The Spread of Information: Comparing Government and Individual Action's Effect on the Spread of Infection

Justin Deterding
Dept. of Physics and Astronomy
University of New Mexico
jdeterding@unm.edu

Catherine Wright
Dept. of Computer Science
University of New Mexico
wrightc@unm.edu

***Abstract*—COVID-19 has had a profound impact on lives worldwide, and researchers from multiple disciplines have been coming together to attempt the daunting feat of predicting what actions are best for slowing the spread of infection. In this paper we present the results of an epidemiological model designed to compare the effects of different types of decision making, modeling isolation by the individual or by their community. Our results show that countries acting fast to close borders works especially well when the individuals within the country also try to isolate themselves from others, even if some individuals go against the guidelines occasionally. We also show by use of a toy model that when countries act to close borders early, the effect that the collective population has on a country reduces.**

I. Introduction

Contagions have shaped the human population for as long as a human population has existed, and we are at a much greater risk today to disease exposure than centuries ago due to the highly connected structure of our civilization. The global population continuously increases, with 80% of people living in dense urban environments [1] and hundreds of thousands of people traveling internationally every day [2]. This deeply connected network of individuals is advantageous for information dissemination and economic growth, however, it is highly detrimental for containing infectious diseases.

We can use network theory to capture some of the traits exhibited by the global population network. For example, analysis of Facebook's social network has shown that it is well-connected, although sparse. Well-connected since over 99.9% of nodes are part of a large central component, however sparse because none of these nodes are connected to even 1% of the total nodes. Even without any high-degree nodes in the network (also called *hubs*), there are still extremely short paths that connect pairs of users, with an average distance of 4.7 edges [2]. This means that Facebook's social network exhibits the characteristics of a "small-world" network. Small-world networks have a property that if a few long-distance connections are introduced into the network, the average path length between every pair of nodes reduces dramatically [3]. It is probable that the global population network behaves similarly like a small-world network, an example being when someone travels internationally they create an edge that drastically reduces the diameter of the total network. Unfortunately, if this traveler carries with them an infectious disease it can then spread in a new area. This has happened countless times throughout history, such as with the introduction of smallpox to the Aztecs and Incas, Hawaii, and the Americas [2]. Interestingly enough, hubs may accelerate the spreading but are not necessary for an infection to disperse. The driving force of infection over a population is the overall average degree of a node [4], meaning that social distancing is vital for slowing the spread.

Since the COVID-19 pandemic started in December epidemiologists, politicians, and economists have been scrambling to understand how best to preserve the standard of human life we are accustomed to. We are told to wear masks when in public, limit social interaction, wash our hands multiple times per day, and self isolate when possible. These are actions that we can take as individuals to slow the spread of infection, and we can think of these efforts as a *bottom-up* approach where the actions of the lower-level agents control the outcome of the global collective. We have also seen world leaders take action by closing international borders, banning large social gatherings, and forcing nonessential businesses to close. Such measures can be interpreted as *top-down* control, with the collective making decisions about how connected individuals are in the global network. Anthony Eagan from the Santa Fe Institute describes a similar relationship of power between local and federal government [5], arguing that top-down control has the obvious benefit of a faster, more consistent reaction to the changing situation, while bottom-up control leverages the fact that every individual and community know their unique circumstances best.

Modeling a disease infecting a closed population of individuals with like characteristics is typically handled by some version of a SIR model, classifying each individual in the population as (S)usceptible, (I)nfected, or (R)ecovered with respect to a transmittable disease. What we are experiencing with COVID-19 is a dynamic that spans different cultures, governments, and population densities. It is therefore necessary to make simplifying assumptions that do not cheapen our model or its results. Our model creates a network of connected

individuals with clusters representing various countries. We modified an existing codebase that simulates infection through a population to incorporate both bottom-up and top-down efforts to slow the spread. Bottom-up control was modeled by giving individual nodes the power to break connections, imitating self-isolation and social distancing without omniscience regarding which individuals truly have the disease. Top-down control was modeled by giving clusters the ability to break connections with other clusters based on a tolerance to the number of infected individuals in each cluster. We compared the outcomes of using both effects and found that unsurprisingly a combination of individual and country isolation results in the fewest individuals being infected. We then modified our model slightly, using an SIS model to measure the transfer entropy from the global population to each country and vice versa for different thresholds of top-down control. Our results show that countries that are sensitive to infection rates and monitor international travel closely have fewer infected individuals over time, and experience less residual effect from the state of the global population.

