

# Analyzing Epidemic Thresholds on Dynamic Network Structures

Keegan Kresge  
*Rochester Institute of Technology*  
*kjk6690@rit.edu*

Natalie Petruzelli  
*St. John Fisher College*  
*ngp01978@sjfc.edu*

Sponsors: Drs. Eben Kenah and Grzegorz A. Rempala  
*The Ohio State University*  
*kenah.1@osu.edu*  
*rempala.3@osu.edu*

June 13, 2021

## Abstract

COVID-19 epidemics in many parts of the United States and the world have shown unexpected shifts from exponential to linear growth in the number of daily new cases. Epidemics on configuration model networks typically produce exponential growth, while epidemics on lattices produce linear growth. We explore a network-based epidemic model that interpolates between lattice-like and configuration model networks while keeping the degree distribution and basic reproduction number ( $R_0$ ) constant. This model starts with nodes assigned random locations in a unit square and connected to their nearest neighbors. A proportion  $p$  of the edges are disconnected and reconnected in a configuration model subnetwork. As  $p$  increases, we observe a shift from linear to exponential growth. Realistic human contact networks involve many local interactions and fewer long-distance interactions, so social distancing affects both the effective reproduction number  $R_t$  and the proportion of long-distance connections in the network. While the impact of changes in  $R_t$  is well-understood, far less is understood about the effect of more subtle changes in network structure. Our analysis indicates that the threshold between linear and exponential growth may occur even with a small percentage of reconfigured edges. Additionally, the number of total infected individuals in an epidemic substantially increases around this threshold even when  $R_0$  remains constant. This study reveals that implementing and relaxing social distancing restrictions can have more complex and dramatic effects on epidemic dynamics than previously thought.

## 1 Introduction

In a pandemic, accurate predictive models of epidemic dynamics are critical for preparation of public health and healthcare systems, planning interventions, and evaluating the effectiveness of these interventions. However, traditional modeling techniques for predicting the behavior of epidemics have failed to reflect the patterns of infection in the United States. This project uses a network-based SIR model that interpolates between lattice-like networks and configuration model networks in an effort to explain unexpected patterns in the daily incidence of new COVID-19 cases seen at local and regional levels in the United States and other countries.

In an epidemic model, a network may be used to represent stable interactions among individuals [13]. Many network-based epidemic models use configuration model networks, where nodes are assigned degrees (i.e. numbers of neighbors) and then these “stubs” are connected completely at random. Epidemics simulated on configuration model-based networks fit the expected growth pattern seen in the likes of China

and South Korea, where daily new case numbers (epidemic curves) show a clear period of exponential growth before peaking and decaying. Case resurgences are able to be implemented via change points in the case of South Korea, while still fitting the expected trends in the data (Eben Kenah and Grzegorz A. Rempala, private communication, June 2020). However, U.S. case data only initially exemplifies the expected exponential growth pattern before devolving into unexpected patterns of linear growth and plateaus. This is exemplified in Figure 1, which shows the new cases per day in the state of Ohio (Grzegorz A. Rempala, private communication, June 2020).

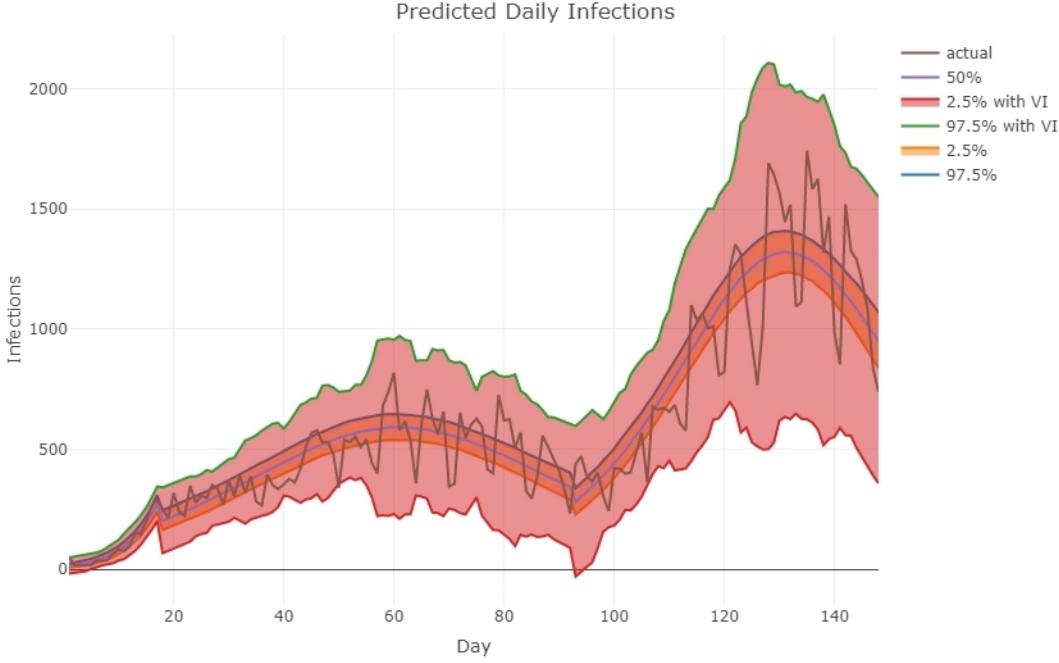


Figure 1: Ohio: New cases per day (epidemic curve) as of August 1<sup>st</sup>

Previous research shows that the pattern is reflective of social changes (caused by COVID-19 regulation and restrictions), not by the network structure or testing saturation [13]. Furthermore, while the patterns largely appear to be a result of the aforementioned social changes, further investigation would be required to determine if the patterns stem from some structure within the epidemic, such as a fractal. This is currently beyond the scope of the project.

While these differing patterns of growth can be implemented into a model by using change points, adding many parameter changes can be difficult and may raise the level of uncertainty in the results. The fact that these patterns are common but difficult to replicate using change points suggests that they are caused by something more fundamental.

Here, we develop and explore a model that is able to reproduce the switch in epidemic curve dynamics from linear to exponential growth without the addition of a change point. Instead, we introduce a lattice-like structure to the network by assigning nodes a random position in the unit square and connecting them to their nearest neighbors with available stubs. This creates networks that are able to reflect linear growth patterns. This new model recreates patterns seen in the COVID-19 data by altering the network structure that the epidemic runs on, rearranging from a lattice-like structure to a configuration model with probability  $p$ , which represents the proportion of interactions one has with someone outside of their immediate community. We analyze the transition between linear and exponential growth by investigating the effect of different values of  $p$ , the level of connectivity of the network,  $R_0$  and  $R_t$  values, and other attributes.

