Pandemic Informatics: Preparation, Robustness, and Resilience

A Computing Community Consortium (CCC) Quadrennial Paper

Elizabeth Bradley (University of Colorado Boulder), Madhav Marathe (University of Virginia), Melanie Moses (The University of New Mexico), William D Gropp (University of Illinois Urbana-Champaign), and Daniel Lopresti (Lehigh University)

Overview

Infectious diseases cause more than 13 million deaths a year, worldwide. Globalization, urbanization, climate change, and ecological pressures have significantly increased the risk of a global pandemic. The ongoing COVID-19 pandemic—the first since the H1N1 outbreak more than a decade ago and the worst since the 1918 influenza pandemic—illustrates these matters vividly. More than 47M confirmed infections and 1M deaths have been reported worldwide as of November 4, 2020 and the global markets have lost trillions of dollars. The pandemic will continue to have significant disruptive impacts upon the United States and the world for years; its secondary and tertiary impacts might be felt for more than a decade.

An effective strategy to reduce the national and global burden of pandemics must: (*i*) detect timing and location of occurrence, taking into account the many interdependent driving factors; (*ii*) anticipate public reaction to an outbreak, including panic behaviors that obstruct responders and spread contagion; and (*iii*) develop actionable policies that enable targeted and effective responses. These three aims will require advances in a number of areas, including:

- The development of models that are not just scientifically effective, but that support understanding on the part of the public, as well as actionable insights for policy makers.
- Identification and preparation of computational and data resources (data, computational power, expertise) that will allow us to respond quickly and predict effectively in a crisis situation.
- Real-time collection and updating of data, models, and model assumptions in rapidly changing environments.

These are not purely technological problems. Effective preparation for and response to future pandemics will require integration of solutions that span the full sociotechnical spectrum of challenges that are posed by these devastating events. This will require systemic, national-level support and a coordinated effort by the computing research community, in tandem with a broad coalition of experts from the social and political sciences, economics and the humanities. Such a framework will allow us to develop an understanding across scales, from cells and RNA to epidemic spread through communities and across countries. Only with such a comprehensive understanding will we be prepared to more effectively manage the next pandemic.

Pandemic Modeling

The first COVID-19 models communicated the extraordinary rates of illness and death that could result from an uncontrolled pandemic. That understanding led many governments to implement control measures early enough to mitigate the disease's impact and save millions of lives.

However, models are simplifications of a very complex reality. For instance, age, socio-economic status, race, and environmental conditions create disparities in infection and mortality rates, but these realities are often not incorporated in the models that have informed policy in the COVID pandemic. More broadly, few epidemiological models factor in antimicrobial resistance, zoonosis, climate change, or increased urbanization, all of which will play increasingly important roles in future pandemic outbreaks.

Models and modeling frameworks that incorporated these realities could help us more effectively prepare for, and respond to, pandemics—e.g., identifying targeted interventions that will save the most lives, designing methods for allocation of scarce resources, forecasting the trajectory of the pandemic, and understanding the complex interactions between the pandemic and the intertwined social, political and economic systems of the 21st century. For example, pandemic spread can be stopped by having everyone stay home for an extended period of time (as was done in Wuhan), but this strategy is not feasible in open societies. Balancing the response with economic activity, social interactions, and political realities remains a significant challenge.

Foundations exist for the next generation of pandemic models. Compared to the coupled differential equations originally used to simulate how a virus spreads over time through homogenous populations, network models can easily incorporate more-realistic interactions among heterogeneous groups of people. Agent-based (networked) models work at an even finer grain, across time and space.

These sophisticated modeling approaches can be transformative: they can predict the course of the pandemic and the impact of various interventions, including testing and social distancing at different scales—although those predictions rapidly become outdated because of the effects of the interventions. Even more importantly, they are potentially more explainable, which is critical both to the public and to decision makers. Significant challenges remain, however:

- How to build models that are robust in the face of different assumptions, and that effectively communicate uncertainty in projections as assumptions are violated.
- How to gather and incorporate relevant data in real time to validate past predictions, update future projections and actively learn when modeling assumptions cause models to fail to capture real-world dynamics.
- How to model the sensing and monitoring systems that gather these data (e.g., testing and syndromic surveillance).
- How to understand and model the evolution of the pathogen in space and time, including its interaction with humans and their immune systems, as well as the effects of interventions and policies.

- How to rapidly incorporate the changing scientific understanding of the disease into the models and their underlying assumptions—and how to know when there is a need to do so.
- How to ensure that the models are sufficiently transparent and explainable so that the general population and policy makers understand how their actions and behaviors protect or expose themselves, their families, and their communities.
- How to incorporate socioeconomic factors such as occupation, race, income, and access to affordable health services that drive inequities in infection, hospitalization and death rates.
- How to effectively incorporate human behavior into the models, in the context of pandemic spread within economic and sociopolitical systems.
- How to validate these models. Traditional ways of using retrospective and predictive validity are rarely useful in crisis situations, when both the data about the situation, and our understanding of that situation, are limited.
- How to identify and balance possible tradeoffs between mitigation, privacy, security, and intellectual property in the face of public health emergencies.

