-3-319-93289-7 ISBN 978-3-319-93290-3 (eBook)
<https://doi.org/10.1007/978-3-319-93290-3>

© Springer International Publishing AG, part of Springer Nature 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is in memory of Duygu Balçan, who has been a fantastic scientist and friend. She was suddenly taken away from us and from science, but she will always be with us in our dearest memories and through her outstanding scientific contributions.

.....

ACKNOWLEDGMENTS

It is the work on many collaborative projects that led to the tools and methodologies presented in the pages of this volume. We benefited from more than 10 years of collaborations and scientific exchanges with a long list of colleagues who helped us in shaping our view and knowledge of computational epidemiology through invaluable scientific discussions.

First of all, we want to thank Vittoria Colizza for her invaluable scientific collaborations. She has also been one of the engines behind the development of the GLEAM framework. We can never thank her enough for the work and contributions to many of the ideas and concepts illustrated in this book.

We are also grateful to Alain Barrat, Marc Barthelemy, and Romualdo Pastor-Satorras, who have been early companions in the analysis of global mobility networks and global epidemics.

Paolo Bajardi, Hao Hu, José J. Ramasco, Daniela Paolotti, Chiara Poletto, Michele Tizzoni, and Qian Zhang have been key collaborators of ours for many years and have greatly contributed in many tangible and intangible ways to this volume. We have grown with and learned so much from them.

Marco Ajelli and Stefano Merler contributed to the development of computational models and data-driven agent-based models that have literally defined the state of the art in the field. We are very fortunate to have the privilege of collaborating and working with them through several common projects.

We are simply honored and humbled by the possibility of collaborating with pioneers in the field of the caliber of M. Elizabeth Halloran and Ira M. Longini. We are thankful to them for the many projects and ideas that have truly shaped our way of looking at epidemic modeling.

We started to think about global epidemics as a data visualization problem because of Wouter Van den Broeck. He is the one who translated the GLEAM framework into images in the first place. Marco Quaggiotto is the one who, for many years, has created the GLEAMviz visual appearance. We are thankful to them for their constant inspiration and lessons in working with data visualization as an art.

We thank Dina Mistry for her invaluable comments and careful proofreading of the book manuscript.

A special thanks goes to all colleagues, students, and coauthors who have shared our journey and excitement in the science of epidemic modeling: Luca Cappa, Claudio Castellano, Ciro Cattuto, Dennis Chao, Matteo Chinazzi, Gerardo Chowell, Fabio Ciulla, Natalie E. Dean, Laura Fumanelli, Maria Litvinova, Paolo Milano, Delia Mocanu, Yamir Moreno, Evangelia (Evelyn) Panagakou, Daniela Perrotta, Piero Poletti, Diana Rojas, Michele Roncaglione, Mauricio Santillana, Mohamed Selim, Lone Simonsen, Kaiyuan Sun, Alain Jaques Valleron, Piet Van Mieghen, and Cécile Viboud. We are infinitely grateful for their outstanding support, ideas, and work.

Vibrant research environments and occasions of discussion with colleagues of all ages and expertise are the key to good science. We have been very lucky to be part of the Center for Inference and Dynamics of Infectious Diseases (CIDID), a NIH-funded Center of Excellence, the Summer Institute in Statistics and Modeling in Infectious Diseases (SISMID), and the Network Science Institute at Northeastern University and their wonderful scientific communities.

We also want to thank all those institutions that have supported our work and the completion of this book. Northeastern University in Boston and the Institute for Scientific Interchange in Turin have provided the authors of this book with great environments for their research and work. We are incredibly grateful to the Sternberg Family for the outstanding support of the MOBS Laboratory at Northeastern University. Finally we acknowledge the generous funding support at various stages of our research in the area of predictive epidemic modeling from the Models of Infectious Disease Agent Study, National Institute of General Medical Sciences Grant U54GM111274, and the Fondazione-CRT Lagrange laboratory.

.....

Boston, USA
Torino, Italy
London, UK

April 2018

About the authors

Ana Pastore y Piontti is an Associate Research Scientist in the Laboratory for the Modeling of Biological and Socio-technical Systems (MOBS Lab) at Northeastern University in Boston, USA. Her research focuses on the characterization and modeling of the spread of infectious diseases, by integrating methods of complex systems with statistical physics approaches.