We find that the shift between linear and exponential epidemic curve growth occurs at low values of  $p$ , regardless of the reproductive potential of the epidemic ( $R_0$ ). When examining the final sizes of epidemics as a function of  $p$ , the threshold where there is a clear separation between small, self-limited outbreaks and large epidemics was found to emerge faster at higher values of  $R_0$ . Once we determine an epidemic size that accurately classifies small epidemics and large epidemics, we can investigate the probability of a large epidemic. The results show that small increases in the  $p$  can have large effects on the probability and

final sizes of large epidemics and on the dynamics of an epidemic. Ultimately, this research can potentially explain some of the patterns seen in the COVID-19 pandemic and further explain some of the impacts of a precarious reopening on the U.S. epidemic. This analysis finds that a lattice-like network is likely to be a more accurate representation of the social interactions in the United States during times of social distancing and quarantining behavior, producing the linear growth and plateau patterns seen in epidemic curves like that in Figure 1. This is promising for future research, as much of the network-based modeling for COVID-19 has been simulated on configuration models, which only produce exponential growth dynamics. The addition of the probability parameter  $p$ , which interpolates between these network structures, produces the shift in growth patterns and provides a basis for understanding the transition between linear and exponential case growth.

## 2 Methods

We implemented a network-based Susceptible-Infected-Recovered (SIR) model in Python, where the network can interpolate between a configuration model and a lattice-like structure [16]. The model is structured such that it contains a given number of nodes (representative of a population of people) each having a degree that is meant to show the average number of people they interact with. On a contact network, infection spreads between people in close proximity, represented in the network through edges. Edges can be broken, usually by some social distancing or quarantining behavior, so that infection cannot be sent or received between the two nodes that no longer connect. As infection spreads across the network, each person (or node) is classified as either susceptible (S), infected (I), or recovered (R). Each individual starts in S, moves into I if infected, and moves into R upon recovery or death. Terms and parameters necessary to understand the model are listed in Table 1.

---

Configuration Model	A network created by connecting all stubs in a collection of nodes uniformly at random.
Edge	A link connecting two nodes. In this case, edges represent an instance of contact or proximity between people.
Epidemic Curve	A curve plotted over time that shows the number of new cases of an illness per day.
Lattice-like Network	A network where the nodes are randomly positioned within the unit square, and each node is given a specific degree from some distribution. The edges are then connected based on the location of their nearest neighbors with available connections.
Node	A vertex in a network that can be connected to other nodes by edges. In this case, nodes represent an individual, each with a status defined by the SIR model.
$p$	The probability that any given edge will be rearranged into a configuration model sub-network. In the real world, this represents the percentage of interactions any one person has with someone who lives outside their community.
$R_0$	The basic reproduction number; measures transmission potential of an epidemic. This is the average number of secondary infections produced by a typical case of COVID-19 in a population of susceptible people.
$R_t$	The effective reproductive number at time $t$ ; also measures transmission potential. This considers the fact that the entire population does not remain susceptible over the course of the epidemic. It's the average number of secondary cases in a population where the number of susceptible people decreases over the lifetime of the epidemic.
Rho ( $\rho$ )	The percent of the population chosen at random at the beginning of the simulation that are initially infected (introduces the disease to the rest of the population). This is usually very small.
Stub	A half edge. Two stubs, when attached, form an edge in the network.

---

Table 1: Definitions

## 2.1 Assumptions

We made several assumptions in the development and use of this model, particularly in the use of a compartmental SIR model (susceptible-infectious-recovered) which has commonly been used in the modeling of COVID-19 [13]. They are as follows:

1. Each individual/node in the network has some number of neighbors greater than 0 that they are connected to by an edge, given by a degree (usually determined by a Poisson or negative binomial distribution).
2. Infectious individuals can infect their immediate neighbors at a fixed rate that is greater than 0. The rate of recovery is also assumed to be constant and greater than 0.
3. Infection and recovery times are assumed to be exponentially distributed.
4. There is no latent period between being infected and being infectious to other individuals.
5. Based on the given transmission and recovery rates, nodes can change their status between the compartments of the SIR model.
6. Recovered individuals cannot infect others or be reinfected.

The justification for the above assumptions is as follows. For Assumption 1, the graph-based approach has previously been shown to be a reliable epidemic model [8]. Additionally, Poisson and negative binomial distributions are common degree distributions applied to real world networks. These distributions imitate interacting with your closest neighbors, not the entirety of a social network, thus limiting the number of edges in the model. The fixed rates of infection and recovery are reflective of real-world occupational/spatial interaction. Assumptions 1 and 2 naturally lead into the exponential distribution of assumption 3, since a constant infection rate leads to the probability of infection growing exponentially with time. Although Assumption 3 can be relaxed, this leads to a more complicated model, which is unnecessary given the scope of this paper [17].

In Assumption 4, we remove the possibility of compartments for latent periods, hospitalization, and other states. Ultimately, adding a latent period or more compartments to the SIR model will not affect the results of the model, since adding these will only delay the onset of the infection. This affects the timing of the epidemic, but does not change the overall resulting behavior of the epidemic or the results of the analysis. Thus, leaving them out drastically reduces any unnecessary calculations. Assumption 5 follows from the nature of the SIR model. For Assumption 6, it should be noted that the literature shows viruses similar to COVID-19 give at least partial immunity to future infections [6]. Since the analysis that we conduct focuses on a relatively brief timescale, the possibility of reinfection for the population is negligible. Thus, we assume individuals will not be reinfected in our model.

## 2.2 Model Overview

The model must to be able to show both patterns of exponential growth, typically seen in a configuration model network, and linear growth, produced by a lattice or lattice-like network. Thus, both structures are present in the model and are able to run independently from each other.

### 2.2.1 Configuration Model

The model was first built as a simple stochastic SIR epidemic run on a configuration model network. The configuration model is a network that is made by assigning each node a degree and connecting stubs to form edges until each node meets their desired degree [1, 10, 11]. Using an SIR model from the Epidemics on Networks package [7], we can run epidemics on configuration model networks produced using the NetworkX package [3]. This produced epidemic curves with exponential growth and decay. This portion of the model assumes constant infectiousness and exponentially-distributed infectious periods. This portion of the model

was predominantly tested and used with a Poisson degree distribution with mean  $\lambda$ . Binomial and negative binomial distributions were also tested and produced similar results.

When running SIR epidemics, the `Epidemics` on `Networks` package produces arrays for each of these compartments: susceptible, infectious, and recovered [7]. These arrays update each time a person moves between these compartments. While useful, this produces curves that are more akin to prevalence curves (active cases per day) instead of an incidence curves (new cases per day). In order to find the incidence curve, it is necessary to find the number of new infections in distinct time steps, instead of at the random, often non-integer values where the arrays are updated. This is done by finding the difference between the number of susceptible people at these new time steps and plotting the decrease in the number of susceptible people as the number of new infections. We use “epidemic curve” to refer to incidence curves, not the prevalence curve produced directly from the SIR command.

While building the configuration network for the model, a function was added that allows the user to view the distribution of pairwise distances across the network. Pairwise distances in a network reveal the shortest possible path between any node pair  $(i, j)$ , which are then plotted in a histogram format for convenience. The pairwise distances are found using the `all_pairs_shortest_path_length` function from the `NetworkX` package, which finds the minimum number of edges crossed between any two nodes [3]. For example, if three nodes are connected in a triangle, then the pairwise distance for any pair of nodes is one, even though there exists both a path of length one and of length two between any pair. The ability to analyze the distribution of pairwise distances between nodes permits one to see the level of connectivity of the network. A more connected network has shorter pairwise distances on average, so an infection can spread faster, further, and more reliably. This facet of the model gains further significance later on in the modeling process when we inspect the threshold between linear and exponential growth, as epidemics with different growth patterns have notable dissimilarities in their level of connectivity. Running epidemics on a configuration model thus produces the typical exponential growth-peak-decay pattern that is seen in textbook epidemic curves.

