

Introduction to Computational Social Science assignment 7:

Mechanistic models in disease spreading prediction on social networks- A response paper

Mihaly Hanics, MS in Social Data Science

This paper is about describing how, why and to what extent are mechanistic models useful in predicting disease spreading over social networks. The paper we write a response about the paper: “Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: lessons and the way forward” by Chowell G. et al [1]. After describing the topics stated in the paper, I also mention how it can be used in a different research scheme.

The most key distinct information in building prediction models according the paper is working with epidemiological datasets (having trajectory data, to predict current epidemic trajectory), knowledge on the disease spreading’s most relevant modes of transmission (such as close contact, droplet or airborne) and the settings for most common transmission (e.g. hospital, public transport, school) and mobility patterns, and the natural history of the disease with as broad information on infections as possible. During the 2014-15 West African Ebola crisis, since the Ebola virus has been reported first decades ago; compared to the Covid-19 which had not much historical data due to being a new virus (there were information on SARS-CoV-1 and MERS, but this strain was much more infective and resulted in a huge outbreak, compared to SARS and MERS, which had less than 1000 cases combined). However, due to the disease spreading in third world countries with poor infrastructure, the information on transmission (contact networks) and mobility was limited. This initially caused that whilst prediction models could be built from the very first moment simply on historical data, for long time they could not be “actualized” as there was little (digital) data on where and whom are infected real-time. Thus, if the initial models were not accurate enough, it would take long time to have models with correct parameter-settings as data was gathered much slower than the disease spread.

In the paper, the authors describe how different models were built and can be built to predict the spread and outcomes of the disease, highlighting how multiple types of models differed in results (such as early-stage exponential models highly overshoot the size of the outbreak). They highlighted that whilst quasi real-time models are key to planning (reactive behaviour), it is also very important to have good initial predictions of the importance of mechanism that drive the growth of the epidemic (some included features are host and pathogen, transmission routes, immunity, behavior, interventions), to make correct decisions on dealing with the epidemic. They noted that whilst there were initial models (built on WHO publicly available data), the lack of enough observed real-time epidemiological data, it was challenging to reproduce the exact growth models. They highlighted the difference in results by using different prediction models. For example, after the first 5-6 generations of disease transmission, the growth showed an exponential-like curve and many models used exponential prediction, which turned out to be incorrect. So they noted that it is also worthy to implement models other than SIR-type compartmental models (the ones that always lead to predicted exponential growth in the absence of susceptible depletion, interventions or behavior changes) such as the generalized-growth model, which leads to (high-order) polynomial (sub-exponential) growth dynamics, but can also be used for exponential

modelling. On a diagram they showed that whilst these predictions vary much (even the non-exponential ones), the median and the interquartile region capture the real growth quite well. Therefore, they emphasize the need for 4 things for such predictions: models that allow deviations from theory (often deviating from exponential behavior, such as models implementing a seasonal component), even if data is lacking, the portfolio of models should be wide and span through the gamut of complexity, from abstracted phenomenological models to compartmental models and highly complex models, agent-based models. They suggest that ensemble model predictions are broader and more accurate. For widely available modeling, they suggest the creation of interactive platforms, websites, where multiple research groups would be involved, sharing their results and approaches.

From my point of view, the last point is very important, that model results shall be aggregated (with a “vote system”, or a median/average system). I found that throughout my works, when you have to provide a prediction or a solution to a problem that can never be 100% solved, the most “stable” (low variance, high accuracy) solution tends to be a combination of many “base methods”. Like in machine learning, ensembling many base estimators. When I designed machine learning models to predict height of basketball players just based on their performance, the best model I created was a custom-made ensembling forest, that was specifically designed “biased” to maximize 1-inch accuracy (resulting in less good perfect 0-inch accuracy), and turned out to be a great, very stable model, that had a very thin error-range, and guessed within 1-inch of the correct height half of the times, which is better than expected. And this model beat gradient descent models, those that minimize a loss function and are supposed to be most correct. Somehow ensembling is great to correct for each predictor’s wrong guesses, but lead to the more likely direction of the change of the epidemic growth function than the “most common guess”. This may remind the reader of the phenomenon called “Wisdom of the crowd”, when the average guess of participants is closer to reality than the best guess.

However, I found that this paper makes points that are “clear” on the theoretical side, but are unclear in how to implement practice. Poor regions lack from much more important infrastructure than sensors and transmitters, technology capable of recording and sharing data. In a way, probably all hospitals will need to train a nurse/doctor to learn how to record, and publish data on the internet. This would be a task for not the people working in epidemic prediction. Also, one can ensemble many models in many ways, who decides which models “deserve to be ensembled” and how to aggregate them? In some sense, I think this is in the hands of the decision makers, they probably get tons of predictions thrown at them every week, and if I had to guess they either account for the “average” of them, or account for the worst case.

A part that I found contradicting is that they suggest having a very wide range of models, that be ensembled, and then in the end state that we need interactive platforms (containing the data, and I believe the self-built models) so that groups can interact with each other and share their findings. But if everybody works together, then we’d have one assumed “perfect” model, instead of a very wide range of totally independent models ready to be ensembled.

Technical details: Even after reading the paper and running through it again, I found it so “semmitmondo” (Hungarian word for “nothingsaying”) (says not much new and is very

repetitive), closest meaning is meaningless) that I asked Bing AI to look it up to say something about it:

Prompt: “Summarize this paper: Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: lessons and the way forward.”

Answer: *The paper “Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: lessons and the way forward” provides a comprehensive analysis of the forecasting approaches used during the 2014-2015 Ebola epidemic in West Africa. Here are the key points:*

The paper discusses the unprecedented impact and modeling efforts associated with the Ebola epidemic, which provides a unique opportunity to document the performances and caveats of forecasting approaches used in near-real time for generating evidence and guiding policy.

The authors highlight the challenges and lessons drawn from these efforts, focusing on data availability and accuracy of early forecasts, the ability of different models to capture the profile of early growth dynamics in local outbreaks, and the importance of reactive behavior changes and case clustering.

They also discuss the challenges in forecasting the long-term epidemic impact very early in the outbreak and suggest ways to move forward.

The paper concludes that rapid availability of aggregated population-level data and detailed information on a subset of transmission chains is crucial to characterize transmission patterns.

The authors believe that coordinated forecasting efforts, combined with rapid dissemination of disease predictions and underlying epidemiological data in shared online platforms, will be critical in optimizing the response to current and future infectious disease emergencies.

This paper provides valuable insights into the role of mathematical modeling in managing public health crises like the Ebola epidemic.

The answer had useful parts, I forgot initially to talk about the conclusion. What else can I say, I agree with the answer? (With the exception of “valuable” insights into the role of mathematical modeling, those insights were that exponential is not always the best pick for modeling, and that ensembling is useful, I’m not in the field of real-time prediction but that didn’t say anything new to me. Although I remember seeing early exponential models of Covid predictions being the most popular and even then I already felt they will overestimate, which turned out to be correct. So quite some mathematicians still haven’t picked up this information that other models than exponential shall be given as much weight as exponential too. Scaring people (with lies) into acting safer is not right.) Other uses of LLMs I didn’t have for this assignment.

Therefore, GPT-4 (in Bing AI) helped me remember to write about the conclusion part of the paper (where they mention their recommendations moving forward), and helped me put my thoughts more together. (I used Bing AI, because it has GPT-4 implemented, with free usage.)