Scientific Computing with Python

Group Project

Final Report - 12/09/2020

The SIR Model of Disease Spread: Simulations and Extensions

Cole Irwin

Daniel Gold

Shruthi Ramesh

Abstract

The SIR Model of disease spread captures some basic yet insightful characteristics of a pandemic. However, the basic SIR model does not take into account notions of space, nor the stochastic nature of pandemics. In this report, we test several variations of the basic SIR Model that account for the spatial nature of pandemics, as well as their stochastic nature and assess how well the SIR model fits to real data.

1 Introduction

The SIR model is used to describe the dynamics of infectious disease, and is comprised of three compartments, S for susceptible, I for infected and R for removed. In addition to a brief discussion of our original findings in the midterm checkpoint, which considered the simple discrete and ODE SIR model configurations, this report contains several variations of the basic SIR model that focus on elements of social interaction not captured by the basic SIR model.

First, we have a simple stochastic model for the spread of disease throughout a homogeneous community of a fixed size. Time is discrete and individuals may meet a fixed number plus a random number of other individuals per day. Second, we move on to spatial considerations. For the discrete case, we simulate interactions in grids where infected individuals transmit the disease to agents within an infection radius. For the continuous ODE/PDE case, we introduce a spatial PDE system, create vectorized operations to turn it into an ODE system, and then solve for a variety of initial conditions and diffusion parametric values. Third, we propose an extension to consider an SIR model built on a network, generated by a stochastic block model. Finally, we fit the SIR model parameters b, k to COVID-19 data in the US.

This report is structured as follows. Section 2 reports the results from the basic SIR model and its variations. Section 3 discusses the performance of the SIR Model with real data. Section 4 concludes with limitations in our analysis and directions for future research.

2 Variations of the SIR Model

2.1 Results from the basic SIR Model

In the simple discrete configuration of the SIR model, we infect 0.1% percent of the population. Under the assumption that everyone interacts with everyone else, every infected person spreads the disease to a susceptible individual with probability b per given time period. Similarly, for each time period, infected agents recover with probability k. It is to be noted that the three states are mutually exclusive and collectively exhaustive, and the population is fixed at a constant N. We were able to show that various choices of parameters b and k provided us with seemingly accurate depictions of the behaviour of infectious disease.

The non-spatial (original) ODE system at hand is defined as below:

$$\dot{x} = -bxy,\ \dot{y} = bxy - ky, \text{ and } \dot{z} = ky$$

where x(t), y(t) and z(t) are the proportion of the population that is S, I, or R at time t, respectively. The parameter $b > 0$ is the rate at which individuals go from susceptible to infected and k > 0 is the rate that individuals go from infected to removed. There is also related parameter, $R_0 = b/k$ which is the 'basic reproduction number' as in Weiss (2013). This is because it is estimating the number of people an infected person will infect (based on b) before they recover (based on k).

We found that there are 3 behavioral outputs of this model when beginning with initial condition with most of the population susceptible and very few infected: 1) Everyone gets infected at some point, 2) The virus is eradicated off after herd immunity is established, and 3) The virus is eradicated immediately. The behavior of 1) is found for combinations of parameters b, k such that R_0 ¡ 1, while the behavior of 2 is found for b, k such that $R_0 > 3$. Finally, the behavior 2) occurs for b, k such that R_0 is between 1 and 3.

2.2 Discrete Time Stochastic SIR Model

The SIR model, although useful, ignores the apparent stochastic nature of epidemic growth and spread arising from random social interactions. This can be captured by including a probabilistic element of social interaction i.e. to modify it as a simple discrete time stochastic epidemic model, based on Tuckwell and Williams (2007).

The model at hand is inspired from the classical discrete time stochastic model of the chain-binomial type which is explained here. We assume that the population is fixed at n, and time is measured in discrete epochs. Individual i encounters a fixed number n_i of other individuals each day, drawn randomly from the population. Individual i also meets a randomly chosen and random number $M_i(t)$ of other individuals over $(t, t+1]$. If an individual becomes infected, he remains in such a state for R consecutive time points and then recovers.

Let $Y_i(t)$ be the number of individuals who are infected at t and have been infected for exactly i time units, $i = 0, 1, \ldots, R-1$. Let $X(t)$ and Z(t) be the number of individuals who are at time

susceptible and recovered respectively. These set of states can be represented as a Markov Chain. If an individual is infected or recovered at time t, then the state space is determined with probability 1 at time t+1. However, if he is susceptible, it can be shown that

$$P(Y_0^i(t+1) = 1 | N_i(t), X^i(t) = 1, Y(t) = y, \dots) = 1 - (1 - \frac{py}{n-1})^{N_i(t)}$$

We now run simulations of this model and present the key results, as shown in figures 1 through 4.

