CSE 8803 EPI: Data Science for Epidemiology

Prakash, Fall 2022, Georgia Institute of Technology

1. General Information

Instructor: Prof. B. Aditya Prakash, School of CSE, College of Computing
Contact: Please email badityap@cc.gatech.edu, with 'CSE 8803' in the subject
Course website: https://www.cc.gatech.edu/~badityap/classes/cse8803-epi-Fall22/
Class Timings: Tuesdays and Thursdays, 5-6:15pm, Van Leer C456

2. Description

The devastating impact of the currently unfolding global COVID-19 pandemic and those of the Zika, SARS, MERS, and Ebola outbreaks over the past decade has sharply illustrated our enormous vulnerability to emerging infectious diseases. Examples of fundamental questions studied by epidemiologists and public officials during these epidemic outbreaks, include: (1) Where did the disease originate? (2) How does it spread, and how does it compare with earlier outbreaks? (3) Where has it spread so far? (4) How can the disease spread be controlled? (5) What are its impacts?

Data science and machine learning have an important role to play in this regard. Coupled with increasing data generation across multiple domains (like electronic medical records and social media), there is a clear need for analyzing them to inform public health policies and outcomes. Recent advances in disease surveillance and forecasting, and initiatives such as CDC forecasting projects have brought these disciplines closer --- public health practitioners seek to use novel datasets and techniques whereas researchers from data mining and machine learning develop novel tools for solving many crucial problems in the public health policy planning process.

This course will cover foundations of computational+networked epidemiology and data science algorithms+systems in context of public health applications. The objective of the course is to introduce students to this emerging multi-disciplinary domain. The course will touch the following from an application viewpoint: (a) Foundations of modeling disease dynamics, (b) Calibration, Surveillance and Forecasting of disease spread (c) Detection, Reverse-engineering and Control, and (d) Additional topics such as Phylodynamics, Disparities, Tracing and Data collection. From a methodology viewpoint the course will feature (i) Non-linear systems, (ii) Network algorithms, (iii) Stochastic Optimization, (iv) ML and neural models for spatio-temporal, graphs and social media data, (v) HPC simulations, and (vii) Visualization techniques.

Students will be expected to participate in class discussion, submit assigned HWs and work on a project. Problems and datasets related to recent outbreaks will be used in some of the course topics and projects. There will also be a focus on problems of intense current interest for combatting COVID-19.

3. Background and Prerequisites

This is a graduate course and will be highly multi-disciplinary, and all the topics will span multiple areas. Students should have background in **at least one** of the following areas: data analytics, network science, algorithms, parallel computation, mathematical modeling, statistics and optimization. Students will be encouraged to work in teams with complementary expertise, allowing them to explore new areas. Programming proficiency in at least one of Matlab, R, Python, C++, Java needed.

4. Tentative Syllabus

Epidemic Models

Dynamical Analysis of Models

Network Construction

Reverse Engineering and Inference

Outbreak Detection

Surveillance

Forecasting

Resource Allocation and Distribution

Interventions

Ethics, Disparities and Other topics