STOR-672 Final Project

Davis Berlind

April 2020

1 Introduction

Starting from the initial outbreak in December 2019, the spread of COVID-19 continues to be the defining feature of 2020 now nearly half-way into the year. With no clear timeline for the development of a vaccine, understanding how this disease will continue to spread remains a critically important question, and one that is well suited for study by simulation. One established approach for studying the spread of infectious disease is the use of compartmental models, which first gained popularity in the 20^{th} century with the foundational work of Kermack and McKendrick [1927]. The key idea in this class of models is to break the population up into discrete compartments, e.g. susceptible, infected, and recovered individuals in the case of epidemiology, and model the movement of individuals between compartments.

One relatively simple example is the Susceptible-Infected-Recovered-Susceptible (SIRS) model, which aims to model pandemics wherein immunity is not permanent, (e.g. the case with the flu, and possibly with COVID-19). Let S, I, and R be the sizes of the susceptible, infected, and recovered populations respectively. Then the total population is given by $N = S + I + R$. Furthermore, let β be the infection rate, i.e. the probability that an infected agent transfers the disease to a susceptible agent in a period, let γ be the rate of recovery, and let ξ be the rate at which agents lose immunity. Then we can characterize the evolution of the infection over time t with the following system of differential equations,

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \xi R \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I - \xi R. \tag{3}$$

We can characterize the steady state of this system by setting each equation to zero, and if we let $p \equiv I/N$

be the disease prevalence, then dividing each equation in the above expression by N gives

$$\frac{dS/N}{dt} = -\frac{S}{N}\beta p + \xi\frac{R}{N} = 0 \tag{4}$$

$$\frac{dp}{dt} = \frac{S}{N}\beta p - \gamma p = 0 \tag{5}$$

$$\frac{dR/N}{dt} = \gamma p - \xi\frac{R}{N} = 0. \tag{6}$$

The last equation implies that in the steady state(s),

$$\frac{R}{N} = p\frac{\gamma}{\xi}. \tag{7}$$

Furthermore, because we have assumed a fixed population $N = S + I + R$, then dividing by N we have

$$\frac{S}{N} = 1 - p - \frac{R}{N} \tag{8}$$

Substituting (7) into (8) and (8) into (5), we get that $\frac{dp}{dt} = 0$ when

$$p = \begin{cases} 0, & \text{if } \frac{\beta}{\gamma} < 1 \\ \max\left\{\frac{\xi(\beta-\gamma)}{\beta(\xi+\gamma)}, 1\right\}, & \text{if } \frac{\beta}{\gamma} > 1. \end{cases} \tag{9}$$

For the second equilibrium to exist, we need $\beta - \gamma > 0$, which is where the basic reproduction number $R_0 = \beta/\gamma$ comes from. So the system admits two steady states,

$$(S^*, I^*, R^*) = \begin{cases} (N, 0, 0) \\ \left(\frac{\gamma}{\beta}, \frac{N-\frac{\gamma}{\beta}}{1+\frac{\gamma}{\xi}}, \frac{N-\frac{\gamma}{\beta}}{1+\frac{\xi}{\gamma}}\right). \end{cases} \tag{10}$$

The disadvantage of the SIRS model is that it assumes homogeneous mixing amongst the agents. In other words, at each time step, an infected agent can interact with any other agent in the population and, given that they are susceptible, infect them with probability β. If we were to represent the system of interactions as a graph, it would be fully connected. This of course does not reflect how individuals actually interact, and presents a real drawback for studying the effects of an intervention like social distancing, which alters how agents interact and has become the primary policy response to COVID-19. My goal with this project is to improve upon the SIRS model by adding structure to the interactions of individuals via random graphs in a continuous time simulation of a pandemic.

2 Model

2.1 Social Structure

In order to impose some level of structure on social interactions, we can first generate a random graph using the Erased Configuration model described in the following subroutine:

Algorithm 1: Erased Configuration Model Subroutine

1 Define population size n.
2 Sample degree sequence $\{d_i\}_{i=1}^n$ i.i.d. from degree distribution D.
3 Let $L_n := \sum_{i=1}^n d_i$ and generate sequence of half-edges $\mathbf{x} = \left\{x_1^{(1)}, x_2^{(1)}, \ldots, x_j^{(i)}, x_{j+1}^{(i)}, \ldots, x_{L_n-1}^{(n)}, x_{L_n}^{(n)}\right\}$.
4 Sample a random permutation $\mathbf{y}$ from $\mathbf{x}$.
5 Pair corresponding half-edges of $\mathbf{x}$ and $\mathbf{y}$ to generate the edges of random graph G.
6 Remove self-loops and duplicate edges from G
7 **return** G.

For simplicity, we will generate $\{d_i\}_{i=1}^n \overset{i.i.d.}{\sim} \text{Poisson}(\nu)$, where ν represents the average number of contacts (edges) each individual (node) has.