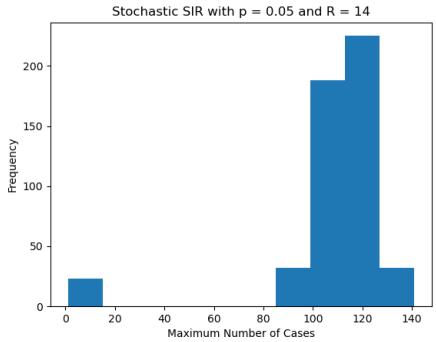

Figure 1: Peak Cases for p=0.05 and R=14

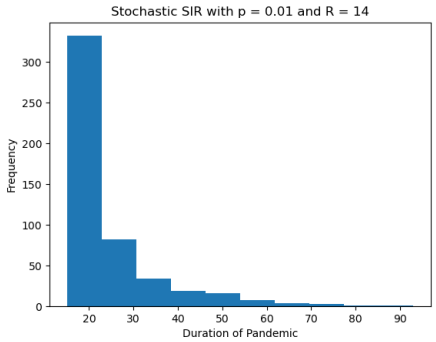

Figure 2: Duration for p=0.01 and R=14

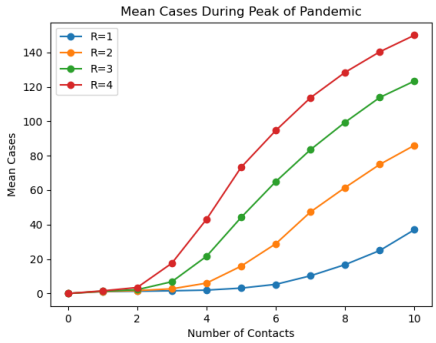

Figure 3: Mean Peaks for Rs and Ns

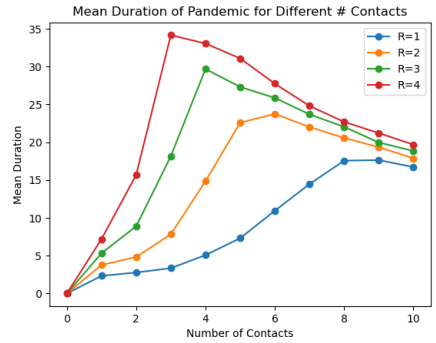

Figure 4: Mean Durations for Rs and Ns

In figures 1 and 2, we have $n = 200$, $N_i(t) = 4$ and one initially infected person. This simplification does capture the basic points we want to establish and we only report the results for this case here. It is striking how the same parameter set can lead to very few new cases or to a large outbreak in which nearly one third of the population becomes infected. Moreover, the variability in the duration is also striking, especially given that contacts are constant.

Figures 3 and 4 show the effects of increasing the number of contacts per day on the mean maximum

number of cases and the mean total duration of the epidemic, with $p = 0.1$ and $n = 200$ averaged over 500 trials. It appears that there is little benefit in reducing R when the number of contacts is small or very large. For intermediate values of N, large reductions in total number of infected individuals can be effected by reducing R. These results present some important effects that cannot be discerned by the deterministic model.

2.3 Spatial Model

Another way of modelling social interaction is by adding a spatial component to the basic SIR model. We do this for both the discrete and ODE models.

2.3.1 Discrete Model

This model contains, in addition to the disease state, a position for each agent given by $pos = (x, y)$, where $0 <= x <= 1$ and $0 <= y <= 1$. At each time step, each individual takes a step of length p in a random direction. If the individual would step outside the domain, they remain in the same location. After an individual takes the random step, they interact with all other individuals within radius q. It can be shown that $b = \pi N q^2$, where N is the population size. We now run simulations of this model. Using this relationship and based on our results from the basic SIR Model, we choose $b = 0.075$ and $k = 0.1$. The results from our simulations are summarized in figures 5 and 6.