### 2.2.2 Lattice-like Model

The next step was running epidemics in the same manner on a lattice-like network in order to show patterns of linear growth. While lattice structures show linear growth dynamics when an epidemic is run on them, using a pure lattice would not accurately capture variation in the number of social connections that individuals have. In a pure two-dimensional lattice, nodes are positioned equidistant from each other and all have degree four, except for the nodes at the corners (which have degree two) and the edges (which have degree three). By creating a network that is lattice-like, we are able to create linear growth dynamics while retaining variation in the degrees of the nodes. Given a graph, a positional dictionary, and a degree list, this algorithm creates a lattice-like network by connecting nodes sequentially by the shortest relative distance between them and their neighbors until they have reached the degree  $k$  assigned to them by a degree distribution [9, 5]. This algorithm is titled `build_lattice_like_edges` within the code, and it is available in the appendix.

The `build_lattice_like_edges` algorithm takes in a given graph,  $G$ , a positional dictionary that relates every node in  $G$  to a coordinate in the unit square, and finally, a list of desired degrees for each node in  $G$ . First, a dictionary is created that relates each node in  $G$  to a list of squared distances from itself to all other nodes. Since two is the maximum distance squared for any two points within the unit square, the number two is designated as a place holder for the index that would be the distance between a node and itself. This will become useful when looking for minimum distances. Then, a loop iterates over every node in  $G$ . While a node has not yet reached the desired number of connections according to the given degree list, the algorithm does the following: it finds the node with the minimum value in its distance dictionary entry, connects the nodes together with an edge, and replaces the minimum distance with two so that the algorithm cannot connect the nodes together again. If either node reaches its desired number of nodes, the algorithm turns all of its distance dictionary entries to two, thus preventing it from connecting to other nodes.

It is important to note that the nodes are not connected by the shortest absolute distances, as `build_lattice_like_edges` iterates through the nodes in such a way that nodes that are the closest neighbors to another particular node may have already reached  $k$  connections before they can be connected to their absolute closest neighbor. An algorithm based off absolute distance connects the two closest nodes first, and then the second two closest nodes until all nodes have their desired  $k$  connections. While an algorithm was

constructed that did connect nodes according to this method (see Figure 2), it produced networks that were over-connected to the point where it would always produce fast exponential epidemic growth, which was ultimately undesirable.

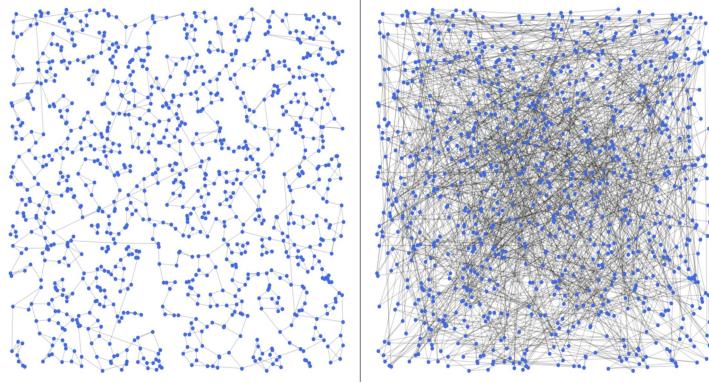


Figure 2: Above are two networks made with the same nodes, node positions, degree distribution, and number of edges. The left algorithm is `build_lattice_like_edges`, which connects edges by iterating through the nodes. The right algorithm connects edges based off absolute distances.

Like the configuration model network, epidemics were run on the lattice-like network created by `build_lattice_like_edges` and the pairwise distances across the network were also calculated. This lattice-like network showed the desired linear growth dynamics.

### 2.2.3 Interpolation Between the Two Models

Since the U.S. COVID-19 epidemic data reveals instances of both linear and exponential growth, it is necessary to use both the configuration model and the lattice-like network to reflect these dynamics. The model combines the two networks by beginning with a lattice-like network with a given degree distribution, usually Poisson. Edges on the network are broken with probability  $p$ , and the resulting stubs are reconnected into a configuration model subnetwork. These reconnected edges will be referred to as long-distance edges. This type of edge does not change the probability of infection across them, as the name might imply. This reconfiguration of edges creates a hybrid of the configuration model and lattice-like networks depending on the desired  $p$ . Thus, a pure lattice-like network is produced when  $p = 0$ , and a pure configuration model network is produced when  $p = 1$ , as seen in Figure 3. No matter what value of  $p$  is used, the degree distribution is the same. Since networks can exemplify distinct characteristics despite having the same degree distribution, we can anticipate varying results from fluctuating values of  $p$  [14].

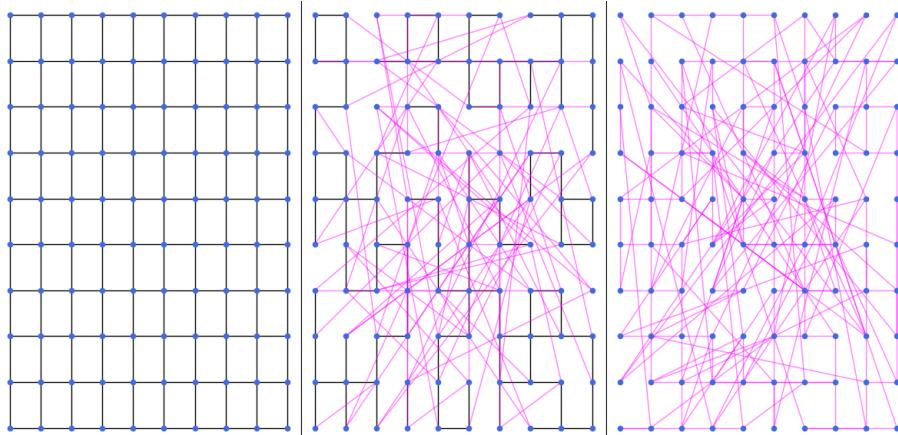


Figure 3: This shows a pure lattice graph changing into a configuration model-based network with probabilities (from left to right) 0, 0.5, and 1.

Ultimately, this model allows us to run epidemics and extract their epidemic curves. We are also able to produce a histogram of pairwise distances, as well as an R-effective curve, revealing the change in  $R_t$  over the course of the epidemic. The final model not only creates epidemics that have linear and exponential patterns of growth, but also permits further analysis of the transition between lattice-like dynamics (linear growth) and configuration model dynamics (exponential growth) by changing the mean degree of the Poisson distribution,  $R_0$ , and  $p$ .

The final function of the model, known as `epi_analyze`, allows a user to input a list of degrees (of a distribution of their choice) and a value of  $p$ , and receive their choice of output, including an epidemic curve, the network the epidemic is run on, a histogram of the pairwise distance distribution, and an R-effective curve. One is able to view both the prevalence curve creating by the SIR model as well as the incidence epidemic curve extracted from the data. The epidemics can also be produced on a linear or logarithmic scale for ease of viewing different growth dynamics. Figure 4 shows the `epi_analyze` function with all of the possible plots drawn. Note that the epidemic produced in Figure 4 is a single epidemic simulation, not a representation of a real-life epidemic.