Nicola Perrà is a Senior Lecturer in network science at the University of Greenwich, UK. His research interests revolve around the study of human dynamics, digital epidemiology, and network science.

Luca Rossi is a Senior Researcher at the Institute for Scientific Interchange in Torino, Italy. His research focuses on the mathematical and computational modeling of contagion processes in structured populations, in particular human transmittable infectious diseases.

Nicole Samay is a Senior Graphic Designer in the Network Science Institute at Northeastern University in Boston. She works with researchers to develop and adapt data visualizations, with a particular focus on information design and spreading processes.

Alessandro Vespignani is the Sternberg Family Distinguished Professor at Northeastern University in Boston. He is a fellow of the American Physical Society, member of the Academy of Europe, and fellow of the Institute for Quantitative Social Sciences at Harvard University. Vespignani is focusing his research activity in modeling diffusion phenomena in complex systems, including data-driven computational approaches to infectious disease spread.

This book is also a result of the special support and work at various stages of the editorial process from three fantastic collaborators:

Corrado Gioannini worked for several years in the private sector, in IT companies, developing his skills in software development and management. He is now a research leader at the Complex Systems and Networks group at ISI Foundation, where he coordinates the development of the GLEAMviz Simulator software framework.

Marcelo F. C. Gomes is a researcher on infectious disease modeling and surveillance at Fundação Oswaldo Cruz's Scientific Computing Program (Fiocruz, PROCC). His main research focus is on combining reported cases from public health agencies and human mobility for the development of risk analysis and nowcasting tools.

Bruno Gonçalves is a Moore-Sloan Fellow at NYU's Center for Data Science. With a background in Physics and Computer Science, his career has revolved around the use of datasets from sources as diverse as Apache web logs, Wikipedia edits, Twitter posts, Epidemiological reports, and census data to analyze and model human behavior and mobility. More recently he has focused on the application of machine learning and neural network techniques to analyze large geolocated datasets.

CONTENTS

INTRODUCTION XIII
OUTLOOK 201

PART I: HOW TO MODEL PANDEMICS

- 01** **INFECTIOUS DISEASE SPREADING: FROM DATA TO MODELS** 3
- Computational modeling of infectious disease spreading 5
 - Global Epidemic and Mobility model 8
- 02** **DATA, DATA, AND MORE DATA** 11
- Population 12
 - The age structure of the world population 13
 - Human mobility 14
 - Mobility patterns and epidemic spreading 18
 - Diseases 19
 - Health infrastructures 26
- 03** **DATA MODEL INTEGRATION: THE GLOBAL EPIDEMIC AND MOBILITY FRAMEWORK** 29
- Building a synthetic world 30
 - Moving people around 31
 - The disease dynamic 34
 - Synthetic epidemics 39
- 04** **FROM DATA TO KNOWLEDGE: HOW MODELS CAN BE USED** 45
- Affected countries 46
 - Invasion tree 46
 - Epidemic profiles and activity peak 49
 - Attack rate 51
 - Risk indicators and severity assessment 51
 - Containment and mitigation scenarios 53

05

THE NUMERICAL FORECAST OF PANDEMIC SPREADING: THE CASE STUDY OF THE 2009 A/H1N1 PDM 55

- Modeling the 2009 H1N1 pandemic 56
- Calibration and parameter estimates for specific outbreaks 57
- Forecast output 58
- Invasion tree 59
- Pandemic activity peaks 60
- Risk and severity assessment 62
- Vaccination campaigns 64
- Real-time forecast with computational epidemic models 65



PANDEMIC INFLUENZA 73

- Barcelona, Spain 78
- Buenos Aires, Argentina 94
- Hanoi, Vietnam 110
- Johannesburg, South Africa 142
- Melbourne, Australia 150
- New York City, USA 158



CORONAVIRUS 167

- Barcelona, Spain 170
- Guangzhou, China 174
- Jeddah, Saudi Arabia 178



EBOLA VIRUS 183

- Arua, Uganda 186
- Kisangani, Democratic Republic of the Congo 188
- Lagos, Nigeria 192

06

COMPUTATIONAL MODELING OF “DISEASE X” 197

INTRODUCTION

OUR INTENT IS TO INTRODUCE NON TECHNICAL READERS TO THE PROCESS THAT, STARTING FROM REAL-WORLD DATA, MAKES IT POSSIBLE TO DEVELOP SIMULATED SCENARIOS AND REAL-TIME FORECASTS OF THE GLOBAL SPREADING OF INFECTIOUS DISEASES.

