Data Analytics, Verification and Validation Approaches for Epidemiological Models

Organizers:
Arvind Ramanathan (ramanathana@ornl.gov), Oak Ridge National Laboratory;
Laura L. Pullum (pullumll@ornl.gov), Oak Ridge National Laboratory

Abstract

This mini symposium will highlight the synergistic opportunities that exist in integrating data analytics approaches with rigorous verification and validation techniques for large-scale epidemiological models. In particular, it will focus on recent advances in the areas of machine learning, applied mathematics and model checking to address the data analysis, verification and validation requirements for epidemiological models. Although a number of epidemiological models are widely used to understand and predict how infectious diseases spread within heterogeneous populations, it has been challenging to reliably quantify the predictions made by these models to enable decision makers to take calculated risks in intervention strategies. As these models are being augmented with a variety of non-traditional sources of data to track, predict and control infectious disease spread, there is a need to develop novel data analytic tools to integrate information from heterogeneous resources and verify/validate these predictions from observations, including historical datasets and current events that can help assign rigorous confidence intervals against these predictions.

Session 1:

- **Speaker:** *Madhav Marathe* (mmarathe@vbi.vt.edu), Dept. of Computer Science, Network Dynamics & Simulation Science Laboratory, Virginia Bio-Informatics Institute, Virginia Tech

  Title: Scalable Agent-Based simulations for Epidemiological Models

  Abstract: Agent-based simulation is an example of an interaction based computational technique useful for reasoning about biological, information and social networks. Developing scalable agent-based models raises important computational and conceptual issues, including computational efficiency, necessity of detailed representation and uncertainty quantification. In this talk, I will describe our experience in developing scale agent-based simulations to study large socially-coupled systems. In describing these systems, we will focus issues related to validation and verification that arise when developing such models.

- **Speaker:** *Sreenivas Sukumar* (sukumarsr@ornl.gov), Computational Science and Engineering Division, Oak Ridge National Laboratory

  Title: An Information Theoretic Approach to Validate Disease Spread Models

  Abstract: This paper is motivated by the need to design model validation strategies for epidemiological disease-spread models. We consider both agent-based and equation-based models of pandemic disease spread and study the nuances and complexities one has to consider from the perspective of model validation. For this purpose, we instantiate an equation-based model and an agentbased model of the 1918 Spanish flu and we leverage data published in the literature for our case study. We present our observations from the perspective of each implementation and discuss the application of model-selection criteria to compare the risk in choosing one modeling paradigm to another. We conclude with a discussion of our experience and document future ideas for a model validation framework.

- **Speaker:** *Christopher L. Barrett* (cbarrett@vbi.vt.edu), Network Network Dynamics & Simulation Science Laboratory, Virginia Bioinformatics Institute, Dept. of Computer Science, Virginia Tech

  Title: Believability and Usefulness: Policy Informatics in the age of social data

  Abstract: Pervasive access to previously unavailable economic, social, behavioral and infrastructure-related information, as well as ICT-driven transformation of more traditional information access, changes informatics-based support of policy development and decision making. Emphasis on integration, interaction, action/state

branching and a post-modeling" perspective of decision informatics impose related changes on many analytical approaches and assumptions, including validation. This presentation begins with an example of a contrast of two policy decision problems that share a desired result: informatics supported determination of a best policy in a complex public health environment. Conventional notions of predictive validity of models are not well suited to these studies or methods. New ideas, and recalling some old ones, are required to underwrite the believability and usefulness of informatics-based support of analysis of massively interacting systems. Some approaches using synthetic information systems are introduced.

- **Speaker: Xiaohui Cui** (cuix@nyit.edu), Computer Science Department, New York Institute of Technology

  Title: Crawfish: A Platform for Testing Scenarios for Agent-Based Disease Spread Models

  Abstract: In this presentation, we describe Crawfish, an agent-based disease spread model scenario test platform. The platform utilizes different metamorphic relationships to test epidemic disease spread models. Our experiments demonstrate that Crawfish can not only be used for testing existing models, it can also provide the capability to help developers locate development bugs during the simulation development stage. Some of these logic bugs are difficult to identify using other test methods.

  Joint work with: Laura Pullum (ORNL), Moriel Schottlender (NYIT), Sanem Ucan (NYIT), and Qingwen Xiong (NYIT)

Session 2:

- **Speaker:** *Aditya B. Prakash* (badityap@cs.vt.edu), Computer Science Department, Virginia Tech

  Title: Understanding and Managing Cascades on Large Graphs

  Abstract: How do memes spread over blogs? Which group should we market to, for maximizing product penetration? Who are the most likely culprits of an epidemic? Will a given YouTube video go viral? Who are the best people to vaccinate? What happens when two products compete? In this talk we will focus on modeling and manipulating such propagation-style processes on large networks and present a multi-pronged approach, which includes: (a) Theoretical results on the behavior of fundamental models; (b) Scalable Algorithms based on these processes e.g., immunization, finding 'culprits' etc.; and (c) Empirical Studies of diffusion on blogs and on-line websites like Twitter. We finally conclude with future research directions. The problems we focus on are central in surprisingly diverse areas: from cyber-security, epidemiology and public health, product marketing to information dissemination.

- **Speaker:** *Chakra S. Chennubhotla* (chakracs@pitt.edu), Department of Computational & Systems Biology, University of Pittsburgh

  Title: Markov Models for Multi-Scale Network Community Detection

  Abstract: Detecting clusters or communities in large scale graphs is active area of interest. In this work, we present a Markov model of multiscale network communities. This approach is based on modeling the data as a weighted graph and initiating a random walk on the network. As diffusion probabilities homogenize, an implicit clustering emerges. At each level of the hierarchy, we choose a set of nodes to represent the implicit clusters, and re-initiate Markov diffusion over those nodes to build a multi-scale representation of the network communities. Unlike other community detection algorithms in the literature, no a priori predictions for the total number of communities is required.

  Joint work with: Arvind Ramanathan (ORNL), Virginia Burger (University of Pittsburgh), Shannon Quinn (University of Pittsburgh), Andrej Savol (University of Pittsburgh), Chad Steed (ORNL), Christopher Maness (ORNL)

- **Speaker:** *Sumit K. Jha* (jha@eecs.ucf.edu), Computer Science Division, University of Central Florida

  Title: Formal methods for analyzing epidemiological models

  Abstract: Computational modelling in epidemiology involves the use of a variety of modeling frameworks, including ordinary differential equations, stochastic differential equations, and agent-based models. When such multi-outcome models are used for prediction and planning purposes, the ability to characterize the inherent uncertainty in the outcome of these models is fundamentally important. In this talk, I will discuss the use of formal methods in analysing epidemiological models with a focus on Bayesian statistical model checking and temporal logics.

- **Speaker:** *Ozgur Ozmen* (ozmeno@ornl.gov), Computational Science and Engineering Division, Oak Ridge National Laboratory

  Title: A Taxonomic Study of Disease-spread Models and Simulations

  Abstract: The presentation is motivated by the need to identify underlying modeling principles related to domain-specific knowledge in epidemiological modeling. The aspects a model incorporates depend on the intended purpose of a model. Modeling assumptions differ based on these aspects; hence verification and validation approaches are influenced. In this talk, an initial taxonomy of disease-spread models is introduced, coupled with a taxonomy of epidemiological simulation studies to explore important processes involved in modeling these events.