Modeling country to country disease spread in our modern world: Through COVID-19 spread and travel data

Adam Salyers - The University of Colorado at Boulder

The country-to-country spread of disease is a topic growing in importance every year across the world[1]. With the rise of COVID-19 and the highly criticized handling of the disease response worldwide, it becomes increasingly important to understand exactly how diseases with a variety of spread, infection, and death rates behave. COVID-19 was relatively tame in its ability to kill[2] and with the rise of antibiotic resistance in bacteria[3], as well as more dangerous diseases the response to COVID-19 shows how unprepared we might be. To address this I have created a network-based simulation incorporating modern disease response, and modern human movement with scalable disease factors to simulate a world response to a more deadly disease, specifically how the disease would spread between countries. I argue that this an effective and accurate way to diagnose the potential weak points of country-to-country disease spread.

I. Introduction

Disease is perhaps one of the biggest risks to human mass extinction that exists, pandemics can kill large swaths of the population, in its time in 1918 the Spanish Flu killed between 10%-20% of the infected population topping out at up to 50 million dead[4]. In contrast, the recent COVID-19 epidemic had a mortality rate of just above 2%[2]. The risk of a highly infectious, deadly virus or bacterial infection seriously crippling the world is becoming more likely. However, the data available for pandemic infections come as often as the pandemics themselves. With major pandemics occurring often 40+ years away from each other[4] the spread, vaccine creation, and treatment statistics change quite a lot. It is important to take advantage of the most recent pandemic to understand exactly how diseases spread in our modern world before the data becomes outdated. In just over 50 years the population density of people living in Asia, Africa, and the Americas has exploded[5]. More and more people are living in densely populated cities, prime for massive disease spread. In addition, the number of people traveling between countries worldwide has doubled in just over 20 years[6]. The reality is that our world is continually becoming more and more interconnected, with people consistently circulating the globe and coming in contact with a massive number of people and diseases. During the COVID-19 pandemic, we saw the main way to combat this was from government regulations that lowered the overall spread rate, between and within different countries[7]. Regulations included travel restrictions and bans, mask mandates, lockdowns, movement restrictions, economic restrictions, testing requirements, and domestic travel restrictions. All of these had varying effects on the spread of COVID-19 but provide us with a quality baseline to understand how COVID spreads within and between countries when different responses are implemented. In addition, no two countries or regulations are created equal, an autocracy like China, or Russia can more easily implement a lockdown and get populations to adhere compared to a democracy like Italy. Countries like Afghanistan have far more internal movement due to war and an inability to adequately enforce disease prevention mandates. Simulations that predict the spread of disease often do so on a small scale[8] which can vary widely from place to place. Focusing instead on the global scale of country-to-country spread allows us to understand the global spread of a pandemic, and “plugin” different in-country spread rates and regulations as our understanding of these individual pieces increases with time. In the rest of this article, I will show how modern data can be used to build a network of country-to-country spread rates that can inform us about how diseases spread between nations.

II. Methods

The most important piece of country-to-country network construction was the understanding of government responses. I used a dataset that covered a wide swath of different countries, regulations, and infection/death rates during the times of those regulations[7]. However, it was necessary to clean the data quite a bit. Although the dataset itself was very comprehensive there was a lack of understanding in the categories of the individual regulations. Categories within the school, sport, mass sections and more all had a variety of regulations that changed often and were far more difficult to interpret. While regulations like travel, mask mandates, and testing requirements had a much larger impact on infection rates and made it far easier to interpret. It was necessary to trim down the regulations to a more manageable size, the final regulations chosen were: Travel, Travel Partial, Travel Domestic, Travel Domestic Partial, Masks, Testing, and Testing Narrow. In addition, the dataset had a variety of missing data for certain countries regarding these regulations, and those with missing data were trimmed leaving 175 countries, the final size of our infection network. After that dataset had been properly trimmed the overall infection rates were calculated. I would attempt to create an infection baseline for any country with no regulations for at least 10 days, for countries that did not have the infection data for a baseline or had a very skewed baseline infection I used a spread rate of 25%[9]. I also used this data to calculate the spread rates under each regulation listed above. Using the first instance of a regulation in place I used countries’ current infections, and total deaths to create a “threshold” under which they would implement that regulation. These newly created datasets of threshold and government responses would be the main driver of change in infection rates within each country as the simulation progressed.