THE FIRST SUCCESSFUL WEATHER FORECAST performed by a digital computer can be traced back to 1950. Five years later, a joint project by the US Air Force, Navy, and Weather Bureau began operational numerical weather prediction in the United States. Nowadays, computer-generated weather forecasts are at the fingertips of billions of individuals through government or commercial services, accessible by any mobile platform on the Internet. Weather forecasts can be very specific, focusing on local area models, or global. They can span a few hours or several days. They are part of our daily life and have helped popularize concepts like the “butterfly effect.”¹

A non-expert reader might reasonably think that the use of computer simulations for infectious disease outbreak forecast is as developed as the one

¹ James Gleick, *Chaos: Making a new science* (Viking Press, 1987).

regarding weather forecast; unfortunately this is far from the case. Although the birth of mathematical epidemiology dates back to the eighteenth century, public health scientists and policy-makers started only in the last 20 years or so to increasingly rely on simulated models to understand epidemiological patterns and guide control measures in real time. It may come as a surprise, but while the weather forecast research community has built thousands of weather stations around the world, put satellites in orbit, and connected large supercomputing infrastructures, the modeling of epidemic outbreaks has long suffered from the lack of near-real-time, high-quality data on populations, human and animal mobility/behavior patterns, and pathogens' biology. For instance, the 1985 pioneering work of Rvachev and Longini² on a mathematical model for the global spread of influenza had to wait on a shelf for almost 20 years before a full-fledged implementation integrating the complete International Airlines database³ saw the light. Similarly, the first cross-comparison of three different data-driven, individual-based, stochastic models of pandemic flu examining the consequences of intervention strategies in the United States had to wait until 2008.⁴

In the last 10 years, we have seen dramatic advances in data collection and availability in a number of areas, ranging from pathogen genetic sequences to human mobility patterns and social media data. These advances, often dubbed as the “big data” revolution, have finally lifted many of the limitations affecting epidemic predictive modeling. The big data paradigm is generally associated with “inductivist” approaches such as statistical modeling, phenomenological models, or machine learning-based methodologies. However, data availability also allows the development of detailed mechanistic models based on the construction of synthetic populations that statistically mimic the real world. They explicitly account for the dynamics of epidemics by calculating the future state of the system from its initial state, through time- and space-dependent equations, as well as the stochastic simulation of individual disease transmission processes.

Mechanistic approaches are clearly data hungry, and the amount of data integration depends on the scale of the model—global, regional, and local—as well as the level of detail in the population description. The latter can go down to the level of single households and specifically consider multiple transmission settings, such as schools or workplaces. It is worth stressing, however, that mechanistic models contain assumptions and approximations too. The theory and equations used to describe the system dynamics are often based

2 Leonid A. Rvachev and Ira M. Longini, Jr., “A mathematical model for the global spread of influenza,” *Mathematical Biosciences*, 75:3 22 (1985).

3 Lars Hufnagel et al., “Forecast and control of epidemics in a globalized world,” *Proceedings of the National Academy of Sciences of the United States of America* 101, 15124–15129 (2004).

4 M. Elizabeth Halloran et al., “Modeling targeted layered containment of an influenza pandemic in the United States,” *PNAS* 105, 4639–4644 (2008).

on effective or coarse-grained integration of degrees of freedom that are informed by the questions the model is set up to answer. No model fits all diseases or spans all scales and geographical resolutions.