Section II explains our models in-depth, including necessary assumptions and results from using different methods of disease containment. It also contains an analysis of the information flow between global and local populations. Section III discusses results from each facet of our model and concludes our findings.

II. Methods and Results

A. Modeling top-down and bottom-up infection spread

Modeling infection spread from real-world data is challenging. Our intent, in the following sections, is to compare two directions of control: the individual control, and the collective control, with the goal of understanding how these two scales of intervention influence the spread of an infection. For modeling the spread of an infection and studying the top-down and bottom-up approaches to decreasing the spread of the infection we utilized Epidemics on Networks (EoN) [6]. EoN is a python package implementation of the algorithms and models from Mathematics of Epidemics on Networks. For our work we began by forking the existing EoN repository on Github, and used the discrete SIR model to implement our top-down and bottom-up interventions [7]. The discrete SIR model determines if a susceptible node in a network is infected by an infected neighbor in the network with some probability P equivalent to the r_0 parameter in the continuous SIR model. A node then remains infected for one round before transitioning to a recovered node. The simulation continues until there are no infected members in the population. In our work we introduce two new actors into the discreet SIR model: a graph commander, for top-down rule enforcement, and a node commander for bottom-up rule enforcement. In the discrete SIR algorithm the first action is made by the virus, and the newly infected nodes are stored. This information is NOT made available to the node or graph commanders during the round in which they were infected. Next, individual nodes get to make their decisions followed by the graph commander. The

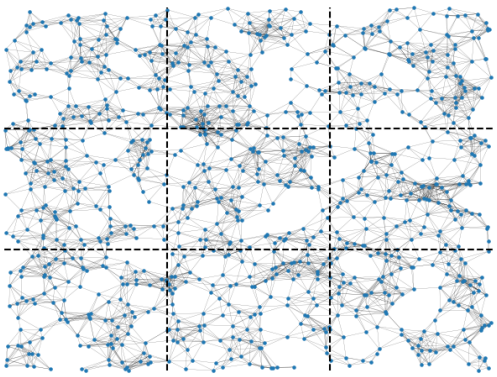

Fig. 1: The SIRD dynamics of the least optimal, optimal, and intermediate rule set.

node commanders only have access to the neighbors assigned to them at the start of the simulation. They can then add or remove edges based on the state of their neighbor nodes. The graph commander at each turn calculates the percent infected for each subgraph (country) including its subgraph. If a particular subgraph has an infection rate higher than a specified external threshold, then the graph commander will remove all edges to that subgraph (borders closure) with the high infection rate. Likewise if a graph commander determines that its internal infection rate is greater than some internal threshold it will remove all external edges, isolating itself from the remainder of the subgraphs. For the following analysis all graphs were composed of 999 nodes with an average node connections between 11.5 and 12.0 connections per node. Node coordinates were generated using a uniform random generator between 0 and 1, and node connections were made by connecting nodes that were within a specified euclidean distance. The network in figure 1 is a network connected in this fashion with nine subgraphs of 111 nodes. Each of the following simulations used a single initially infected node, with a transmission probability of 80%. In our analysis we evaluate 10 executions of the discreet SIR model, then calculate the mean time series of the susceptible, infected, recovered populations and the average mean node degree (AMND) of the graph, which we refer to as the mean SIRD dynamics. From the mean SIRD dynamics we characterize a particular rule set using the peak of the mean infected curve (PMI), and the minimum of the AMND. Then we attempt to classify the optimal strategies by selecting the strategy that minimizes the PMI, while maximizing the AMND.

1) Modeling bottom-up infection spread: To understand the dynamics of a bottom-up approach to controlling the spread of infection we developed a rule matrix that the nodes utilized to make their decisions. In the table below is an example rule table. Each entry in the rule table indicates the probability that a node will change the state of an edge. For example, if a node has a susceptible neighbor (S) that it is currently disconnected from, then it will likely re-connect to that node with a 75% likelihood. You'll note that a majority