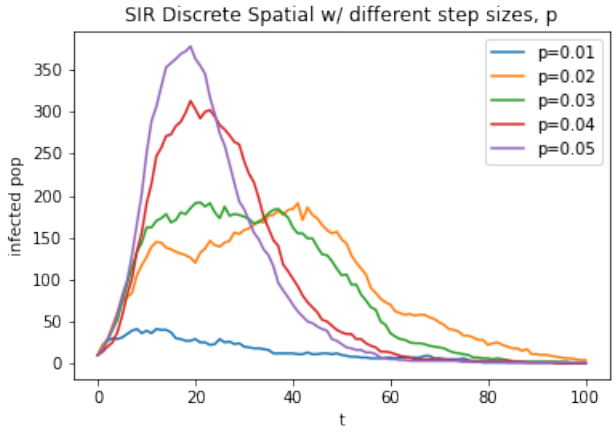

Figure 5: Different Step Sizes

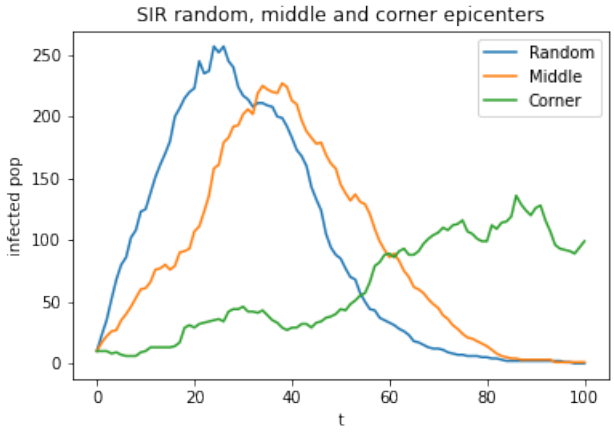

Figure 6: Different Epicenters

Figure 5 shows the counts of infected people throughout the simulation for a range of step sizes. It appears that there are very few infected people throughout the simulation for very low values of p

(0.01). For relatively higher values of p (above 0.04), the number of infected cases rapidly increases in the initial stages of the epidemic and then fall rapidly soon after. For intermediate values of p (0.02 and 0.03), however, the curve is flatter; infection spread and recovery are both more gradual.

We choose $p = 0.03$ for our next simulation, where we analyze the effect of different epicenters on the trajectory of the infected population. We find that the disease spreads rapidly when the infected population is distributed at random in the spatial grid, but also ends the soonest. The case where the initial infected population is in the center has slower spread of infection, a lower peak, and a higher duration of the pandemic. Finally, the disease spreads the slowest when all the infected people are at a corner, but the number of infected people increases steadily. In other words, the pandemic is long drawn out when the initial infection is contained in a corner.

2.3.2 ODE Model

The following PDE system describes an update to our original ODE system, adding a spatial component on a MxM grid:

$$\begin{aligned}
\frac{\partial s(x,t)}{\partial t} &= -bs(x,t)i(x,t) + pLs(x,t) \\
\frac{\partial i(x,t)}{\partial t} &= bs(x,t)i(x,t) - ki(x,t) + pLi(x,t) \\
\frac{\partial r(x,t)}{\partial t} &= ki(x,t) + pLr(x,t)
\end{aligned}$$

where L is the 2-dimensional Laplacian grid of size $M^2 x M^2$. Note that the system is very similar to what we had before, except that 1) s, i, r are now functions of both x (position) and t (time), and 2) Each term has an additional diffusion term with parameter p and the Laplacian grid applied to our s, i, r. Note that the PDE system above can be turned into a system of ODEs if we are able to vectorize S, I, R, removing the calculation's dependency on position x. We can rewrite the above and calculate dSIRdt as follows:

$$\frac{ds(t)}{dt} = -bs(t) * i(t) + pLs(t) \tag{1}$$

$$\frac{di(t)}{dt} = bs(t) * i(t) - ki(t) + pLi(t) \tag{2}$$

$$\frac{dr(t)}{dt} = ki(t) + pLr(t) \tag{3}$$

where s,i,r are m^2x1 vectors and * denotes element wise multiplication.

While considering parameters b=.15, k =.1, p =.5, we look at the role that the initial condition plays in the behavior of our system. We test 3 cases: 1) an initial outbreak (I(0) = .00001) in the top left corner of our grid, 2) an initial outbreak in the center of our grid, and 3) a smaller but randomly distributed outbreak across our grid. In the randomized case 3), each cell has a 40% chance to have an outbreak, and if it has an outbreak, it is calculated as I(0) = .00001/#Outbreaks. This was done so that, across our entire MxM grid, we have the same total percentage of the population infected at the start. For the center case 2), we also look at the role our diffusion parameter p plays in the behavior of our system when we adjust it and hold b, k, and initial conditions constant. We visualize all of this by looking at the final R value of our grid system - this will be the amount of the population removed after 300 days in each respective grid point. Note that all of these simulations were run on a 50x50 grid rather than a 200x200 grid due to the run time of our code.