In order to model spread between countries, I used a multitude of datasets. Two being the outgoing tourism numbers of each country[10], as well as incoming tourism numbers of each country[6]. It should be noted that I was only able to get accurate outgoing tourism numbers for 81 countries, as there is simply not enough data for chaotic countries or autocratic countries that refuse to release their data publicly. However, since these 81 countries make up most of the tourism/business travel populations it did not have that significant of an impact on the simulation. Taking these two datasets along with top tourism "partners" I created a new dataset in which every one of the original 81 countries had a list of tourist numbers to every other country. Some countries could be "infected" by tourists from other countries but could not "infect" other countries with their own tourists. Using data from regions[11] and borders[12] I was also able to get a picture of the direct "spheres" of disease influence that each country had. Countries in the same regions could infect one another and countries that shared borders always had a chance to infect one another. In addition, I took world and country population data[13] in order to cap the infection number in individual countries, as well as determine the odds of infection to other countries whether through tourism or region/border spread. This simulated the reality that island countries, with travel restrictions, and good government mandates would be the last countries infected.

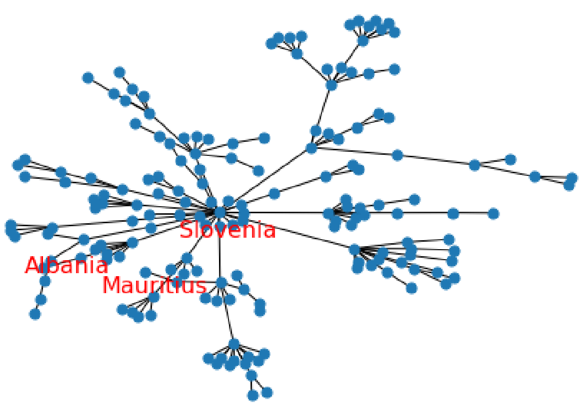
The simulation itself was set up as a turn-based simulation in which each day was a single turn and each country took their turn in order.

- Each country's infection was tracked for total infections, current infections, deaths, recoveries, and the day each person was infected on up to 20. After 20 days of a person being infected, they would become recovered and be removed from the number of people currently infected. Every day each infected person would have certain odds to infect another non-infected or recovered person. These odds were calculated by taking the current infection of the country under its current government regulations (described in paragraph one of methods) multiplied by a number set when the simulation was called.
- Each country would then have certain odds to infect another country based on tourism. These odds were calculated by the following equation, $p = (OT/CP)*(NI/CP)*x$ were OT is outgoing travelers, CP is country population, NI is the number of infected (inside the current country), and x is a value given when the simulation is called. If successful the simulation would choose a random country A from among those our infecting country B travels to with odds determined by, $p = (CT/TT)*Y$, where CT is the travelers to country A from country B, TT is the total travelers from country B, and Y is a value set when the simulation is called if country A is in the same region as country B.
- Each country would then have certain odds to infect each of its regional neighbors and border countries, determined by the functions, $p = (NI/CP)*Z$ and $p = (NI/CP)*T$ respectively. Where CP is the country population, NI is the number of infected (inside the current country), and both Z and T are values set when the simulation is called.
- All odds of country-to-country spread would also be affected by travel regulations. Decreasing if travel is partially banned, and decreasing more if it is fully banned.
- Finally, the simulation would check if the current infected and dead in any given country activated any new