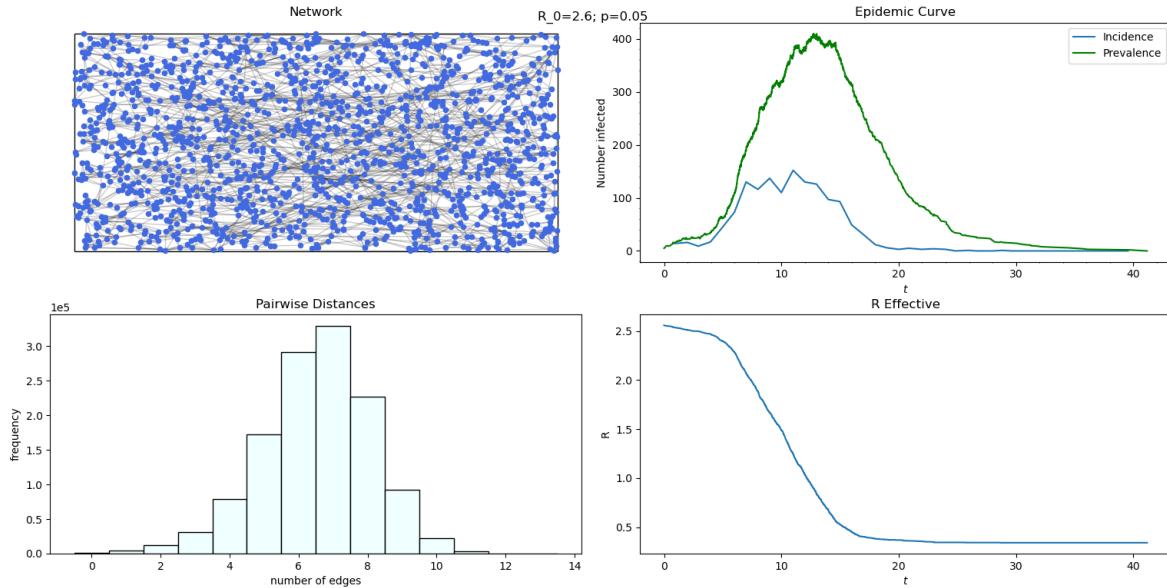


Figure 4: The results from the `epi_analyze` function with all possible outputs

### 3 Analysis

It is important to understand what can cause a shift between the exponential and linear growth of disease incidence, as seen in many parts of the United States and the world, in order to prevent and control such outbreaks in the future. Previous analysis has revealed that the shift seen in Ohio (see Figure 1) is not a result of geographic distribution (i.e. overlapping but distinct epidemics in different regions of the state) or saturation of testing (i.e. reported growth does not represent the true growth of the epidemic) [Eben Kenah and Grzegorz A. Rempala, private communication, June 2020].

The network-based SIR model produced here is successful in producing these shifts in growth dynamics, which is a promising step in the right direction. This section will analyze the threshold between exponential and linear growth in greater detail, and attempt to find a possible explanation for the patterns seen in the U.S. COVID-19 epidemic.

#### 3.1 Final Size and Threshold Analysis

Cursory analysis of the model's outputs revealed that the transition from exponential to linear growth usually occurs at low values of  $p$  that depend on the  $R_0$  given to the model. In order to approximate the location

of the threshold more specifically, the final sizes of a broad range of epidemics at different  $p$  and  $R_0$  values were investigated. The final size of an epidemic is the total number of cases recorded over the lifetime of an epidemic, including the initially infected population. When infection is introduced to the population, there are two possible outcomes: a small outbreak that goes stochastically extinct after infecting a small number of individuals or a large outbreak that infects a substantial fraction of the population. To analyze the probability of a large epidemic, it is important to find an outbreak size that can distinguish between these two possibilities.

Using the `build_lattice_like_edges` algorithm created in Section 2.2, one main data set was simulated to conduct this threshold analysis on. In this simulation, one initially infected person was introduced into a population of 1,600 people. The network had a Poisson degree distribution with a mean of ten. The simulation produced data for three different values of  $R_0$ : 1.5, 2, and 3. These values were chosen in order to show the range of behavior for epidemics of varying levels of severity. An  $R_0$ -value that is around one will naturally show much more variability in epidemic size, with a gradually declining chance of large outbreaks as one gets closer to  $R_0 = 1.0$ . Meanwhile, a higher  $R_0$ -value, like  $R_0 = 3.0$ , is more likely to produce epidemics that impact a greater portion of the population, with a lesser chance of a small outbreak.

Finally, epidemics were simulated at an assortment of probabilities. Probabilities from  $p \in [0, 0.1]$  by increments of 0.0025 and  $p \in [.2, 1.0]$  by increments of 0.1 were chosen. The former range used much smaller increments because the threshold is first seen to emerge at probabilities below 0.1, even at low  $R_0$  values. At each combination of  $p$  and  $R_0$  values, 2,000 epidemics were simulated and their final sizes were recorded, producing a total of 300,000 epidemics for this threshold analysis.

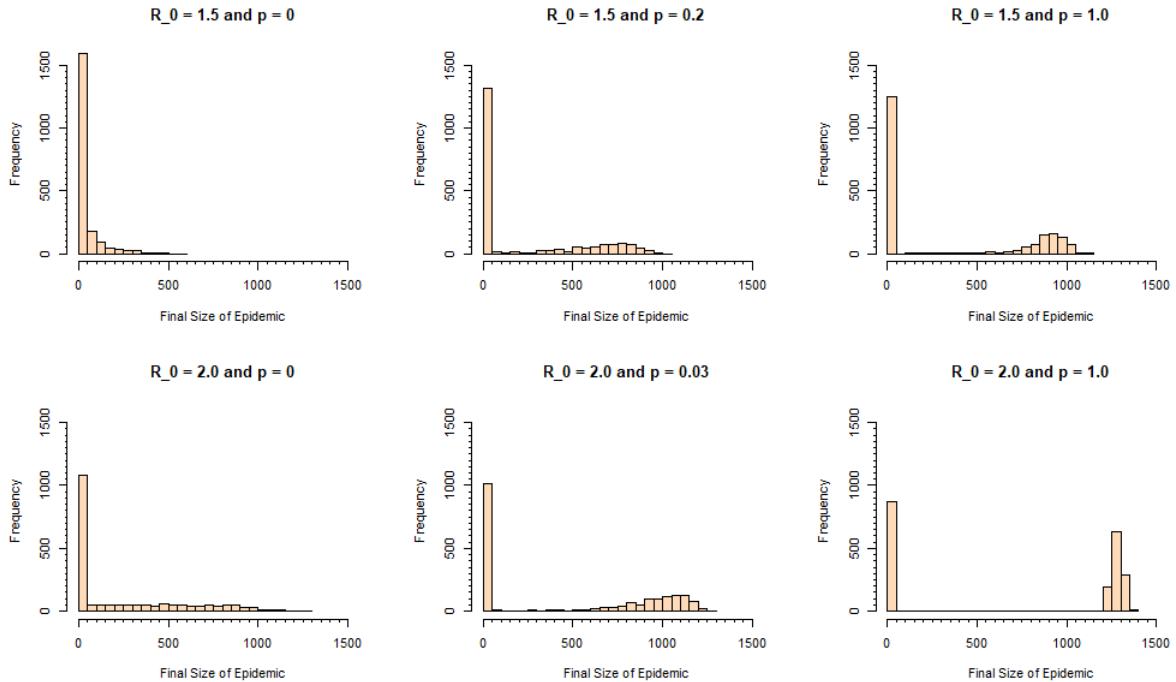


Figure 5: Final Size Distribution Over Three Different Values of  $p$  for  $R_0 = 1.5, 2.0$