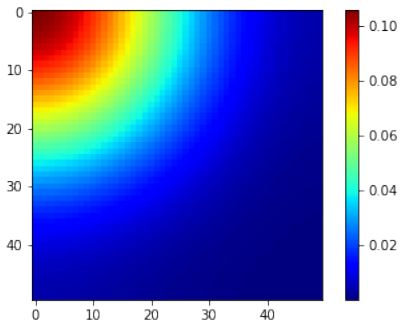

Figure 7: Corner Initialization

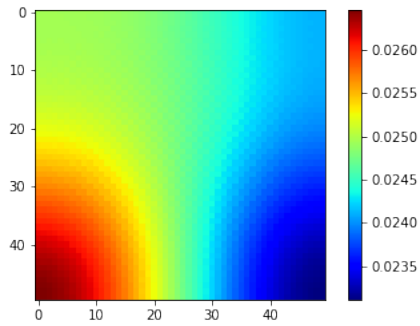

Figure 8: Random Initialization.

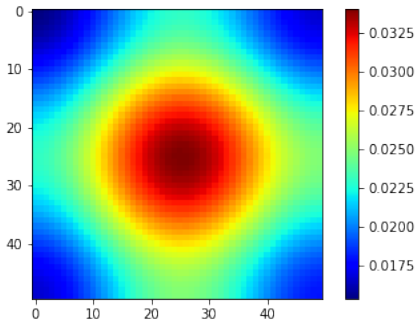

Figure 9: Center Initialization, p = .5.

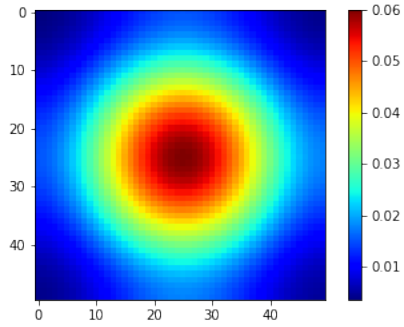

Figure 10: Center Initialization, p = .25

We can see that the initial value plays a large role in the behavior of the entire system. When the initial outbreak is localized, the worst of the infection is radially distributed about the outbreak point. The areas that are far away from the outbreak at hardly impacted, given our parameters. For the randomly distributed case, the behavior is largely determined by any clusters of grid points that may have appeared when the random points were generated. In all trials of our random initialization, there is a smaller variation of end results, which can be seen in the figures' scale axis.

We also find that larger diffusion parameter p values are associated with an faster spread of the infection, which yields a greater radius of impact. This said, population removed is greater when the p value is small. This can be explained logically as the virus being more self contained in a given geography, working its way though that population rather than spreading out elsewhere. In a real world setting, this could be seen as having travel restrictions in place such that infected people can not carry the disease elsewhere. Instead, they are subject to infect others in their original geography.

2.4 Spatial Cluster Model using SBM

In our default spatial model, we consider a spatial domain defined by the grid $[0, 1] \times [0, 1]$, where all agents are allowed to interact with all other agents. This model of spatial interaction might be useful for considering the spread of disease in a closed space, such as an island or among micro-organisms on a petri dish, for example, but is less useful when considering the transmission of disease between different communities, such as between cities and towns. In this section, we describe and implement a spatial SIR model that captures this dynamic, as described in Paeng and Lee (2017).

In this model, we assume the domain is divided into distinct districts and construct a graph, G, where each node represents a district. In order to create a graph that replicates the community dynamics described above, we select the stochastic block model (SBM) which generates k clusters and randomly connects nodes within each cluster with nodes from another cluster with some arbitrary probability (seen in Figure 11).

Let p, q be nodes and, if they are adjacent, they are connected by an edge and are called neighbors. This edge can be thought of as a roadway, for example, or some other type of abstract network connection. For adjacent p, q we define the metric $l(p, q) = |p - q|$.