regulations and if so would update the current infection odds of that country

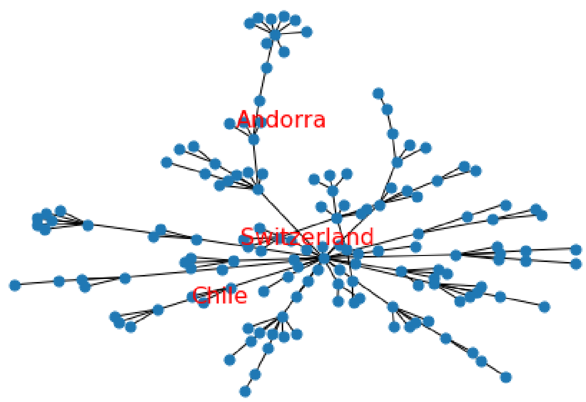
III. Results

The simulation went well overall, it became necessary to implement a baseline for infected regulations as many countries, specifically, autocracies had skewed numbers that may have been contributed by false reporting[14]. The main takeaway from the results was the “superspreader” countries were very much a reality in our current day and age. In the network shown we see Slovenia infecting 20+ countries, more than double what we see in any other country in the network. The network shown



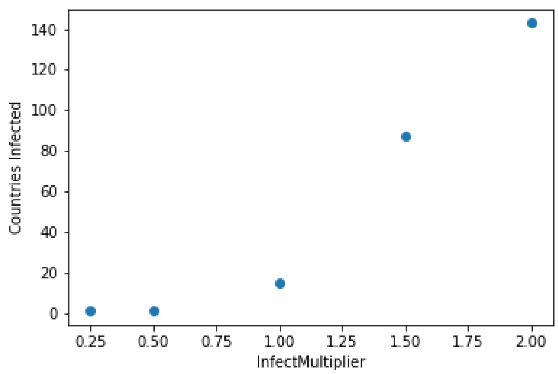
is a simulation run on COVID-19 data for spread rates, tourism, and death rates. We see a mirroring in a reality where Italy, a superspreader, infected over 40 countries, also much more than double what any other country infected[15]. The reality is, these countries with slow enacting regulations, as well as high levels of tourism, and lots of regional/border neighbors act as superspreaders. The disease infects a large number of the country's populace before the government regulation can slow down the infection rate. But during that time the high level of human movement allows the disease to move freely between a large number of different neighbors. After this happens 1-3 times a large majority of the world’s countries are infected and cannot rely on keeping the disease out as a form of health control.

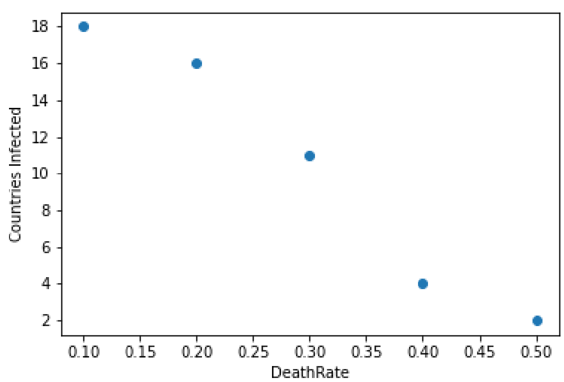
In addition, we also see that Land Borders play a huge role in the spread of one country to another. In the graphic below we see a simulation across a year of a COVID-19 similar disease that started in



Brazil. We can see in the graphic that Switzerland is still the main superspreader. The highest risk superspreaders are European countries because they have large amounts of land borders, regional neighbors, and tourists. In fact on running the simulation 25 times over the course of a year with random start countries outside Europe, a European country ended in the top 3 of country spreaders in over 50% of cases. This also mirrors what we see in reality where COVID-19 starting in China ended up with Italy as the main source of infections for other nations.