2.2 Main Routine

In addition to specifying structure for inhomogeneous interactions, we deviate from the typical SIRS framework by allowing our agents to perish from the infection, and subsequently be removed from G. Therefore, the possible events in the system we are simulating include:

- An infected node recovers and gains immunity.
- A recovered node loses immunity and becomes susceptible.
- A susceptible node is infected by an infected neighbor.
- An infected node dies and is removed from G.

Now, to simulate the spread of a pandemic through a network, we make the following simplifying assumption,

Assumption 1 (Exponential Events) *Assume that the time between each event in this simulation follows an exponential distribution and that events are independent so that,*

A.1.a. Given that a node n_I is infected, the duration of the infection for n_I is an exponential random variable with mean $1/\gamma$ days.

A.1.b. Given that a node n_I is infected, the time at which n_I dies is an exponential random variable with mean $1/\mu$ days.

A.1.c. Given that a node n_S is susceptible and has an infected neighbor, the time until that neighbor infects n_S is an exponential random variable with mean $1/\beta$ days.

A.1.d. Given that a node n_R is immune, the time until n_R loses immunity is an exponential random variable with mean $1/\xi$ days.

The benefit of Assumption 1 should be immediately clear. Suppose that n_I is an infected node and let $R \sim \text{Exp}(\gamma)$ be the time until n_I recovers and $D \sim \text{Exp}(\mu)$ be the time until n_I dies, Then the probability that n_I dies before recovering is given by

$$\begin{aligned}
P(D < R) &= \int_0^\infty P(D < R, R = r)\, dr \\
&= \int_0^\infty P(D < R \mid R = r) f_R(r)\, dr \\
&= \int_0^\infty P(D < r) f_R(r)\, dr \\
&= \int_0^\infty [1 - \exp(-\mu r)]\gamma \exp(-\gamma r)\, dr \\
&= \lim_{r\to\infty} \left[\frac{\gamma}{\gamma+\mu} \exp[-r(\gamma+\mu)] - \exp(-\gamma r)\right]_0^r \\
&= \frac{\mu}{\mu+\gamma}.
\end{aligned}$$

So for modeling COVID-19, we know it takes approximately 25 days on average to fully recover ($\gamma = 1/25$, Ling et al. [2020]), and we know that the mortality rate is approximately 7%, so we can back-out μ as

$$0.07 = \frac{\mu}{\mu + \frac{1}{25}} \implies \mu = 0.003.$$

We can use a similar procedure to back-out β once we specify a probability that a susceptible node is infected by an infected neighbor. For ξ we will simply make an assumption about the duration of immunity.

Suppose now that we just have one node and want to determine the time of the next event $E = \min\{D, R\}$, then another convenient property of assuming exponential events is that $T \sim \text{Exp}(\gamma + \mu)$ and because the type of event E only has support on $\{\text{Recovery}, \text{Death}\} = \left\{\text{Recovery}, \text{Recovery}^C\right\}$, and

$$\begin{gathered}
P(E = \text{Death}) = P(D < R) = \frac{\mu}{\mu+\gamma} \\
\implies \\
P(E = \text{Recovery}) = P\left(E = \text{Death}^C\right) = 1 - \frac{\mu}{\mu+\gamma} = \frac{\gamma}{\mu+\gamma}.
\end{gathered}$$

So E is a Bernoulli random variable with success probability $\mu/(\mu+\gamma)$, and we can simulate the next event by drawing a time T and a type of event E. This argument generalizes to the case of many independent

exponential random variables. If $T_i \sim \text{Exp}(\lambda_i)$ for all $i \in \{1, \ldots, n\}$ and the T_i's are jointly independent, then,

$$T = \min\{T_i, \ldots, T_n\} \sim \text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right),$$

and the index of the event i follows a multinomial distribution where

$$P(i = j) = \frac{\lambda_j}{\sum_{i=1}^{n} \lambda_i}.$$

These properties of the exponential distribution motivate the following routine for simulating pandemic on a network. For notational convenience let $\Lambda(n)$ be the set of neighbors of the node(s) n and let $\Gamma(n_i, n_j)$ be the count of edges between node(s) n_i and n_j.

Algorithm 2: Network Pandemic Simulation

Define end time $T_{\max}$, population size N, initial number of infected nodes $N_{I,0}$, number of simulations S, and γ, μ, β, and ξ.
for $s \leftarrow 0$ **to** S **do**
    Generate graph G.
    Sample $N_{I,0}$ nodes $\{n_1, \ldots, n_{N_{I,0}}\}$ at random from G without replacement.
    $I := \{n_1, \ldots, n_{N_{I,0}}\}$
    $S := \{n \in G \mid n \in \Lambda(I) \cap n \notin I\}$
    $R := \{\}$
    Initialize:
        $t := 0$
        $N_I := N_{I,0}$
        $N_S := \Gamma(S, I)$
        $N_R := 0$
        $N_D := 0$
    while $t < T_{\max}$ **do**
        Simulation Routine.
    end
end