For each  $R_0$  and  $p$ , a histogram was produced in R in order to show the distribution of final epidemic sizes as probability increased [12]. Examining the histograms for each  $R_0$  value reveals how altering the network structure by increasing  $p$  changes the epidemic final size distribution. For each  $R_0$ , a threshold emerges as  $p$  increases and a gap in the final size distribution appears, below which are the small outbreaks and above which are the large epidemics. Animations of these distributions over changing probabilities are available as supplemental files. In Figure 5, distributions of epidemic final sizes at selected values of  $R_0$  and  $p$  show the influence of the probability of reconfiguring connections on an epidemic's final size distribution.  $R_0 = 3.0$  is not pictured, as the gap between small outbreaks and large epidemics is immediately visible even at  $p = 0$ . At the lower  $R_0$ -values, the threshold takes more time to emerge. At  $R_0 = 2.0$ , a noticeable break in epidemic size appears as early as  $p = 0.03$ , and increasing  $p$  skews the distribution towards larger epidemics. At  $R_0 = 1.5$ , the break appears at approximately  $p = 0.2$ , with the distribution skewing right

more slowly with increasing  $p$ . Examining the lower and higher ends of the histograms for each  $R_0$  allows for one to pinpoint the final size where this threshold emerges. For  $R_0 = 1.5$ , the break occurs at around 80 total cases, while the breaks for  $R_0 = 2.0$  and  $R_0 = 3.0$  occur at 60 and 50 total cases, respectively.

## 3.2 Probability and Final Size of Large Epidemics

Using the threshold numbers outlined in the previous section, probabilities of large epidemics can be found for our experimental  $R_0$  values and probabilities. The plot in Figure 6 shows the probability of a large outbreak defined by the thresholds for a given combination of parameters. It should be noted that for Figure 6 and later in Figure 7, the values of  $p$  were incremented by 0.1, excluding the additional data points on the interval of  $p \in (0, 0.1)$  by 0.0025 used in the histograms of Section 3.1. Excluding these points does not impact the overall trend of the graph, and thus have been removed to improve comprehension. Regression lines have been added to adjust for the minute impact caused by removing these points.

As seen in Figure 6, there is an increase in the probability of having a large epidemic by up to 20% for  $R_0$  values of 1.5 and 2.0 within the  $p$  interval  $[0, 0.4]$ . This increase is more dramatic for lower values of  $R_0$  and  $p$ , which is important to note, since we expect the baseline value of  $p$  in a real-world network to be between 0 and 0.1. After this interval, the higher values of  $p$  have roughly the same probability of an outbreak. We anticipate the line of  $R_0 = 3.0$  to be flat, since at  $p = 0$ , the majority of epidemic final sizes are either under 25 or above 1300 infected individuals. Thus, there is little potential for increased probability of infection by changing the graph structures. We expect the lines to plateau at higher values of  $p$ , since at a certain point the random connections overpower any innate structure in the network. Therefore, we can attribute the graph's lack of monotonicity to random variations in the finite number of simulations.

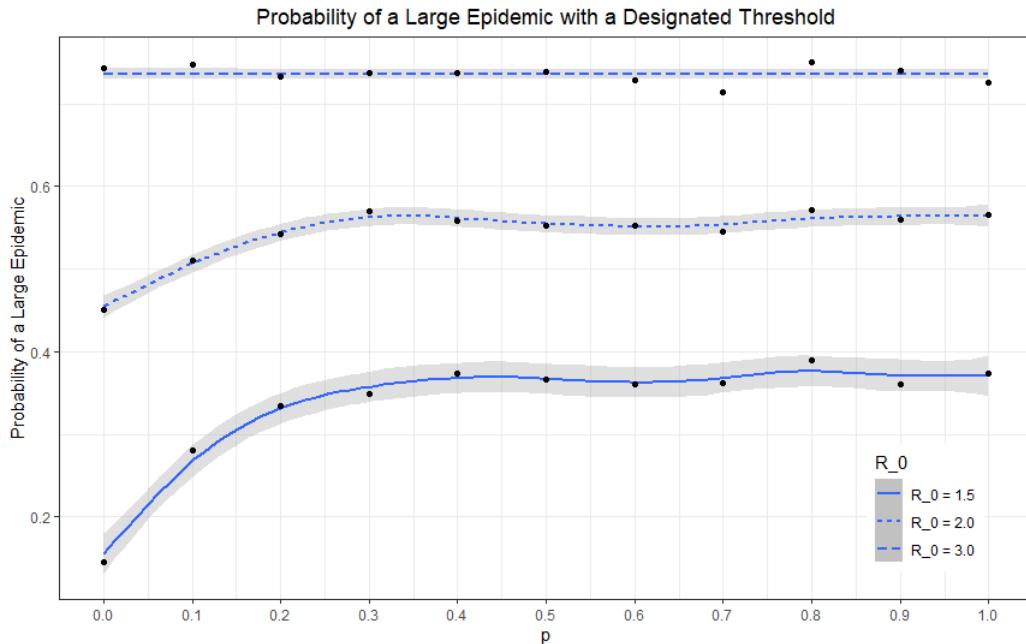


Figure 6: Probability of a Large Epidemic versus  $p$

In a real-world scenario, depending on the approximate  $R_0$  of the disease and what value of  $p$  accurately represents the connections between people through which infection can spread, the probability of a large epidemic may be smaller than that predicted by many simple epidemic models. There appear to have been several early introductions of COVID-19 into the United States that failed to expand into a fully-fledged epidemic with exponential growth, which would be consistent with an  $R_0$  well below three and a relatively low value of  $p$  [2].