In this context, predictive epidemic modeling is emerging as an interdisciplinary field that promises to advance the capabilities of projecting the course of an epidemic already underway or to anticipate the effectiveness of possible interventions or clinical trials. Indeed, computational modeling has been used to support responses to recent outbreaks such as the 2009 H1N1 pandemic, the 2014 West Africa Ebola outbreak, the Zika epidemic in the Americas in 2016, and the highly pathogenic avian influenza A(H7N9) in 2014 and 2015. Although predictive epidemic modeling is not yet as developed as weather forecasting, and many practical and foundational challenges still need to be addressed, the promise of computer simulations to improve epidemic preparedness and response is now recognized.⁵

The many types of descriptive and predictive models that were used during recent, large-scale outbreaks are often hidden behind the veil of technical jargon, mathematical and statistical language, and computational implementation. In order to reach out beyond the circle of practitioners and convey the transformative potential of computer simulations for public health preparedness and response, this book aims to provide a visual journey through the data and model integration process at the core of large-scale computational approaches. This overview is mostly done by using a storyboard that exemplifies data and algorithms through concrete examples and illustrations. Our intent is to introduce non technical readers to the process that, starting from real-world data, makes it possible to develop simulated scenarios and real-time forecasts of the global spreading of infectious diseases.

In its first part, this book guides the reader in the construction of the modern frameworks used to project and analyze the global spread of epidemics and pandemics. The results of these modeling activities are intuitively communicated by powerful infographics; in particular, we present examples of results obtained from numerical simulations concerning the international spreading of potentially pandemic pathogens. The second part of this book is focused on a set of pandemic charts that illustrate, through the infographic tools described in the first part, the possible scenarios of future pandemics. This atlas is meant to show commonalities and patterns in emerging health threats, as well as explore the wide range of possible scenarios that can be used by policy-makers to anticipate trends, evaluate risks, and eventually manage future events. In a nutshell, the second part of this book is a visual catalog that captures the possible evolution of future pandemics and introduces the reader to a vast range of interventions characterizing the fight against infectious diseases.

*THE PROMISE OF DATA-DRIVEN
MODELING APPROACHES
TO IMPROVE EPIDEMIC
PREPAREDNESS AND RESPONSE
IS NOW RECOGNIZED.*

⁵ National Science and Technology Council Report, "Towards Epidemic Prediction: Federal Efforts and Opportunities in Outbreak Modeling" (2016).

*THE RIGOROUS ANALYSIS AND
DISCUSSION OF PANDEMIC RISK
IS NOW ONE OF THE RESEARCH
FRONTIERS OF COMPUTATIONAL
EPIDEMIC MODELING.*

In order to exemplify numerical epidemic modeling, throughout the book we used GLEAM, the global epidemic and mobility framework,⁶ developed and supported by a team of researchers and institutions around the world. This framework is by no means to be considered prototypical; however, it integrates many of the data and concepts common to the many types of descriptive and predictive models available to the scientific community. We feel this framework conveys to a general audience the kind of work and results that can be achieved in computer simulated epidemic models. We also refer to several other models and approaches that attain the same level of complexity and results that we present here; we apologize in advance to all the colleagues that have done extraordinary work in this area if their contributions are not referenced or explained in detail. Indeed, this book is not meant as a technical review of the field, but rather as an introduction for non-practitioners to the richness of this approach, showing the potential of looking to the future of global epidemic modeling through data-driven numerical approaches.

It is also important to stress that the pandemic charts are not to be considered an exhaustive catalog of epidemic events: a full exploration of all the possible scenarios, as well as the risks associated to pandemic events, is far beyond the scope of this book and necessarily involves a large effort from the entire scientific community.