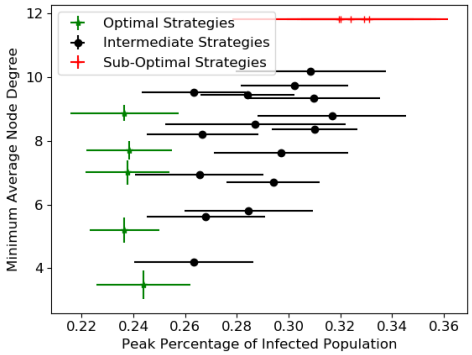

Fig. 2: Scatter plot of PMI vs. AMND for varying node strategies. The optimal point would be a point in the upper left most corner of the scatter plot. This point would minimizes the peak number of infected while maintaining the highest average node connection. Error bars are the 95% confidence interval developed from 10 independent trials.

of the indices are zero or one. The rules for connecting or disconnecting from the infected and recovered simply state to connect to these nodes. For recovered nodes this makes sense, there's no risk in connecting to a recovered node, then to maximize the connectivity of the graph these edges should be restored. Similarly, because of the order in which actors make their decision, it only matters that if you were disconnected from an infected node prior to finding out it was infected. Therefore we treat infected nodes like recovered nodes.

	Disconnected	Connected
S	0.75	0.50
I	1.00	0.00
R	1.00	0.00

In figure 2 we show a scatter plot of various strategies based on varying the probability of connecting and disconnecting from susceptible nodes. To determine the optimal strategy we use a scatter plot of the PMI vs. AMND. The plot represents the various strategies with their corresponding uncertainty, the 95% confidence interval of the 10 averaged runs. Ideally, we would want a point in the uppermost left corner, which corresponds to maintaining maximum graph connectivity, while simultaneously having the lowest number of infected individuals. However we note that that region is empty. We determined the optimal strategies by placing the highest priority on MPI, then AMND. One could place a higher priority on AMND, and lower priority on MPI and get a different set of optimal strategies. The optimal strategies (green) each lie within the 95.0% confidence interval of the others. In all five of the top strategies, shown in the table below, there is one salient feature. They all disconnect with 100.0% probability from their susceptible neighbors.

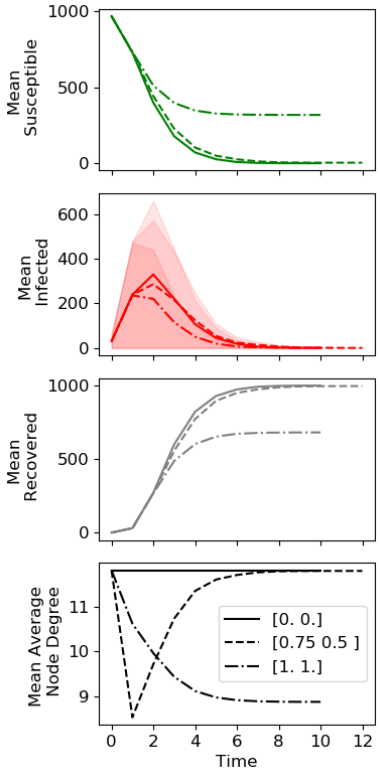

Fig. 3: SIRD dynamics of the least optimal, optimal and intermediate rule set. The legend show the probabilistic rule set for connecting and disconnecting to susceptible nodes explained in the preceding paragraph.

Rank	Disconnected	Connected
1	1.00	1.00
2	0.25	1.00
3	0.50	1.00
4	0.75	1.00
5	0.00	1.00

The top-ranking strategy essentially oscillates between connected and disconnected, and similarly the other rules oscillate between completely disconnecting and reconnecting some percentage of their neighbors. In figure 3 the dynamics of three strategies are compared. We note that the optimal performing should rebound to maximum connectivity, like the other strategies. However since the strategy stops the infection before it reaches the entire population there is a significant susceptible population. Since our nodes don't take inventory of their own state many infected and recovered individuals will continue to disconnect from susceptible neighbors. While this is a deficiency in our model in AMND should remain relatively unaffected by this.