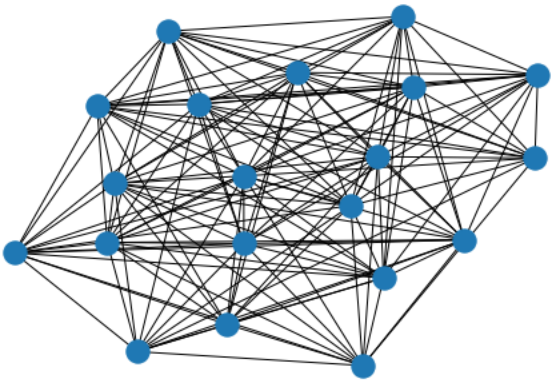

Figure 11: SBM with two clusters of 10 and associated probabilities $p_1 = 1$, $p_2 = 0.5$

We define the weight as:

$$w(p,q) = \begin{cases} \frac{1}{l(p,q)+1} & \text{for } q \in N(p) \\ 0 & \text{otherwise} \end{cases}$$

where $N(p)$ denotes the neighbors of p. Next, we define the degree as:

$$d_p = \sum_{q \in N(p)} w(p,q)$$

with measure defined as follows:

$$M_p(q) = \frac{w(p,q)}{d_p}.$$

We are thus ready to construct the spatial SIR model:

$$\begin{aligned} S_{n+1} &= S_n - \beta \frac{S_n I_n + S_n \Delta_w I_n}{P} \\ I_{n+1} &= (1-\gamma) I_n + \beta \frac{S_n I_n + S_n \Delta_w I_n}{P} \\ R_{n+1} &= R_n + \gamma I_n \end{aligned}$$

where the Laplacian operator, Δ_w, is defined with regards to I_n as:

$$\Delta_w I_n = \sum_{q \in N(p)} M_p(q)|I_n(q) - I_n(p)|.$$

To simulate, we infect a community (or node) with probability $\frac{3}{n}$, where n is the number of communities. After experimentation, we choose an infection rate of 0.01, $\beta = 0.75$, $\gamma = 0.05$ and $t = 1000$ iterations to produce realistic graphs.

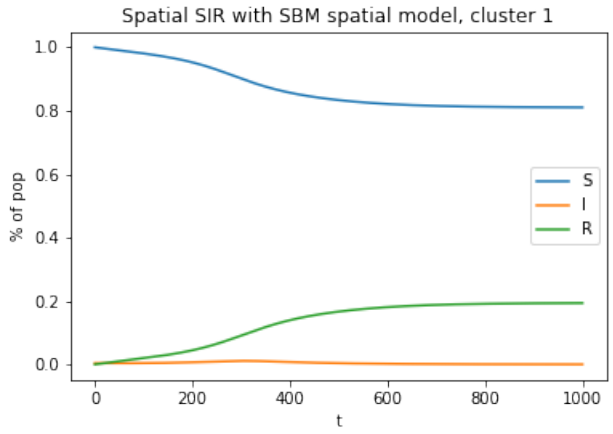

Figure 12: Spatial SIR in Cluster 1

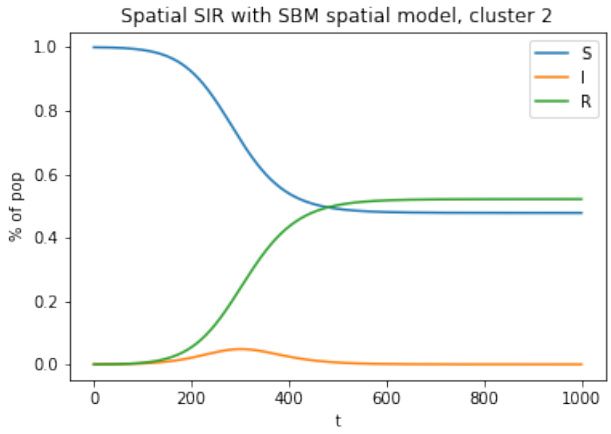

Figure 13: Spatial SIR in Cluster 2

We see in Figure 21 and Figure 22, where Cluster 1 and Cluster 2 comprise 47% and 53% of the population, respectively, that the infection rates of each cluster diverge significantly. Interestingly, 0.2% of Cluster 1's population is infected at $t = 0$, while only 0.14% of Cluster 2's population is infected. Thus, at this early stage of inquiry, we might conclude that population density is a greater determinant of infection than initial infection rate.

Of further interest is the degree to which spatial correlation is present. Indeed, in Figure 14 we note that Cluster 1's infection rate increases only after it is surpassed by Cluster 2's. This relationship is mathematically expressed by the Laplacian term in the above model. However, it is unclear whether in practice this relationship should be characterized by the same coefficient as $S_n I_n$ (ie. β).