In addition to analyzing the probability of a large outbreak occurring, the average final size of large epidemics may also be calculated given the defined threshold from Section 3.1. Figure 7 shows a plot of the average size of epidemics that are larger than the threshold epidemic sizes for each  $R_0$  value. For  $R_0 = 3.0$ ,

the size of outbreaks shows little variability, staying more or less at a constant average size of approximately 1,500 cases for all  $p$ . This aligns with our findings from Figure 6, which determined that the probability of a large epidemic occurring at  $R_0 = 3.0$  is almost certain regardless of the value of  $p$ . For  $R_0 = 2.0$ , the size begins at a starting average of 532 cases, and it more than doubles as  $p$  approaches 0.1, showing significant growth in average final size as  $p$  only marginally increases. We again note that the large increase in final sizes below  $p = 0.1$  is critical, since the real-world baseline value of  $p$  is expected to be in this range. For  $R_0 = 1.5$ , there is a steady increase in average final size until  $p = 0.4$ , where it plateaus from that point onward. The change in epidemic final sizes with  $p$  is more dramatic than the change in the probability of a large epidemic, especially for  $R_0 = 2.0$ , but both are affected by  $p$ . The effect of social distancing and other health related policies would cause the real world network reconfiguration probability,  $p$ , to be lower than average, and as seen in Figures 6 and 7, even a small increase in  $p$  can cause a significant jump in the probability and final size of a large epidemic. This strongly suggests that it is critical that precautions are taken seriously during a pandemic, as a local epidemic could balloon in size even when a small portion of connections are reestablished.

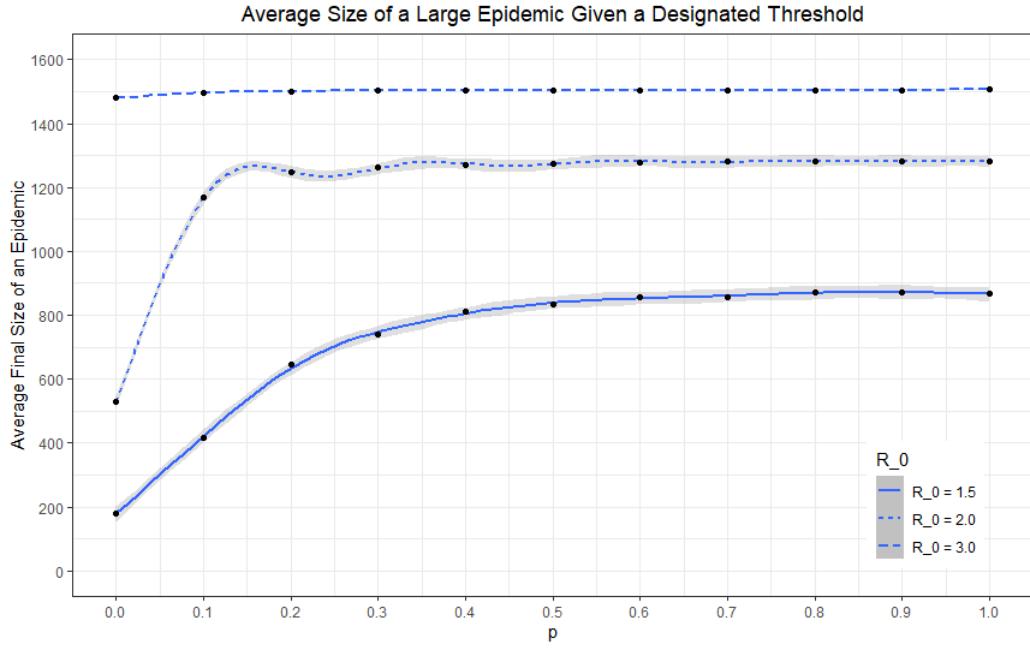


Figure 7: Average Size for Large Epidemics versus  $p$

### 3.3 Regression

In addition to examining the threshold between linear and exponential growth, the epidemic data can be analyzed by estimating a regression trend line. To understand how linear or exponential the growth of new cases of a particular epidemic may be, we first select the portion of data points that are halfway between the start of the epidemic and the peak in infected incidence numbers. We choose this portion near the beginning of the epidemic so that the depletion of susceptible people has not slowed down the spread of the epidemic significantly. Afterwards, an exponential least-squares regression line is calculated according to the formula  $y = \rho N + ae^{bx}$ , where  $\rho$  is the percent of the population that is initially infected,  $N$  is the number of node in the network, and the coefficient  $b$  describes the scale of exponential growth in the epidemic. When acquiring the data, ten epidemics were simulated at every combination of  $R_0$  and  $p$ , with the probabilities selected between 0 and 1 by increments of 0.01 and the same  $R_0$ -values as Section 3.1. Finally, a plot of  $b$  versus  $p$  was produced for each of the three  $R_0$ -values, as seen in Figure 8.

Prior to evaluating the threshold for large epidemics, the curves of these plots did not appear as anticipated, particularly for  $R_0 = 1.5$ . The plots were expected to show an increase in the exponential regression coefficient  $b$  as  $p$  increased, reflecting a transition from linear to exponential growth as more long-distance connections were added into the graph. However, with  $R_0 = 1.5$  there is a bump in the data at very small  $p$  values, creating an apparent increase in the regression exponent at small values of  $p$ . Upon looking

more closely, it was determined that this is due to the fact that there was insufficient data to accurately produce a trend line. To counteract this error, there was a restriction put on the epidemic that would ensure the trend line would be made with sufficient data points. However, the bump in the data remained.

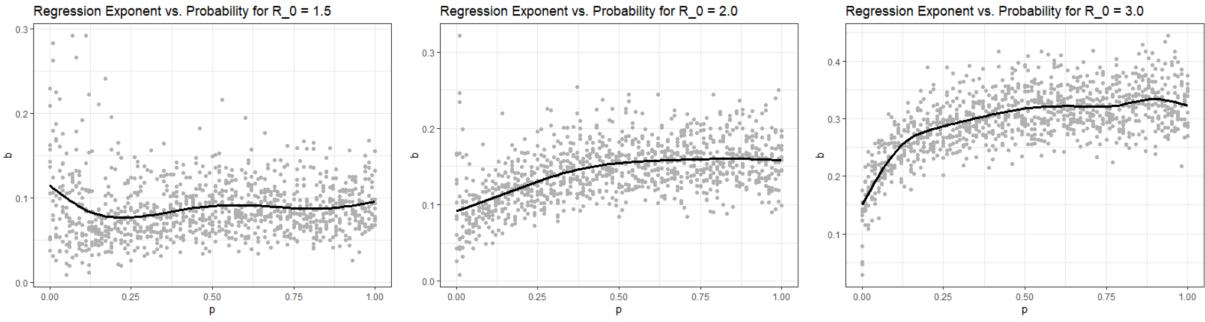


Figure 8: Exponential Regression Coefficient versus  $p$  for  $R_0 = 1.5, 2, 3$  [from left to right] before added restrictions

After analyzing the threshold between large and small epidemics, an additional restriction was put on the epidemic curves to guarantee a large epidemic occurs. This restriction is that the epidemic must achieve a minimum of total infected individuals, which is designated by the threshold values from Section 3.2. This restriction was put in place to help filter out self-limited small outbreaks that were artificially increasing the regression exponent at low values of  $p$ . After rerunning the data set, the bump in the  $R_0 = 1.5$  exponential coefficient plot was no longer there, while the desired increasing data trend remained for the other  $R_0$  values. The corrected regression plots are shown in Figure 9. Interestingly, although the trend line for  $R_0 = 1.5$  stays relatively constant, the variance of the line diminishes with  $p$ . This means that for  $R_0 = 1.5$ , the intensity of the epidemics on average does not change, and outlier data points are suppressed as  $p$  increases. Additionally,  $R_0 = 2.0$  and  $R_0 = 3.0$  have curves that exhibit an increasing exponential trend with constant variance. These trend lines show that as  $p$  increases, the shift from linear to exponential growth occurs faster and with greater intensity at higher  $R_0$  values. This fits the expectations of the model, as it is aimed to produce exponential growth more reliably at high  $p$  and  $R_0$  values and linear growth more frequently at low  $p$  and  $R_0$ .