Importantly, models should not focus narrowly on reducing epidemic spread, but rather be embedded in frameworks that target broader socio-economic impact. The models should also address growing trends in urbanization as well as climate change. They should not only incorporate known factors, but also reveal hidden variables and previously unknown factors—and support mitigations that reduce the risks of **all** causal variables.

This complex and important research agenda will require significant effort from the computing research community, in collaboration with experts from the social, behavioral, epidemic, and economic sciences, with sustained support at the national level and a strong, national-level infrastructure, as detailed in the Recommendations section at the end of this document. (See also the companion CCC Quad Paper entitled "The Rise of AI-Driven Simulators: Building a New Crystal Ball," which discusses the ways in which artificial intelligence strategies can aid in the modeling of complex problems).

Infrastructure for Pandemic Informatics

In a crisis, it is essential to have immediate access to enough computing power, data sources, and expertise to bring state-of-the-art methods to bear on the associated problems. In the current pandemic, the COVID-19 High Performance Computing (HPC) consortium has shown the value of advanced computing resources. However, this consortium was an ad hoc effort that required significant effort to establish and operate, and any delay can be critical in the context of a pandemic. Moreover, computing power is only one part of the solution. An even more fundamental need is information: comprehensive, clean data about all of the salient features of the situation, modified automatically and dynamically as that situation evolves, and distributed in a manner that preserves individual privacy. Computational infrastructure to manage data associated with vaccines, treatments, and long-term health effects of COVID-19 will be a particularly important element here. Finally, there is a need for experts in the applications, software, data, and system operations to be "on call" to minimize the barriers to entry for these resources.

Models and resources are only part of the landscape for pandemic preparation, resilience, and robustness, of course; research is also needed into privacy-preserving contact tracing; algorithms for optimal pooled testing; strategies to mitigate the effects of mis- and disinformation; understanding of the short- and long-term sociotechnical issues that arise when societies move to a distributed mode of life, learning, and work; and decision-support systems that help people make sense of the massive amounts of data from the world and from the models.

Recommendations

Bringing together the resources needed to support crisis response in the inevitable next pandemic, as well as, to carry out forward-looking research that will allow us to anticipate and perhaps divert its onset, would address many of these challenges thereby creating major benefits for the United States and the world. To that end, we recommend a multi-pronged approach that combines distributed and centralized resources, as well as, coordinated action from agencies, foundations, corporate research labs, academic departments, and national labs.

As a foundation for this, we recommend the creation of a distributed, pervasive *National Pandemic Informatics Infrastructure*. A dedicated *Pandemic Informatics Institute,* staffed on a full-time basis by a strong, interdisciplinary group of scientists, engineers, and support staff, would act as the nexus of this infrastructure, serving as a coordinator and central source of resources, best practices, etc. Critical resources—data sets, software, testbeds, and computational resources, as well as the essential expert support for accessing and using those resources—would be provisioned at various infrastructure sites across the nation to support both ongoing research regarding pandemics and the rapid ramp-up that accompanies these events. As highlighted above, pandemic response is a sociotechnical problem, and one in which policy plays a critical role. This has implications for the infrastructure: e.g., the development of privacy-preserving policies for rapid data sharing among various groups during crisis, or providing useful information to decision makers.

The research challenges outlined above fall at the boundaries between computing and other areas of work, notably the social sciences. The second element of our recommendations is a coordinated, multidisciplinary, multi-agency research effort to foster the necessary advances at these boundaries. Importantly, funding for this should not be diverted from existing agency budgets; rather, this demands wholly new funding streams. At the same time, agencies and foundations should continue the recently initiated programs that target pandemic informatics, extending them indefinitely and prescribing award levels and time horizons that allow for sustained work on these difficult problems. Importantly, all of these efforts must be deeply and fundamentally interdisciplinary, bringing together computing researchers with scientists, humanists, social scientists, and political scientists. A multi-agency initiative that brought together the NSF, DOE, DOD, CDC, IARPA, and the NIH, among others, would be useful in accomplishing this at a high level, but finer-grained policies and practices are also essential: e.g., requiring co-PIs from computing and the social sciences on a solicitation about models that incorporate human behavior.

This white paper is part of a series of papers compiled every four years by the CCC Council and members of the computing research community to inform policymakers, community members and the public on important research opportunities in areas of national priority. The topics chosen represent areas of pressing national need spanning various subdisciplines of the computing research field. The white papers attempt to portray a comprehensive picture of the computing research field detailing potential research directions, challenges and recommendations.

This material is based upon work supported by the National Science Foundation under Grant No. 1734706. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

For citation use: Bradley E., Marathe M., Moses M., Gropp W. & Lopresti D. (2020) Pandemic Informatics: Preparation, Robustness, and Resilience.
https://cra.org/ccc/resources/ccc-led-whitepapers/#2020-quadrennial-papers