In terms of next steps, it would be of interest to consider a real life network — whether that of cities, as presented by Paeng and Lee, or other abstract notions of network, such as one of

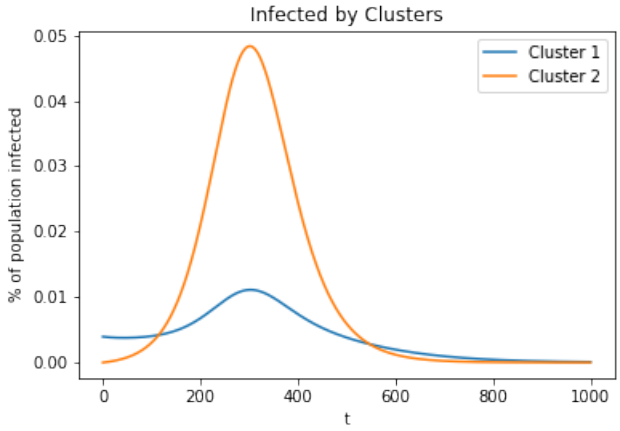

Figure 14: Infected Pop. of Cluster 1 v. Cluster 2

ideological affiliation. Further, it would be useful to consider the degree to which the Laplacian term is independent of β.

3 USA COVID-19 Non-Spatial ODE Fitting

We used Worldometer data for the USA's COVID data points *Worldometers.info* (N.d.). We chose to cover a specific country rather than world wide due to the fact that data reporting may differ vastly across countries. Specifically, we was able to pull Active Cases to calculate I with a fixed N (from the same source, calculated on July, 1, 2020). To calculate S, we compile an cumulative amount of cases from Daily Cases, and subtracted this from N at every point. To calculate R, we simply subtracted S and I from the total population N, and then we divided all values by N to turn these into proportions. See the US SIR data as pulled and calculated in figure 15. Note that it may appear a bit trivial due to the fact that the population N is large and (thankfully) I never grows too large).

$$N = ConstantPopulation = 331,002,651$$
$$S = \frac{N - CumulativeCases}{N}$$
$$I = ActiveCases/N$$
$$R = N - SN - IN$$

I used the lmfit (https://lmfit.github.io/lmfit-py/index.html) tool to fit parameters b, k into our ODE system. This is done through a few steps. First, the ODE function is defined in terms of S, I, R, b, k. Next, we define an ode solver to solve the ODE system with initial conditions S[0], I[0], R[0] and parameters b, k. Then, we define an error residual function that calculates the absolute error between the solved S, I, R curves and the real data curves for a given b, k. Finally, we run a lmfit.minimize operation on the error function with respect to our parameters b,k. Effectively, this is looking for the most optimum b, k combinations such that the least square error between our fitted solution and real solution is minimized Chowell (2017). With an initial guess of b = .15, k = .1 yielded a fitted curve of b = 0.03612338, k = 0.02448164. We tested other initial starting values and received similar results. This alludes towards a COVID-19 R_0, with consideration of data quality and the USA's response over time, as 1.475529417.

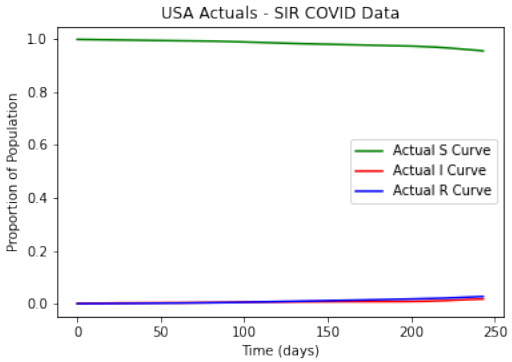

Figure 15: Actual USA Covid Curve. Data from Worldometer

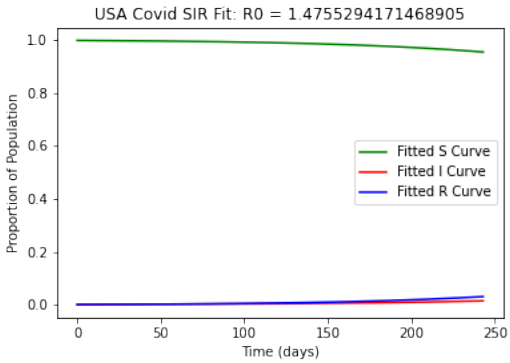

Figure 16: Fitted USA COVID Curve.