It is important to note that the shift from linear to exponential growth cannot be explained by shift in  $R_0$  values. Although, the intensity may change with  $R_0$ , the linear growth phenomenon is due to the shift in pairwise distance distributions and intrinsic network structure. As  $p$  increases, edges that were once locally connected are reconfigured anywhere in the network. As a consequence, less edges are needed to be crossed to get from one part of the network to another and this is defined by a shift in pairwise distances. Within the terms of the epidemic, this means that infection spreads faster and more reliably.

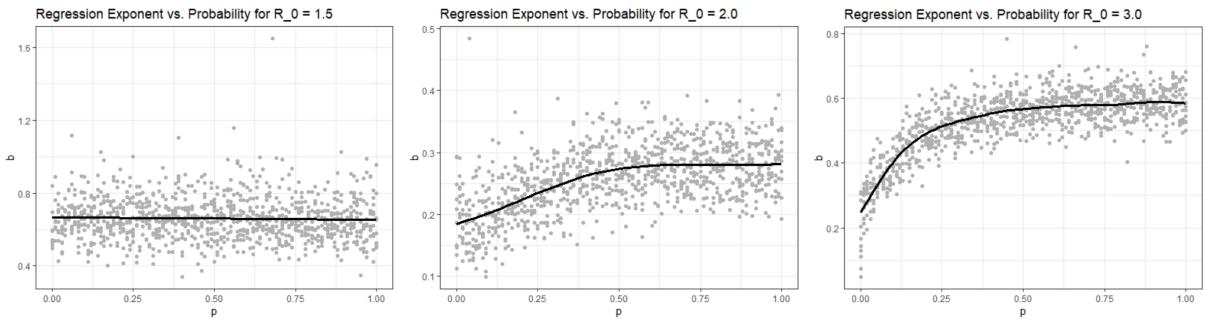


Figure 9: Exponential Regression Coefficient versus  $p$  for  $R_0 = 1.5, 2, 3$  [from left to right] after added restrictions

## 4 Conclusion

The model produced in this research was not only able to produce epidemics of linear or exponential growth, but was also able to show a transition in growth patterns by interpolating between different network structures. By creating a parameter  $p$  for the probability of long-distance connections, the model is able to show these different growth dynamics as well as the transition between them. Beyond this, the model is able to produce other aspects of an epidemic for analysis. This includes both incidence and prevalence curves, R-effective curves, and pairwise distance distributions. The model thus allows one to see a multifaceted view of an epidemic, further shedding light on the relationship between epidemic dynamics and the structure of the underlying network.

Analyzing the final size distributions of epidemics in this model allowed us to identify a clear threshold epidemic size at each  $R_0$  value. By determining the final size at which this separation between large epidemics and small outbreaks occurs for each  $R_0$ , we are able to understand how this threshold relates to the probability and final size of a large epidemic. This model was able to replicate the transition between exponential and linear growth that occurred in the state of Ohio around the time that social distancing measures began. It is possible that the network of connections in the state resembled a lattice-like network with enough long-distance connections to support exponential growth. Under social distancing, these connections became more local (i.e. lower  $p$ ) and there was a transition to linear growth with a much longer plateau than would be expected under exponential growth and decay. If true, this would also explain why multiple early introductions of COVID-19 to the United States resulted in apparently self-limited outbreaks. This analysis suggests that the United States' network of interactions is best modeled by a lattice-like structure instead of the configuration model typically utilized in epidemic modeling.

These results have important implications in the real world COVID-19 epidemic. If the U.S. epidemic is modeled using a network structure that is inaccurate compared to the social interactions and distancing measures that are occurring in real life, then predictions of case numbers and predictions could be skewed, leaving health care systems and populations alike expecting an epidemic that is far different than the one that will ultimately happen. In addition to social networks being better represented by a lattice-like structure during this time, this research suggests that a small value of  $p$  could exist in the real world, as increased social distancing measures means that there is a low probability of long-distance contact between people. As the United States moves towards reopening, the value of  $p$  in the network structure will increase due to increasingly diverse social interactions, creating a shift towards exponential growth of new cases. As seen in Figures 6 and 7 respectively, there is a greater probability of experiencing large outbreaks and a much higher final case count as  $p$  increases by even small amounts. As these connections are reestablished in the United States, the network best representing our social interactions will move further away from the lattice-like shape and more towards a configuration model, yielding exponential growth and a more severe epidemic. It is crucial to keep this in mind as the United States and other countries ease social distancing restrictions.

## 5 Further Research

Further research can be conducted to examine why these trends occur. While the simulations were able to model a shift between linear and exponential growth as a function of  $p$ , it is still not entirely certain what real world events may have caused the  $p$  value to change. While the introduction of social distancing and quarantining measures definitely contributed to the switch in network structure, other events may have played a part in this. Other countries have also engaged in these measures, many of them committing to social distancing with far more zeal than the United States, and are still seeing mainly exponential growth patterns in their epidemic curves. Although it could just be that the United States has a remarkably different geographic arrangement of its population, it is also possible that other factors could be exacerbating these unexpected growth dynamics.

Regardless, most epidemic research is conducted on the assumption that configuration models are the most accurate representations of contact-spacial networks, thus producing the expected exponential patterns of growth seen in textbook epidemic curves. As the present growth of the epidemic in the United States seems to be better-represented by a lattice-like network, it would be useful for epidemiologists and

epidemic-models to explore network structures like the ones analyzed here. This could provide a more accurate description of the expected dynamics of epidemics for the future, and better prepare the United States for future epidemic scenarios.

Furthermore, analysis of the epidemic simulations could be refined by utilizing dynamic survival analysis [4]. While this research maintained constant transmission and recovery rates, these parameters could be changed in order to reflect the transmission and recovery of COVID-19 in different areas of the nation (Eben Kenah and Grzegorz A. Rempala, private communication, June 2020) [13]. Dynamic survival analysis has the potential to more accurately predict parameters for the model based on real-time epidemic data from specific areas of the United States. While this was utilized briefly within the research, it warrants more exploration.

Finally, models using small-world epidemics could be looked at with a similar analysis [15]. It has already been shown in Figure 10 that the small world model exhibits linear and exponential epidemic growth similar to our lattice-like interpolation model. The most important difference between the two is that nodes in our model can have different degree distributions. A more detailed investigation of the relationship between these two models is also warranted given the discoveries of this research.