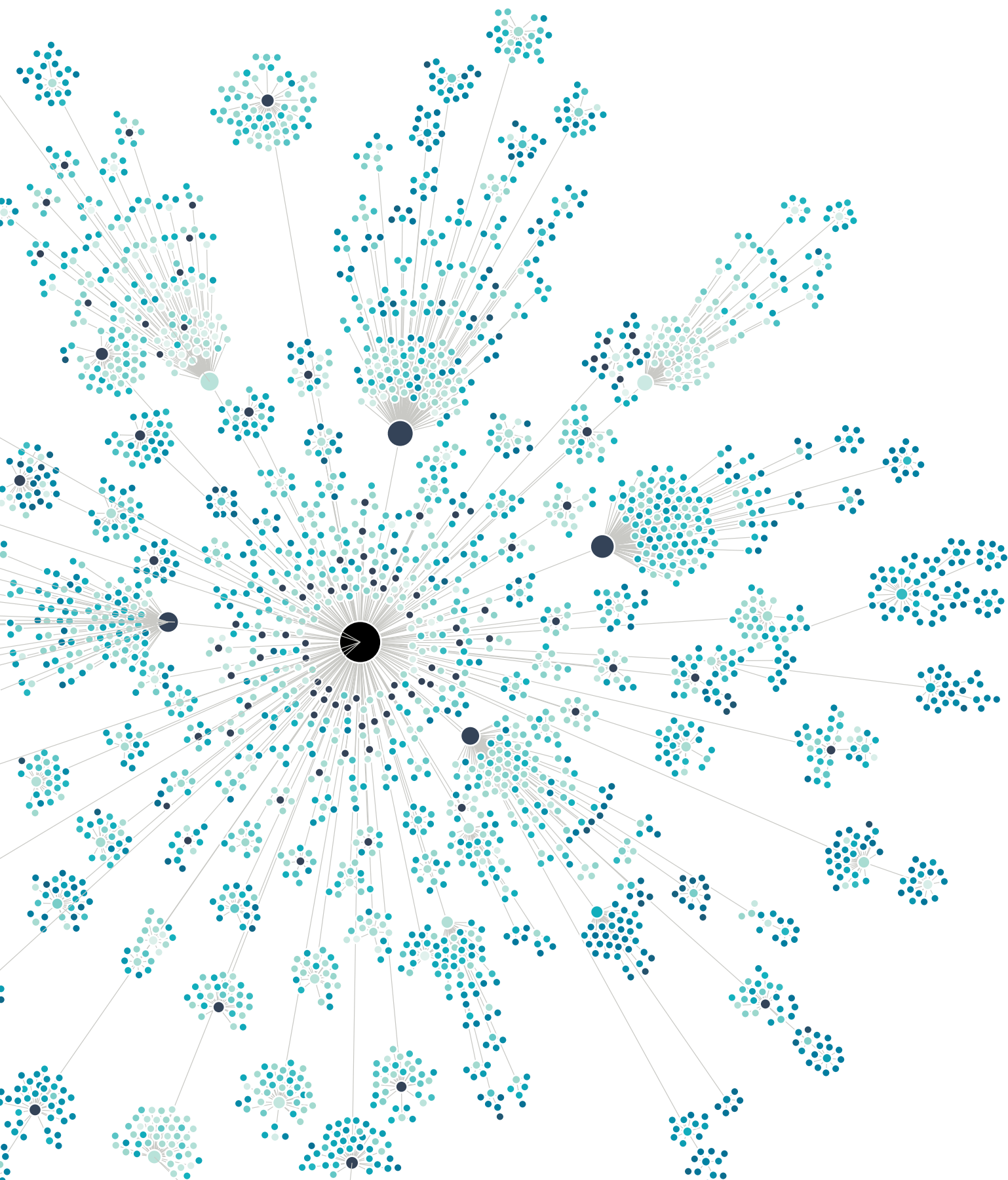
The rigorous analysis and discussion of pandemic risk is now one of the research frontiers of computational epidemic modeling, where a number of major scientific challenges still need to be addressed in the coming years, as acknowledged in the book's final outlook chapter. Indeed, we hope that this book will contribute to fueling the interest in solving these challenges and advancing numerical epidemic modeling to the point of an operational framework analogous to the one used for numerical weather forecasting.

⁶ The Global Epidemic and Mobility model, www.gleamviz.org



PART I

HOW TO MODEL PANDEMICS



CHAPTER 1

INFECTIOUS DISEASE SPREADING: FROM DATA TO MODELS

WE LIVE IN AN INCREASINGLY INTERCONNECTED WORLD where every day one billion cars take the road and more than two billion people travel each year by plane. Urbanization, growing populations, and global migrations are creating a new and complex battlefield in the fight against new and old diseases. As a result, we demand ever-increasing predictive power to anticipate future epidemic outbreaks and evaluate associated risks. In scientific terms, this power corresponds to the mathematical description of patterns found in real-world data needed to develop models that can be used to predict future events.