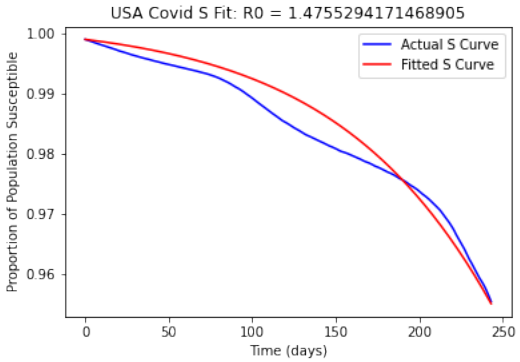

Figure 17: Fitted USA COVID S Curve

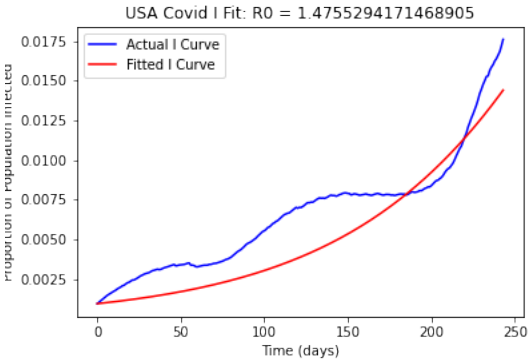

Figure 18: Fitted USA COVID I Curve.

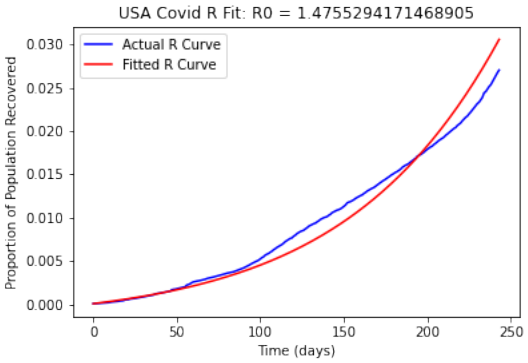

Figure 19: Fitted USA COVID R Curve

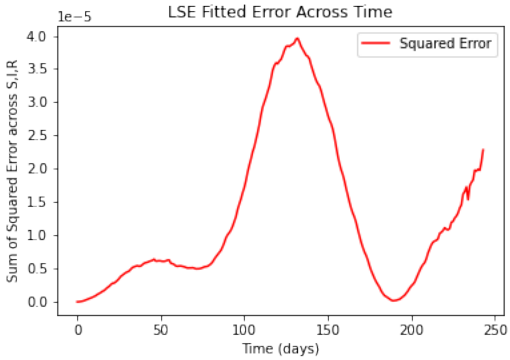

Figure 20: Least Squared Error of fit at Different Times.

4 Conclusion

This report examined the behavior of the SIR model with slight variations and summarized key results. It also explored the performance of the SIR model when fit to COVID-19 data.

The stochastic SIR model pointed to some important findings that have practical applications in controlling the size and duration of epidemics. However, this model may be oversimplifying the complex matrix of social interactions. A serious limitation is that contacts are not defined symmetrically (if the group randomly chosen to meet individual i contains individual j, the group chosen to meet individual j need not contain individual i). It also does not consider the possibility of immigration. Nevertheless, the above findings may have important pharmacological and social implications as they shed light the human and economic costs of epidemics. Future research could focus on analyzing the impact of varying the other parameters of the model through more simulations. It would also be interesting to consider the asymptotic case for large n, or the case where $R = \infty$. Analytical estimation, as opposed to the estimation by simulation presented here, may be a useful direction for further research.

In the discrete default extension, we were able to show that the SIR model built on a 2D-grid is able to express various spatial dynamics of the spread of infectious disease. For example, infection that is randomly distributed on the graph leads to a much higher initial infection rate, while disease spread from a corner outward generates a slower infection rate growth. Further, the degree to which agents move around has a strong positive correlation with the spread of disease.

The spatial cluster extension that considered an SIR model built on a SBM graph developed the notion of spatial interaction in a more sophisticated, and perhaps realistic way by which to model how infectious disease spreads through communities. It was shown that the model was useful in expressing the degree to which disease was spread within given "clusters", as well as the presence of spatial correlation or the spread of disease from one cluster to another. It was discussed that this model could be developed further to describe a real network, such as a grouping of cities. Further, it may be possible to develop a richer model by describing a spatial coefficient.

For the spatial PDE standard extension, we find that the the grid point where the initial infected population starts is the most severely impacted, having the largest amounts of the grid point's population removed after the total time course. The diffusion parameter p has a positive relationship to the rate that an infected population in one gridpoint will spread to another grid point and hence, a higher p relates to a larger spatial region (geography) impacted. A smaller p value alludes towards geographic containment, which yields a larger final R value in the impacted areas than for larger p values.