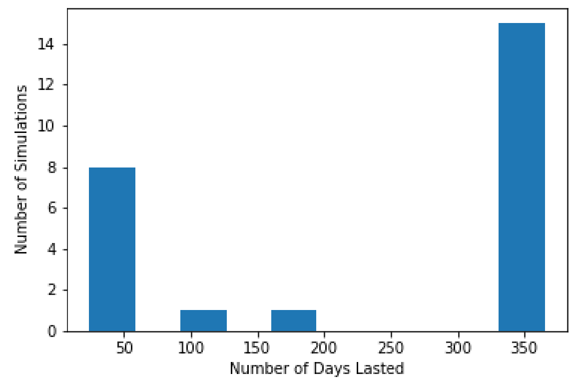
The simulation also revealed the contrast between infectivity and death rates for maintaining the life of disease.
The figures below show the number of countries infected





after 200 days of disease starting in Mexico. With a death rate of 2%, we can see as infectivity rises, the number of countries infected rises dramatically. Compare this to death rates whereas the death rate increases the number of countries infected after 200 days decreases. Showing that a virus that kills has to have a large infectivity rate in order to sustain itself.

The simulation also showed how often diseases themselves die out when the infectivity to death rate are not well balanced. The chart below shows the lengths of the simulation when the death rate is at 0.3 and the infect multiplier is at 0.5, running for one year. We can see that often the disease would die out, most of the time in less than a month. For a lot of the simulations the disease did carry on, but it shows how important it is to understand the infectivity and death rates of diseases when



understanding how they spread.

IV. Discussion

When looking at the results from the simulation the discovery of superspreader countries is most relevant to what we can do to stop future outbreaks. The fact that most countries get infected from just a handful of centrally located, slow regulating, and regionally centered nations tells us that regulations outside of those countries must come swiftly. In an ideal world, these “superspreaders” would put fast-moving regulations on themselves, clamping down on their citizens' business and leisure travel to other countries and states. This may be possible in autocratic countries where political control is focused on one central authority. But for countries that are most often super-spreaders, chaotic or democratic countries, it is unrealistic. Instead, it is up to the world to view tackling a pandemic as a global effort. Regulations for countries at risk, that is countries with lots of tourism or with infected neighbors must begin putting regulations on themselves before they get infected. It may be impossible to get superspreaders under control once they are infected, but if those super spreading countries can start implementing regulations before they get infected it is possible to slow the spread of superspreaders.

Seeing how these superspreaders share many of the same characteristics, centrally located, regionally connected, and with higher tourism populations it may be an option to identify countries at risk of being superspreaders. Countries like Italy may end up infecting 40+ countries in a truly deadly pandemic like another Spanish Flu. If it is possible to identify these countries before a pandemic even starts countries can identify which nations pose the highest risk and close off travel to those nations as soon as the pandemic starts anywhere. Countries that are at a high risk of being superspreaders can also be prepared for a pandemic. Those that are determined high risk, can be encouraged by other nations or the UN to implement effective pandemic response plans. These are potential solutions but all hinge on the world understanding the threat diseases pose and the country's willingness to act.

Seeing diseases die out “quickly” in the simulation is mirrored in real life. Diseases like Zika, Swine Flu, and more were all potential pandemic level diseases but never got there. Yet COVID-19 infected many countries before we understood its spread and death rates. It is very important to understand the death rates and spread rates of these diseases. As shown in the simulation these factors can determine whether a disease itself will become a pandemic. In the simulations, the slowest spreading time of any disease was in the beginning. We in the world need to use the time of slow-spreading disease to fully understand it. So long as countries withhold or fabricate information regarding the spread of diseases that originate in their nations[14] we as a world will be at a disadvantage. It is vital that we use the time when a disease first appears to determine its threat level.

In an ideal world, we would have fast-moving government regulations, perfect healthcare, and human life would be valued over economic prosperity, but that is not the case. As diseases of varying infection and danger rates appear it may save lives to shut down the entire world but that is unrealistic for world economies. A valuable, although callous, next step would be to create a model that determined the economic impact of government regulation as compared to the threat level of any disease. We cannot all go into lockdown as soon as a new virus appears even if it would save lives. But if we can determine the difference between the next Zika and the next Spanish Flu before it is too late we can save lives and potential catastrophe while not bankrupting the world economy.