In the natural sciences, we are used to predicting the complex properties of new materials through precise measurements of physical quantities and the implementation of numerical models or studying the performance of a new airplane by means of computers before we even assemble one of

parts. One of the most successful examples of predictive power is that of weather forecasts, which are generated using sophisticated simulations on supercomputer infrastructures which integrate present data and huge collections of previous meteorological patterns into large systems of non-linear equations. Although weather forecasts may help us plan for our next barbecue, more importantly the science behind this allows us to prepare for extreme and disruptive natural events like tornadoes, storms, and hurricanes, saving countless lives every year.

In sharp contrast, for a long time, achieving predictive power in areas such as epidemic spreading meant confronting the insurmountable obstacle of a lack of high-quality data on human behavior and mobility at all scales. In fact, although mathematical models have been important tools in analyzing the spread and control of infectious diseases for more than a century, in most cases it was impossible to gather detailed data concerning individual and collective behavior and turn those models into real-time predictive computational tools. In the last 20 years, this foundational limitation has started to lift: indeed, every 1.2 years, more human-driven socioeconomic data is produced than during all preceding human history.¹ This data revolution has started the quest of a

¹ James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela H Byers, "Big data: The next frontier for innovation, competition, and productivity," (2011).

A MULTIDISCIPLINARY APPROACH TO EPIDEMIC ANALYSIS

01

MODELING

Elaborate stochastic infectious disease models to support a wide range of epidemiological studies, covering different types of infections and intervention scenarios.

03

COMPUTATIONAL THINKING

The computer is the laboratory. Models run on high-performance computers to create *in silico* experiments that would be hardly feasible in real systems, to guide our understanding of typical non-linear behavior and tipping points of epidemic phenomena.

02

REAL-WORLD DATA

Real-world data on population and mobility networks are integrated in structured spatial epidemic models to generate data-driven simulations of the worldwide spread of infectious diseases.

04

TOOLS DEVELOPMENT

Computational tools help in modeling the spread of a disease, understanding observed epidemic patterns, and studying the effectiveness of different intervention strategies. These tools are available to researchers, healthcare professionals, and policy-makers.

new mathematical and data-driven understanding of human networks and dynamics and finally ignited a transformative science cycle based on the following components:

- Collection, acquisition, and integration of human dynamics data from the individual to the societal scale
- Development of data infrastructures, collaborative information platforms, enabling the production of knowledge from data
- Identification of general principles and laws that characterize social complexity and capture the essence of human dynamics (data analysis and integration)
- Development of mathematical and data-driven models endowed with a high level of realism, able to offer novel quantitative understanding and predictive power in the area of socio-technical systems

*THE COMPUTER IS NOT JUST
A NUMBER CRUNCHING TOOL:
IT ALLOWS RESEARCHERS TO
SIMULATE AND STUDY SYSTEMS
THAT DO NOT FIT IN
A LABORATORY.*

The spreading of infectious diseases has been one of the first scientific areas where this research program has been put in motion. The scientific community is finally in the position to envision the development of large-scale models that, by combining theory, data, and computational thinking, can deliver real-time or near-real-time situational awareness during infectious disease outbreaks.

COMPUTATIONAL MODELING OF INFECTIOUS DISEASE SPREADING

The development of data-driven models is rooted in the combination of large-scale data mining techniques, computational approaches, and the mathematical modeling of infectious diseases. Data-driven computational approaches have solid foundations on the wealth of data that can be integrated at different scales, from the biology of the pathogen and the behavior of the single individual to the “social aggregate,” where the “social aggregate” is a large-scale social system consisting of millions of individuals whose dynamics can be characterized in space (geographic and social) and time. Those data are integrated in mathematical epidemic models, which are adapted to deal with the multi-scale and complex nature of the real world. In this context, the mathematical and statistical modeling framework has evolved from simple compartmental models to structured approaches in which the heterogeneities and details of the population and system under study are becoming increasingly important features (FIGURE 1.1). Modeling approaches explicitly include spatial structures consisting of multiple sub populations coupled by traveling fluxes, while the epidemic within each sub population is described according to approximations depending on the specific case studied. This patch, or meta population modeling framework, has then grown into a multi-scale framework in which the various possible granularities of the system (country,

