Thesis Project Portfolio

Investigating the Effectiveness of Wastewater Surveillance Through Computer Simulation
(Technical Report)

The Pitfalls of Predictive Modeling: Investigating How Inaccurate Models Can be Useful Through the Lens of Data Relativity
(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Colin Crowe

Fall, 2022

Department of Computer Science

Table of Contents

Sociotechnical Synthesis
(Executive Summary)

Improving Disease Monitoring Through Effective Use of Computer Modeling

Computer modeling as a field focuses on creating simulations of complex systems to improve our understanding of how they work. Such simulations gained additional importance at the onset of the COVID-19 pandemic, which caused researchers and policymakers alike to turn towards computer models as a means of assessing the impact of the disease and predicting the effectiveness of possible responses. Both halves of my thesis focus on this application of computer modeling. For my technical report, I developed an SIR model (a model composed of agents who are either susceptible, infected, or recovered) that would simulate disease spread through a small community. The data this model created was then used to evaluate the effectiveness of wastewater surveillance, a method of disease control that involves taking samples from sewer water and testing them for traces of disease. Insights from computer models can be difficult to directly translate into the real world, however, owing to simplifying assumptions made during the model creation process. My STS research therefore focuses on determining how exactly the results of computer modeling should be used to create effective real-world policy, given that their creation involves subjectivities and uncertainties.

Beginning with the technical topic, this half of the report focuses around creating an SIR model to test the effectiveness of wastewater surveillance. Monitoring wastewater has limitations because, while obtaining an infected sample is an indication that someone has the disease, it can be difficult to trace that sample back to any specific individuals who need to quarantine. Importantly, unlike in the real world, creating a computer model allows me to know exactly who

in the simulation is infected and who isn't at any given time. This means that any strategy for "reverse-engineering" the wastewater data into information about individual infections can be easily verified. While my model did not produce any specific strategies for using wastewater surveillance, it did yield insights into what conditions lead to the most effective sampling. Specifically, I found that in systems where individuals were relatively isolated (simulating quarantine conditions), detecting traces of disease in sewer water was highly effective in that it implied only a few individuals were likely to be infected. In contrast, systems in which individuals had many connections between each other (simulating non-quarantine conditions) displayed the opposite results; sampling was ineffective because detecting traces of disease seemed to imply that almost anyone could be infected, rather than a few specific people (Fig 1).

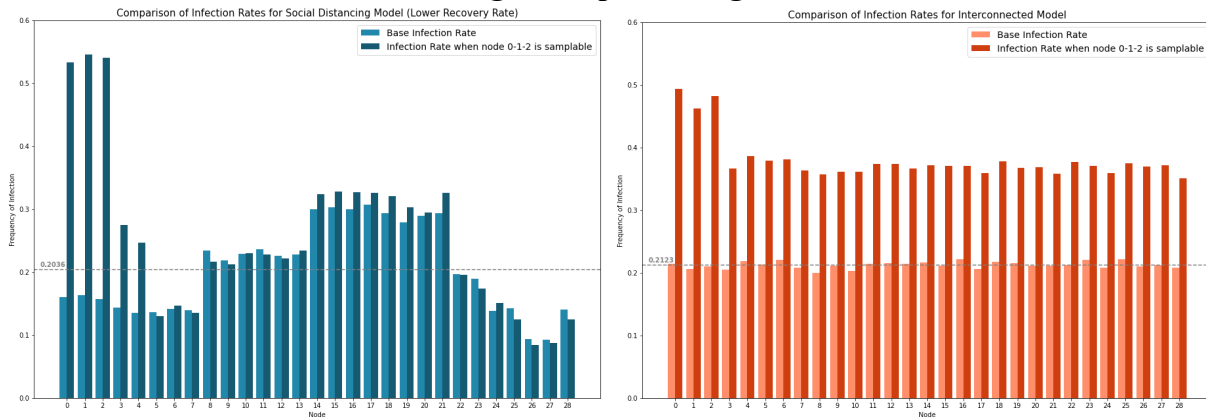**Information Gain From Collecting a Sample During Quarantine vs. Non-Quarantine**

Fig 1: The difference between light color bars and dark color bars represents information gained from a sample. In the blue graph (quarantine), only some nodes demonstrate gain, while in the orange graph (non-quarantine), every node demonstrates high gain. Infections are therefore easier to identify in the blue scenario than the orange one.

This connects nicely to my STS research, which sought to use a framework known as data relativity to investigate how the data produced by computer models should be translated into real-world action. Simplifying assumptions and errors in underlying datasets can skew the results of models and potentially make them inaccurate. My research first attempted to resolve this through tests and techniques to verify the accuracy of a computer model, but found that these

often fall short of providing confidence in model accuracy. The truth is that inaccurate models are frequently employed when determining real-world action, and that, strangely, this rarely leads to disaster. I found the answer to this conundrum by applying data relativity, a framework that mandates that the techniques used to create data matter just as much as the data itself, and discovered that inaccurate models can still be useful if they are used in ways that comply with this framework. Specifically, I found that evaluating models is often difficult because many techniques require sourcing a lot of data, and if that data is not compiled through means congruent with data relativity, then it will not be a successful evaluation. Similarly, successful presentation of model results typically follow data relativity by placing findings relative to the parameters and assumptions that generated them, while unsuccessful ones do not.

As is plain to see, these two topics are not separate entities. The process of creating and using a computer model of my own design gave me experience in the field of computer modeling, and that experience shaped my perspective while conducting STS research. The information I obtained from that research then informed how I would go on to present the results of my simulation. Reflecting on this project as a whole, I am surprised to see just how much of my undergraduate experience was shaped by this most recent global pandemic. I became interested in computer modeling during that time for exactly that reason. I became interested in both the topics presented in this project because I believed that finding the answers to them would improve our ability to respond when such an event inevitably occurs once again. It is my hope that, in some small part, the insights gained from my work may help in exactly that.