2) Modeling top-down infection spread: In studying the spread of infection using a top-down approach we used graph similar to the one in figure 1 with nine countries. There are two parameters to set for each of the graph commanders, the internal and external tolerances. In varying these parameters

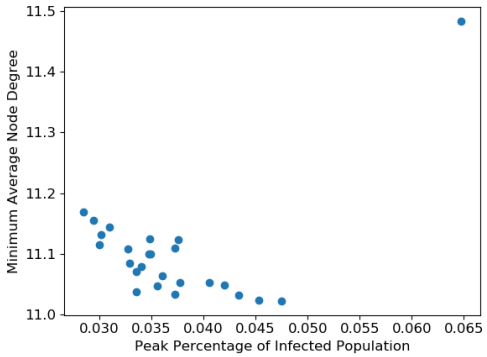

Fig. 4: Scatter plot of the results of various various internal and external threshold combinations of a graph commander. The upper right most point corresponds to no graph commander. The lower right points correspond to varying combination of border closing thresholds. The optimal, upper left most point, corresponds to the lowest possible threshold combination (0.01, 0.01).

we assumed that all countries had the same internal and external tolerances. While this is not necessarily true, it was necessary for this work to narrow the scope of the study such that the study could be completed in a reasonable amount of time. In figure 4 we varied the internal and external tolerance from $1.0\% - 8.0\%$ in addition to including a 100% tolerance which deactivates the graph commanders. In the scatter plot below the upper right point is when none of the graph commanders respond (1.0,1.0) to the spread of the infection, and is our point of comparison for an uncontrolled infection spread. The lower left is varying combinations of external and internal tolerances. The most effective strategy is to have a very low internal tolerance to the infection. This maximizes the PMI and while maximizing the AMND. This is because if a country can manage to maintain the infection within its borders, then other countries don't get infected and don't have to disconnect.

In the scatter plot above we did not include error bars, because they were larger than the spread of the plotted points. Since only one node is initially infected it depends where that node is. Whether or not it's close to a border or if the initial node is more or less connected. While we could increase the number of initial infected this defeats the possible effects of early containment. In short our results show that early containment can be beneficial to all countries, however the results are highly dependent on where the initial infection begins. In the SIRD dynamics plots below we see that the least effective strategy is to do nothing, and the best effective strategy is to have a low internal and external tolerance and to act quickly to contain the spread of the virus.

3) Modeling bottom-up and top-down infection spread: In modeling the effects of the combined top-down and bottom-up interventions we utilized a slightly different rule than the one in II-A1. Instead we implemented a quarantine rule,

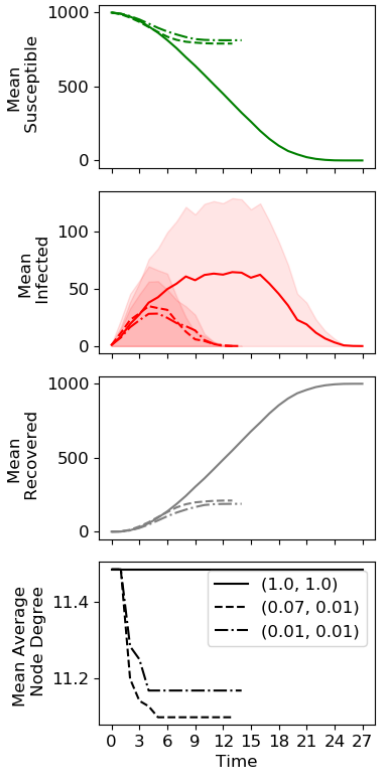

Fig. 5: The SIRD dynamics of the least optimal, optimal and an intermediate rule set. The legend corresponds to (Internal Threshold value, External Threshold value). The comparison of the do nothing rule set (1.0,1.0) compared to an early response (0.01,0.01) underscores the effect of an early response.

that is only activated when the graph commander closes its borders. Furthermore, to simplify the search space, so that the simulation completes in a reasonable amount of time the internal and external tolerance of the graph commander was set to the same value. For the node quarantine rule there is a specified quarantine level (QL). The allowed percentage number of connections a node to possible connections is then set to $1 - QL$. However, nodes are also allowed to cheat the quarantine at some rate (CR). When a node cheats it randomly samples form it neighbor nodes and reconnects. At the next time step, unless it cheats again, it will return to its quarantine specified number of connections. In figure 6 we simulated various border closing thresholds as in II-A2, while varying the QL and CR parameters. Again we have a scatter plot of the PMI vs. the AMND however this time while selecting the optimal strategies by sorting with priority given to strategies that minimize peak infection while maximizing the minimum average node degree, we also sorted by strategies with the lowest uncertainty in their outcome. In the table below the top five strategies are displayed. For reasons discussed previously the results of these simulations had very large variations. Of the strategies shown below the ratio of $I_{std}/I_{max} < 18\%$, Whereas for many of the other optimal algorithms without