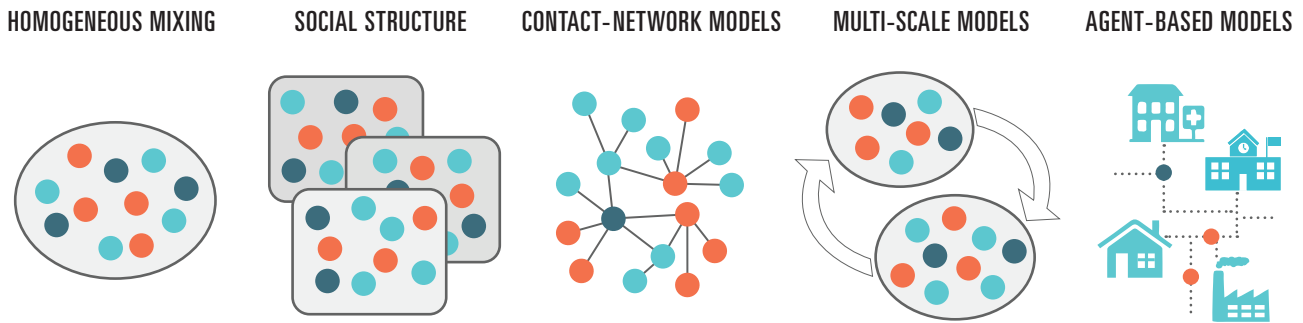


Figure 1.1 | Different scale structures used in epidemic modeling

Circles represent individuals and each color corresponds to a different stage of the disease. From left to right: homogeneous mixing, in which individuals are assumed to homogeneously interact with each other at random; social structure, where people are classified according to demographic information (age, gender, etc.); contact-network models, in which the detailed network of social interactions between individuals provides the possible disease propagation paths; multi-scale models, which consider sub population coupled by movements of individuals, while homogeneous mixing is assumed on the lower scale; agent-based models, which recreate the characteristic movements and interactions of any single individual on a very detailed scale (a schematic representation of a city is shown).

Adapted from V. Colizza et al., C. R. Biologies 330 (2007)

inter city, and intra city) are considered through different approximations and combined with interaction networks describing the flows of people and/or animals. At the most detailed level, the introduction of agent-based models (ABMs) has enabled the usual modeling perspective to stretch even further by achieving a full description of the society and a complete characterization (household, workplace, etc.) of each individual.

The advances in data-model integration have highlighted complex properties and heterogeneities, which cannot be neglected in the description of epidemics. Although these characteristics have long been acknowledged as relevant factors in determining the properties of dynamical processes, it is clear now that the large-scale epidemics often elude the straightforward linear thinking we are used to and surprise us with tipping points, emergent behaviors, and unexpected shifts in dynamical regime that characterize complex phenomena. In this context, the computer plays a fundamental role; on the one hand, it allows the creation of *in silico* experiments hardly feasible in real systems, and on the other hand, it can access and compare quantities and observables across many different models. This computational approach is the guide for understanding typical non-linear behaviors and tipping points not accessible by analytical means. In this way, the computer is not just a number crunching tool: it allows researchers to simulate and study systems that do not fit in a laboratory, such as the traffic patterns in a big city or the spreading of a new pandemic. In this perspective, it has been crucial to introduce computational algorithms that rely on the repeated computation of stochastic processes. These techniques, generally named Monte Carlo methods or stochastic algorithms, are especially useful in studying systems with a large number of coupled degrees of freedom and where stochastic effects are a fundamental component of the system. While these algorithms have been first used in physical sciences, their use in the area of human systems, economics, and social sciences has become increasingly popular. The possibility of handling significant uncertainty in inputs and occurrences allows us to finally study large-scale systems where stochasticity is an intrinsic element of the system.

1.1 INFOBOX

DATA-INTENSIVE COMPUTATIONAL TOOL FOR THE ANALYSIS OF INFECTIOUS DISEASES

The Global Epidemic and Mobility (GLEAM) model is a publicly available computational framework based on a metapopulation approach in which the population of the world is spatially structured into geolocalized patches or subpopulations (e.g., cities) where individuals mix. These patches are connected by the mobility patterns of individuals. GLEAM is capable of generating stochastic simulations of epidemic spread worldwide, yielding (among other measures) the incidence and seeding events at a daily resolution for 3,253 subpopulations in 232 countries and dependent territories (www.gleamviz.org).

