CS:4980 Computational Epidemiology, Spring 2020

Project Overview (Draft)

Deliverables

Here are the *deliverables* for the project, along with a time line.

1. **Pre-proposal:** Due 3/3.
   Should contain (a) background, (b) problem statement, and (c) methods the project might use to solve the problem, including possible data sources and pointers to literature. At most 1 page long.

   Instructors will do a "sanity check" on pre-proposals and provide quick feedback.

2. **Proposal:** Due 3/12.
   Should be an updated version of the pre-proposal, updated based on instructors' feedback. This document will serve as a guide for the project. Also, at most 1 page long.

3. **Paper draft 1:** Due 4/16.
   Should be a well-written, technical paper containing 5 sections: (a) an *introduction* that includes background and problem statement, (b) *methods*, including description of data, (c) preliminary *results* with appropriate plots, graphs, and other visuals, (d) *discussion*, including an interpretation of results, and (d) *limitations* of work. At most 6 pages long.

   Instructors will review and provide feedback.

4. **Paper draft 2:** Due 4/30.
   Should be updated in response to instructors' feedback. This draft of the paper will be reviewed by your "peers" in a double-blind review process. At most 6 pages long.

5. **Peer-reviews:** Due 5/7.
   Each of you will be assigned some number of papers to write reviews for. Each review needs to be at most 1 page long. More details on this process will be provided later.

6. **Project presentation and final paper:** During finals week 5/11-15
   Each group will have a 10-min slot to present their work. Final project reports (at most 8 pages long) are also due at this time

Topics

The course content is organized into the following *topics*. Your project will likely fall into one of these categories.

- **Compartmental modeling and simulation:** The Wuhan Coronavirus (Covid-19) modeling paper you just read is an excellent example of this. The main goal of such a project would be to predict total number of cases, infection peak, etc. Such a project would be evaluated on the merits of modeling choices and how data might have been used to learn model parameters.
- **Contact network modeling and simulation:** A project in this space could be similar to a compartmental modeling project, expect that individuals are modeled as nodes in a network and their contact patterns are modeled via edges. Contact networks could be "real," i.e., obtained from data or they could be synthetic, i.e., generated by an algorithm. Such a project would need to reveal something that could not have been revealed by just using compartmental models.
- **Inference problems:** Using disease incidence data (e.g., number of cases, where they occur), we can learn something about the disease (e.g., R_0, period of infectivity). If the underlying model is assumed to be a contact network, we can learn aspects of the flow of infection such as where the infection started and which edges the infection passed along. We could also learn about risk factors for specific diseases. Such a project will likely need disease incidence data and some understanding of data mining and/or machine learning methods.
- **Modeling and evaluating interventions:** In planning for a potential Covid-19 outbreak, the CDC is considering the impact of "social distancing" measures such as school closures, voluntary isolation, voluntary quarantining, etc. In projects in this space, you would model and evaluate interventions such as these and others such as vaccinations.
- **Surveillance:** The CDC is also planning for potential Covid-19 surveillance sites in Chicago, NYC, LA, and San Francisco. Surveillance methods should be low-cost, but at the same time should provide early detection and should "cover" a large fraction of the cases. A project in this space would model the problem of doing low-cost surveillance that provides early-detection of broad coverage and evaluate different surveillance strategies.

Things to think about

You should choose a topic that interests you. Besides interest in a topic, there are a number of other things to consider, in choosing a project.

- What data do I need? How will I get this data?
- What coding skills (e.g., Python, R, Matlab, Tableau) will I need? Do I have these skills? Do my group-mates have these skills?
- What data mining or machine learning skills do I need? Do I know how to use Python packages such as SciPy or Scikit-learn? Or can I learn these quickly? Do my group-mates have these skills?
- Can I read and understand the papers that seem to have the information I need? Do I have the math or statistics knowledge to extract what I need from these papers? What about my group mates?
- Can I write a technical paper? If I am not a good technical writer, do I know where to go for help? What about my group mates?