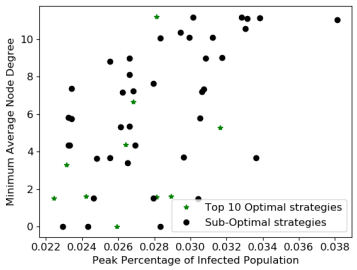

Fig. 6: A scatter plot of the MPI vs. AMND, the optimal point would be in the upper right corner. The optimal nodes are selected based on strategies with minimal uncertainty, in addition to minimizing MPI and maximizing AMND.

considering the uncertainty this measurement varied as much as 60% about the mean. From the table below it arguable that for consistent results a strong quarantine measure is necessary. However we note that there are reasonably high cheating rates. Similar to II-A1 it seems an oscillation between highly connected and sparsely connected graphs can work to effectively limit the spread of the infection.

Tolerance	CR	QL	I max	D min
0.03	0.25	0.95	0.024	1.61
0.05	0.50	0.75	0.032	5.29
0.01	0.00	0.75	0.022	1.51
0.08	0.25	0.95	0.029	1.61
0.01	0.75	0.95	0.027	6.67

From the SIRD dynamics we see that early border closings, and strong quarantine restrictions the infection can be effectively stopped. However for late closures and weak quarantine procedures the infection can exist in the population for a much longer period of time infecting a far larger portion of the population.

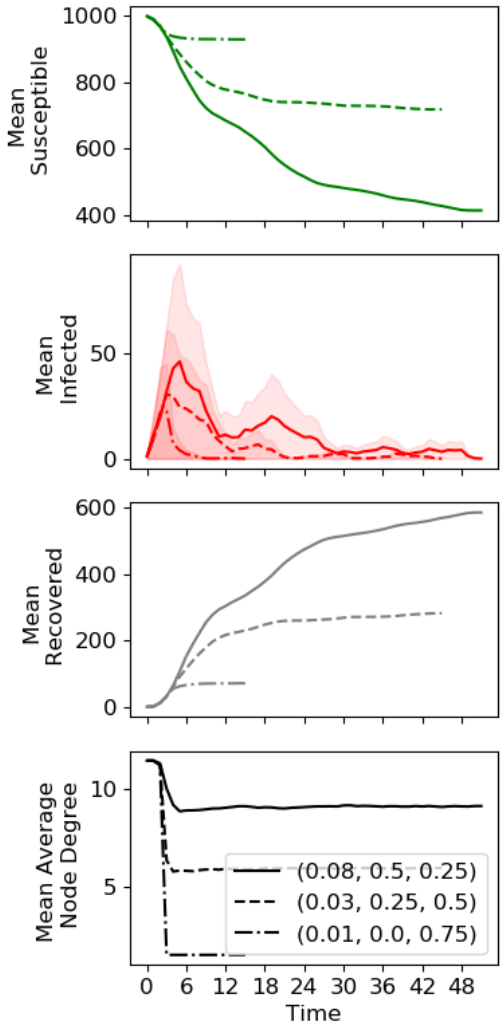

Fig. 7: The SIRD dynamics of the least optimal, optimal and intermediate rule set.

B. Information flow: the global effect on local populations

It's intuitive that closing borders early prevents the spread of infection across countries while closing borders too late does not. The difficulty is in finding at what point in a pandemic is the right time to close borders in order to maximize citizen safety and minimize travel restriction. To attempt to find an answer to this question we measured the flow of information between the global population and each country during the spread of an infectious disease. Our model mimics a toy model developed by Sara Walker et. al. for analyzing causal information transfer in evolutionary transitions [8]. Using a modified version of our model described in section II-A we produced a generalized global population consisting of 10 countries, each with an equal population and population density δ such that δ is the average degree of a node not including edges that connect to a different country. The subgraphs are connected by a global connectivity parameter ϵ, which is the average degree of each node over edges connecting two countries. For greater global connectivity ϵ, the global population is more inter-connected and the diameter of the network is much smaller, making infection travel faster and reach more susceptible individuals. While before we used a discrete SIR model, we needed much more infection data to measure information so we utilized an SIS model for which individuals who get infected can only return to being susceptible, making it possible to collect time series data for any length of time. This model uses the same parameters as before: the number of individuals initially infected (ρ), and the rate of infection (r_0).