In terms of the next steps for this model, better data needs to be collected. Autocracies the world over put forward skewed data that makes it very difficult to monitor disease spread within their nations. Guarded tourism and migration numbers force estimates within the model that lead to spreading patterns that do not mirror reality. Change in this respect looks unlikely barring a change of heart in those governmental regimes. An alternative would be simulations for different countries spread on a smaller scale. There are quite a few simulations available[8] but they often focus on specific circumstances and are overdone for wealthy countries. Simulations and understanding of spread for poorer, and more autocratic countries are necessary to better understand the world spread of disease.

V. References

[1] Bloom, David E, and Daniel Cadarette. “Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response.” *Frontiers in immunology* vol. 10 549. 28 Mar. 2019, doi:10.3389/fimmu.2019.00549

[2] ScienceAlert. "Fatality Rate of Major Virus Outbreaks Worldwide in The Last 50 Years as of 2020." Statista, Statista Inc., 31 Jan 2020, https://www-statista-com.colorado.idm.oclc.org/statistics/1095129/worldwide-fatality-rate-of-major-virus-outbreaks-in-the-last-50-years/

[3] Ventola, C Lee. “The antibiotic resistance crisis: part 1: causes and threats.” *P & T : a peer-reviewed journal for formulary management* vol. 40,4 (2015): 277-83.

[4] *Outbreak: 10 of the Worst Pandemics in History by Staff*. www.mphonline.org/worst-pandemics-in-history/.

[5] Max Roser, Hannah Ritchie and Esteban Ortiz-Ospina (2013) - "World Population Growth". *Published online at OurWorldInData.org.* Retrieved from: 'https://ourworldindata.org/world-population-growth' [Online Resource]

[6] “International Tourism, Number of Arrivals.” *World Bank*, 'https://data.worldbank.org/indicator/ST.INT.ARVL.

[7] Porcher, Simon "A novel dataset of governments' responses to COVID-19 all around the world", Chaire EPPP discussion paper 2020-03, 2020.

[8] Guo, Z., Xu, S., Tong, L. *et al.* An artificially simulated outbreak of a respiratory infectious disease. *BMC Public Health* **20,** 135 (2020). https://doi.org/10.1186/s12889-020-8243-6

[9] Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina and Joe Hasell (2020) - "Coronavirus Pandemic (COVID-19)". *Published online at OurWorldInData.org.* Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

[10] “International Tourism, Number of Arrivals.” *World Bank*, 'https://data.worldbank.org/indicator/ST.INT.ARVL.

[11] *World Region Data*, UN Stats, 'https://unstats.un.org/unsd/demographic-social/products/worldswomen/annex_tables/List%20of%20countries%20-%20Table%209.xls.

[12] GeoDataSource. “Country Borders Database.” *GeoDataSource*, www.geodatasource.com/addon/country-borders.

[13] “Countries in the World by Population (2021).” *Worldometer*, www.worldometers.info/world-population/population-by-country/.

[14] Thomas, Dr. Liji. “Cremation Numbers Reveal Possible Suppression of True COVID-19 Data in China.” *News*, 9 June 2020, www.news-medical.net/news/20200608/Cre

mation-numbers-reveal-possible-suppression-of-true-COVID-19-data-in-China.aspx.

[15] Hasmot Ali, Md. Fahad Hossain, Md. Mehedi Hasan, Sheikh Abujar, Covid-19 Dataset: Worldwide spread log including countries first case and first death, Data in Brief, Volume 32, 2020, 106173, SSN 2352-3409, https://doi.org/10.1016/j.dib.2020.106173.

More Code and Pickle Files can be downloaded from the GitHub:

https://github.com/AdamSalyers/Final-Paper-CSCI3352