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

Springer

Authors
Ana Pastore y Piontti
Northeastern University
Boston, MA
USA

Nicola Perra
University of Greenwich
London
UK

Luca Rossi
Institute for Scientific Interchange
Torino
Italy

Nicole Samay
Northeastern University
Boston, MA
USA

Alessandro Vespignani
Northeastern University
Boston, MA
USA

Collaborators
Corrado Gioannini
Institute of Scientific Interchange
Torino
Italy

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

Bruno Gonçalves
Center for Data Science
New York, NY
USA

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

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 Perra 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

05

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

PART II: PANDEMIC CHARTS

PANDEMIC INFLUENZA 73

CORONAVIRUS 167

EBOLA VIRUS 183

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 Ravchev and Longini[2] 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 database[3] 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.

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

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.

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