To compare and contrast the effectiveness of top-down control we vary the tolerance threshold τ. To simplify our model we assume each country has the same tolerance to the disease in the country as they do to the disease in other countries. In order to view the dynamics of the disease over time we added to our model the possibility for countries to reopen their borders; if a country has less infected individuals than their tolerance allows they can open borders to any other country

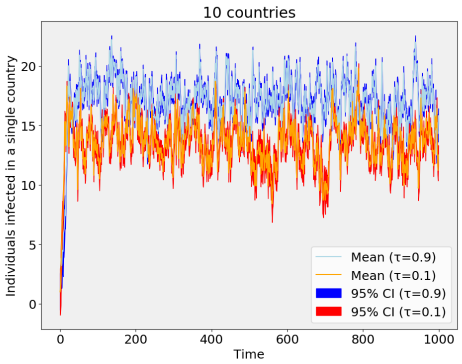

Fig. 8: Average infected individuals in countries with population of 100. Parameters: $\delta = 4$, $\epsilon = 0.3$, $\rho = 1\%$, $r_0 = 15\%$

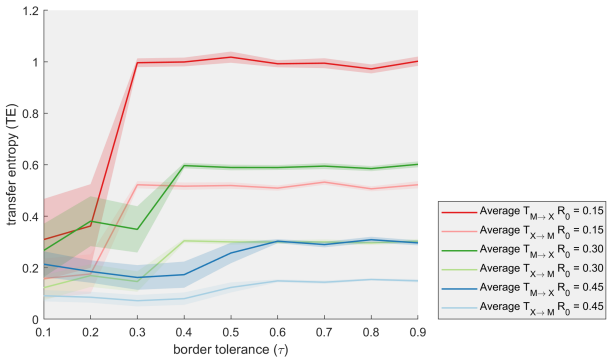

Fig. 9: Transfer entropy for varying tolerance τ with infection rates r_0 increasing from .15 (red/pink) to .30 (green/light green) to .45 (blue/light blue). Transparent bars indicate 95% confidence intervals. For all plots the global connectivity $\epsilon = 0.3$, with $\rho = 1\%$ of the population initially infected

that also has an infection rate below their tolerance threshold. To see the difference between eagerly closing borders and never closing borders, we ran this simulation over 1000 time steps and gathered the number of infected individuals for two tolerance thresholds: $\tau = 0.1$ and $\tau = 0.9$. The average individuals infected in every country is shown in figure 8, with high tolerance shown in blue and low tolerance shown in red with 95% confidence intervals. It is clear that if borders are eager to close then the infection is slowed, and countries will consistently have fewer citizens ill.

To calculate information flow we collected time series data of the number of infected individuals for each country. The average number of infected individuals of all countries was also collected as the global state M. For each parameter we performed 10 runs with a random sample of 3 countries chosen each run to represent local behavior. To measure the directional flow of information from the collective to an individual country ($T_{M\to X}$) and from an individual country to the collective ($T_{X\to M}$) we utilized transfer entropy, which measures the amount of information exchanged from one system to another [9]. The equation for the transfer entropy measured in bits from system Y to system X is

$$T^{(k)}_{Y\to X} = \sum_n p(x_{n+1}, x^{(k)}_n, y^{(k)}_n) log\left[\frac{p(x_{n+1}|x^{(k)}_n, y^{(k)}_n)}{p(x_{n+1}|x^{(k)}_n)}\right]$$

such that k is the dimension of embedding, where similar to the order of a Markov process, k signifies how many generations previous are used to calculate the information transferred for a given generation. In our model we used JIDT [10] to generate MatLab code for a discrete transfer entropy calculator. In [8], Walker uses many different values of k, choosing $T_{Y\to X} = Max\{T^{(k)}_{Y\to X}\}$. Due to issues in Java heap space using MatLab our simulation could not calculate transfer entropy for $k \geq 3$, so our results are found taking the maximum transfer entropy of $k = 1$ and $k = 2$. From this we attained 30 values for information transfer at each tolerance level (0.1 through 0.9), and simulated diseases with three varying rates of spread. Results are shown in figure 9, with $r_0 = 0.15$ shown in red (top-down) and pink (bottom-up), $r_0 = 0.30$ shown in green and light green, and $r_0 = 0.45$ shown in blue and light blue. 95% confidence intervals are shown as shaded bars around each average, and all other parameters were held constant.