We found a relatively good model fit of the USA COVID-19 Active Cases curve as reported by *Worldometers.info* (N.d.). One drawback here was the quality of data, and the lack standardization across country. For example, there are different data sets available (Daily Cases, Active Cases, Daily Recoveries, Daily Deaths, etc.) across different countries, which calls for slightly different ways of piecing together all of the S, I, R actual curves. An early (albeit) ambitious goal was to compare results across multiple countries. If we could assume similar data quality and reporting standards across countries, then the fitted parameters b,k are not only indicators of the disease's base infection stats (which should not change across countries), but furthermore are indicative of a given country's response to the virus. However, the lack of data quality and standardization in reporting removed this exploration from the scope of our investigation. Hence, our final b, k are a function of the virus' base stats, the USA's response to the virus, and the quality of data provided.

References

Chowell, Gerardo. 2017. "Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts." *Infectious Disease Modelling* 2(3).

Paeng and Lee. 2017. "Continuous and discrete SIR-models with spatial distributions." *Journal of Mathematical Biology* 74:1709–1727.

Tuckwell, Henry C and Ruth J Williams. 2007. "Some properties of a simple stochastic epidemic model of SIR type." *Mathematical biosciences* 208(1):76–97.

Weiss, Howard Howie. 2013. "The SIR model and the foundations of public health." *Materials matematics* pp. 0001–17.

Worldometers.info. N.d. https://www.worldometers.info/coronavirus/.

5 Appendix

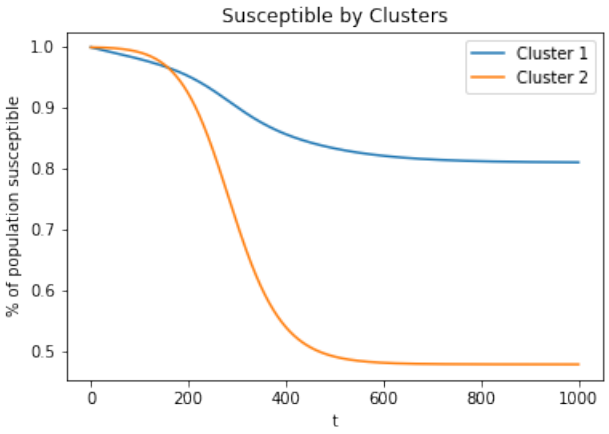

Figure 21: Susc. Pop, Cluster 1 v. Cluster 2

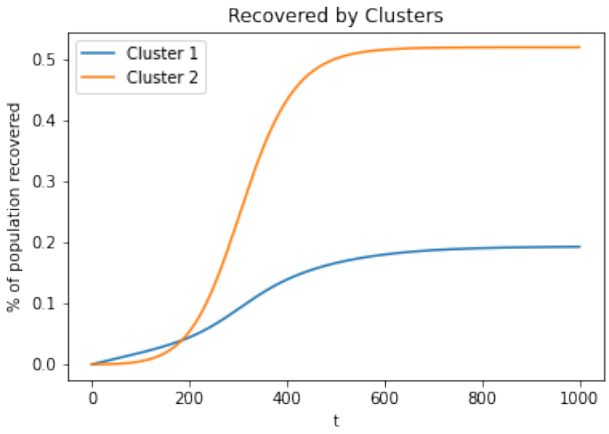

Figure 22: Rec. Pop, Cluster 1 v. Cluster 2

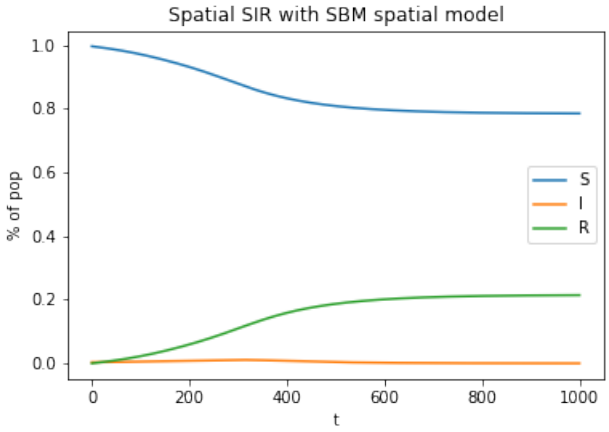

Figure 23: Spatial Cluster SIR Model with SBM