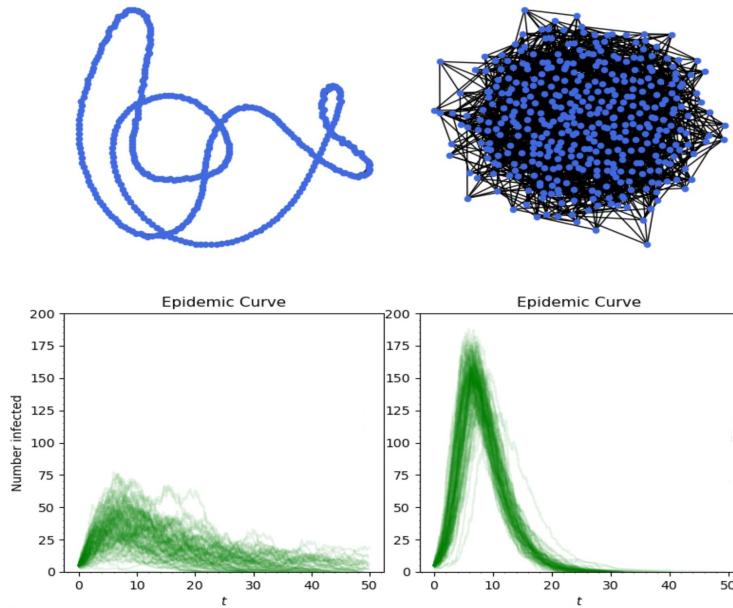


Figure 10: Small-world epidemic prevalence curves with  $p = 0$  on the left and  $p = 1$  on the right.

## 6 Appendix: Lattice-like Network Algorithm

Below is the code to build a graph with lattice-like edges as described in section 2.3.2.

```

1  def build_lattice_like_edges(G, pos, deg_lst):
2      """
3          This function adds edges to a network based on a given degree list, such that the closer
4              two nodes are to one another, the more likely they are to be connected. The graph with
5                  the added edges are returned
6      """
7
8      #Raise an error if there is not a degree given for each node
9      if len(deg_lst) != G.number_of_nodes(): raise ValueError("The degree list must have
10          the same number of values as nodes in the graph")
11      G, relabel_dict = relabel_nodes(G)
12      pos = relabel_pos(pos)
13      # A distance squared dictionary is made with format dist[node]=[{(d_0)**2, (d_1)**2, ...,
14          (d_n)**2}]
15      dist_dict = get_node_to_dist_dict(pos)
16      #For every node, check that it does not already have its desired number of
17      #connections, and if not find the next closest node with an available stub and
18      #connect them together. Repeat until the node has meet its desired number of

```

```

15 #connections. If there are no nodes to connect to, move onto the next node in the
16 #network.
17 for source in G.nodes():
18     while G.degree[source] < deg_lst[source]:
19         if dist_dict[source] != [2] * G.number_of_nodes():
20             target = dist_dict[source].index(min(dist_dict[source]))
21             if G.degree[target] == deg_lst[target]:
22                 dist_dict[target] = [2] * G.number_of_nodes()
23                 for node in G.nodes():
24                     if node == source:
25                         dist_dict[node][target] = 2
26                     else:
27                         dist_dict[node][target] = 2
28             else:
29                 break
30     G = nx.relabel_nodes(G, relabel_dict)
31 return G

```

## References

- [1] E. A. BENDER AND E. CANFIELD, *The asymptotic number of labeled graphs with given degree sequences*, Journal of Combinatorial Theory, Series A, 24 (1978), pp. 296–307.
- [2] A. S. GONZALEZ-REICHE, M. M. HERNANDEZ, M. J. SULLIVAN, B. CIFERRI, H. ALSHAMMARY, A. OBLA, S. FABRE, G. KLEINER, J. POLANCO, Z. KHAN, B. ALBURQUERQUE, A. VAN DE GUCHTE, J. DUTTA, N. FRANCOEUR, B. S. MELO, I. OUSSENKO, G. DEIKUS, J. SOTO, S. H. SRIDHAR, Y.-C. WANG, K. TWYMAN, A. KASARSKIS, D. R. ALTMAN, M. SMITH, R. SEBRA, J. ABERG, F. KRAMMER, A. GARCÍA-SASTRE, M. LUKSZA, G. PATEL, A. PANIZ-MONDOLFI, M. GITMAN, E. M. SORDILLO, V. SIMON, AND H. VAN BAKEL, *Introductions and early spread of sars-cov-2 in the new york city area*, Science, 369 (2020), pp. 297–301.
- [3] A. A. HAGBERG, D. A. SCHULT, AND P. J. SWART, *Exploring network structure, dynamics, and function using networkx*, in Proceedings of the 7th Python in Science Conference, G. Varoquaux, T. Vaught, and J. Millman, eds., Pasadena, CA USA, 2008, pp. 11 – 15.
- [4] W. R. KHUDABUKHSH, *Dynamic survival analysis*, 2020.
- [5] K. KUULASMAA, *The spatial general epidemic and locally dependent random graphs*, Journal of Applied Probability, 19 (1982), pp. 745–758.
- [6] S. F. LUMLEY, J. WEI, D. O'DONNELL, N. E. STOESSER, P. C. MATTHEWS, A. HOWARTH, S. B. HATCH, B. D. MARSDEN, S. COX, T. JAMES, L. J. PECK, T. G. RITTER, Z. D. TOLEDO, R. J. CORNALL, E. Y. JONES, D. I. STUART, G. SCREATON, D. EBNER, S. HOOSDALLY, D. W. CROOK, C. P. CONLON, K. B. POUWELS, A. S. WALKER, T. E. A. PETO, T. M. WALKER, K. JEFFERY, D. W. EYRE, AND O. U. H. S. T. GROUP, *The Duration, Dynamics, and Determinants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibody Responses in Individual Healthcare Workers*, Clinical Infectious Diseases, (2021). ciab004.
- [7] 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*, Journal of Open Source Software, 4 (2019), p. 1731.
- [8] J. C. MILLER AND E. M. VOLZ, *Incorporating disease and population structure into models of sir disease in contact networks*, PLoS ONE, 8 (2013), p. e69162.
- [9] D. MOLLISON, *Spatial contact models for ecological and epidemic spread (with discussion)*, J Roy Stat Soc, 39 (1977), pp. 283–326.
- [10] M. MOLLOY AND B. REED, *A critical point for random graphs with a given degree sequence*, Random Struct. Algorithms, 6 (1995), pp. 161–180.

- [11] M. E. J. NEWMAN, S. H. STROGATZ, AND D. J. WATTS, *Random graphs with arbitrary degree distributions and their applications*, Physical Review E, 64 (2001).
- [12] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [13] OSU/IDI COVID-19 RESPONSE MODELING TEAM, *Predicting covid-19 cases and subsequent hospital burden in ohio*. [https://idi.osu.edu/assets/pdfs/covid\\_response\\_white\\_paper.pdf](https://idi.osu.edu/assets/pdfs/covid_response_white_paper.pdf), April 2020.
- [14] Y. SHANG, *Distinct clusterings and characteristic path lengths in dynamic small-world networks with identical limit degree distribution*, Journal of Statistical Physics, (2012).
- [15] S. H. STROGATZ AND D. J. WATTS, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), pp. 440–442.
- [16] G. VAN ROSSUM AND F. L. DRAKE JR, *Python reference manual*, Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [17] H. Z. ZHILAN FENG, DASHUN XU, *Epidemiological models with non-exponentially distributed disease stages and applications to disease control.*, Bulletin of mathematical biology, (2007).