The first observation is that top-down transfer entropy is always greater than bottom-up transfer entropy, meaning that the global population affects individual countries more than individual countries effect the global population. For every observed rate of transmission we see that transfer information eventually converges to a state at which top-down transfer entropy is double that of bottom-up transfer entropy, with little variance. The point at which this convergence happens seems related to the spread of infection, such that for diseases that spread slowly there is convergence at $\tau = 0.3$ while for diseases that spread quickly we don't see convergence until $\tau = 0.6$. This shows that for slow-spreading diseases the only way to reduce the top-down causal effect is to have highly sensitive borders. For faster spreading diseases however, tolerance can be higher while still reducing top-down causal effect.

III. Discussion and Conclusions

We've produced a model that simulates a disease spreading through a closed population with different strategies of control. By using only bottom-up control by the individual we saw that the optimal strategy is for every individual to disconnect from their susceptible neighbors. This follows general intuition of how to slow the spread of disease, and is something that most of us are trying to accomplish by social distancing and working from home. By using only top-down control we saw that while early containment may be beneficial, it depends where infection started and where it has reached. Often times when we realize a rapid increase of infection it is already too late. By combining both types of control we arrive to a situation similar to the one we are in today. Closing borders

and quarantining individuals does help to slow the spread, even if some individuals "cheat" in quarantining.

Analysis of the directional flow of information between the global collective and individual populations shows that there is a dominant causal effect from the global population. However, when countries are highly sensitive and close borders early this effect is reduced based on how spreadable the disease is. A difficulty here is that the rate of spread for COVID-19 is still undetermined, and seems to vary by country [11].

Moving forward it is hard to justify maintaining a strict form of control like the ones we have modeled. As humanity attempts to juggle the health of individuals with the needs of our economy we are worried that prioritizing one will be disastrous for the other. Based on our results we can only suggest that maintaining some social distancing is the most easily accomplished method of control that produces significant differences in the spread of infection.

REFERENCES

[1] G. West and C. P. Kempes, "Policies for responding to pandemics should be rooted in a scientific understanding of cities," *SFI Transmission: Complexity Science for COVID-19*, 2020. [Online]. Available: http://sfi-edu.s3.amazonaws.com/sfi-edu/production/uploads/ckeditor/2020/05/01/t-025-kempes-west.pdf

[2] M. O. Jackson, *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*. Pantheon, 2019.

[3] M. Mitchell, *Complexity: A guided tour*. Oxford University Press, 2009.

[4] B. J. Cowling, K. H. Chan, V. J. Fang, L. L. Lau, H. C. So, R. O. Fung, E. S. Ma, A. S. Kwong, C.-W. Chan, W. W. Tsui *et al.*, "Comparative epidemiology of pandemic and seasonal influenza a in households," *New England journal of medicine*, vol. 362, no. 23, pp. 2175–2184, 2010.

[5] A. Eagan, "It is important to keep in mind that as agents we maintain bottom-up control, even if we lack decisive power." *SFI Transmission: Complexity Science for COVID-19*, 2020. [Online]. Available: https://sfi-edu.s3.amazonaws.com/sfi-edu/production/uploads/ckeditor/2020/04/15/t-015-eagan.pdf

[6] J. C. Miller and T. Ting, "Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks," *arXiv preprint arXiv:2001.02436*, 2020.

[7] "Github repository with source code, forked from mathematics of epodemics on networks:," https://github.com/judeter/Mathematics-of-Epidemics-on-Networks.git.

[8] S. I. Walker, "Evolutionary transitions and top-down causation," in *Artificial Life Conference Proceedings 12*. MIT Press, 2012, pp. 283–290.

[9] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[10] J. T. Lizier, "Jidt: an information-theoretic toolkit for studying the dynamics of complex systems," *Frontiers in Robotics and AI*, vol. 1, p. 11, 2014.

[11] K. Linka, M. Peirlinck, and E. Kuhl, "The reproduction number of covid-19 and its correlation with public health interventions," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/05/06/2020.05.01.20088047

[12] "Github repository with source code:," https://github.com/judeter/ASII.git.

IV. CONTRIBUTIONS

Work done on this project was evenly distributed between the two group members. Justin produced the majority of the model, results, and write-up for Section II-A. Catherine provided the abstract, introduction, and discussion, along with the modified model, results, and write-up of Section II-B. All code used in this project was written by Justin and Catherine or cited appropriately in the paper and commented as such in our programs.

The source code for all results generated in this project can be found at [12] and [7].