Other high-performance computational tools are also available to the public for the spatial analysis and modeling of epidemics. These tools differ in their underlying modeling approaches and in the implementation, flexibility, and accessibility of the software itself.

The Spatiotemporal Epidemiological Modeler (STEM) is a modeling system for simulating the spread of an infectious disease in a spatially structured population. Contrary to other approaches, STEM is based on an extensible software platform, which promotes the contribution of data and algorithms by users (www.eclipse.org/stem).

Agent-based models describe the stochastic propagation of a disease at the individual level, thus taking into account the explicit social and spatial structure of the population under consideration. In this respect, CommunityFlu is a software tool that simulates the spread of influenza in a structured population of approximately 1,000 households with 2,500 persons (www.cdc.gov/flu/tools/communityflu).

A larger population is considered in FluTe, a publicly available tool for the stochastic simulation of an epidemic in the United States at the individual level (www.cs.unm.edu/~dlchao/flute). The model is based on a synthetic population, structured in a hierarchy of mixing social groups, such as households, household clusters, neighborhoods, and nation-wide communities.

The Bruno Kessler Foundation hosts the most detailed European-wide agent-based model for infectious diseases (dpcs.fbk.eu). The model is informed by routine socio demographic data collected throughout all European countries (e.g., school and workplace attendance, household structure, etc.).

EpiFast involves a parallel algorithm implemented using a master-slave approach, which provides for scalability on distributed memory systems. This tool allows for the detailed representation and simulation of a disease in social contact networks among individuals that dynamically evolve over time and adapt to actions taken by individuals and public health interventions (www.vbi.vt.edu).

FRED (A Framework for Reconstructing Epidemiological Dynamics) is an open source modeling system developed by the University of Pittsburgh Public Health Dynamics Laboratory. The system uses agent-based modeling derived from census-based synthetic populations that capture the demographic and geographic distributions of the population, as well as detailed household, school, and workplace social networks (www.phdl.pitt.edu/index.php/research/software18).

Computational modeling has led to a qualitative change in the ways we model epidemic and social contagion processes. Visualization and analysis tools that are able to cope with multiple levels of representation are being developed along with computer simulations that provide experiments not feasible in the real world. For the first time, epidemic processes can be studied in a comprehensive fashion that addresses the complexity inherent to real-world health problems. Whereas data availability is pointing out the limits of our conceptual and modeling frameworks, it is also allowing the validation of results across different modeling approaches, mathematical techniques, and approximation schemes.

GLOBAL EPIDEMIC AND MOBILITY MODEL

In this book, we want to illustrate the methodology described earlier and convince you that this approach is mature enough to be encapsulated in actionable tools for the simulation of case studies, the analysis of risk through model scenarios, and the forecasts of newly emerging infectious diseases (**INFOBOX 1.1**).

In order to tell this story, we use the Global Epidemic and Mobility (GLEAM) framework. GLEAM combines real-world data on populations and human mobility with complex stochastic models of disease transmission. This computational model is the product of the multidisciplinary work of a 10-year collaboration among several teams across the world. The model has been integrated in a computational platform that can model the world wide spread of epidemics for human transmissible diseases, offering extensive flexibility in the design of the compartmental model and scenario setup, including computationally optimized numerical simulations based on high-resolution global demographic and mobility data.

In the following chapters, we show the reader the what, how, and why of the computational approaches to modeling infectious diseases. First, we present what data we can leverage in informing state-of-the-art modeling approaches. Next, we show how the data-model integration is made; in other words, we take a look under the hood of the computational modeling approach. Finally, and most importantly, we make the case for why computational modeling is important to support public health decisions and intervention plans. Indeed, computational modeling does more than forecast and provides rationales and quantitative analysis in a number of areas that we briefly summarize here.

Preparation and contingency planning

The decision-making process in the fight against diseases relies on a multitude of options such as vaccination, contact tracing, quarantine, administration of drugs, and social distancing, which must be implemented in a manner that addresses the complexity inherent to the biological, social, and behavioral

*COMPUTER SIMULATIONS
PROVIDE EXPERIMENTS NOT
FEASIBLE IN THE REAL WORLD
AND ADDRESS THE COMPLEXITY
INHERENT TO REAL-WORLD
HEALTH PROBLEMS.*