Algorithm 3: Simulation Routine

$\alpha := (\gamma + \mu)N_I + \beta N_S + \xi N_R$.
Sample $T \sim \text{Exp}(\alpha)$.
$t := t + T$.
Sample $E \sim \text{Multinomial}\left(\frac{\gamma N_I}{\alpha}, \frac{\mu N_I}{\alpha}, \frac{\beta N_S}{\alpha}, \frac{\xi N_R}{\alpha}\right)$.
if $E =$ *Recovery* **then**
    $N_I := N_I - 1$.
    $N_R := N_R + 1$.
    Sample n_I uniform at random from I.
    Remove n_I from I and add n_I to R.
    $N_s := N_s - \Gamma(n_I, S)$.
    $S := \{n \in G \mid n \in \Lambda(I) \cap n \notin I\}$.
else if $E =$ *Death* **then**
    $N_I := N_I - 1$.
    $N_D := N_D + 1$.
    Sample n_I uniform at random from I.
    Remove n_I from I.
    Remove n_I from G.
    $N_s := N_s - \Gamma(n_I, S)$.
    $S := \{n \in G \mid n \in \Lambda(I) \cap n \notin I\}$.
else if $E =$ *Infection* **then**
    $N_I := N_I + 1$.
    Sample n_s from S according to $n_s \sim \text{Multinomial}\left(\frac{\Gamma(n_S^1, I)}{\Gamma(S,I)}, \ldots, \frac{\Gamma(n_S^{|S|}, I)}{\Gamma(S,I)}\right)$.
    Add n_s to I.
    $N_s := N_s + \Gamma(n_s, S)$.
    $S := \{n \in G \mid n \in \Lambda(I) \cap n \notin I\}$.
else
    $N_R := N_R - 1$.
    Sample n_R uniform at random from R.
    Remove n_R from R.
    $N_s := N_s + \Gamma(n_R, I)$.
    $S := \{n \in G \mid n \in \Lambda(I) \cap n \notin I\}$.
end

In practice, we do not actually determine S each loop, but use shortcuts for finding the new susceptible nodes. The notation used here is just for convenience.

3 Results

3.1 Parameter Specification

For the purpose of the experiments in this section, the following simulation parameters are used:

1. Population size: 50,000.
2. Initially infected nodes: 50 (0.1% of the population).
3. $\gamma = 1/25$ – Average disease duration of 25 days.[1]
4. $\mu = 0.003$ – Corresponds to an average mortality rate of 7%.
5. $\xi = 1/270$ – Immunity lasts for an average of 270 days.[2]

In addition to the parameters specified above, we experiment with two different values of β corresponding to infection rates of 10% and 50% respectively. We also experiment with three configurations of the Poisson degree distribution with $\lambda \in \{5, 10, 20\}$ corresponding to strong, mild, and no social distancing scenarios.[3]

3.2 Estimation

The tables below displays experimental results from 1,000 simulations with 95% confidence intervals.

	5 Contacts			
	Infection Rate: 10%		Infection Rate: 50%	
P(No Infections on 500$^{\text{th}}$ Day)	0.868	(0.847, 0.889)	0.0	(0.0, 0.0)
Max. # Infected	44.24	(43.51, 44.97)	29,723.33	(29,715.28, 29,731.38)
Time of Max. # Infected	19.39	(15.83, 22.96)	26.66	(26.63, 26.70)
Max. # Immune	111.23	(105.57, 116.89)	38,170.88	(38,165.18, 38,176.57)
Time of Max. # Immune	165.81	(160.01, 171.62)	96.69	(96.54, 96.84)
# Infected on 500$^{\text{th}}$ Day	2.28	(1.72, 2.85)	2, 655.29	(2,648.46, 2,662.12)
# Immune on 500$^{\text{th}}$ Day	55.80	(50.09, 61.51)	30,120.95	(30,110.77, 30,131.12)
# Susceptible on 500$^{\text{th}}$ Day	20.81	(15.54, 26.07)	2,434.82	(2,428.89, 2,440.75)
# Deaths on 500$^{\text{th}}$ Day	14.59	(13.61, 15.57)	6,540.32	(6,535.62, 6,545.01)
Time # Immune > # Infected	23.02	(22.67, 23.37)	38.34	(38.30, 38.37)

Table 1: Experimental Results with n = 1,000 and 5 Contacts

[1] It has been reported that COVID-19 can take 1-14 days from the initial exposure to manifest symptoms, and there is some evidence it takes around 9.5 days on average to stop shedding the virus (i.e. when the patient is no longer infectious). Therefore; 25 days is a good, and perhaps conservative, first approximation of the average disease duration [Ling et al., 2020].

[2] In previous outbreaks similar to COVID-19, e.g. SARS and MERS, immunity was estimated to last from 2-3 years, so 9 months of immunity appears to be a conservative estimate. https://www.npr.org/sections/health-shots/2020/04/13/833412729/how-long-does-it-take-to-recover-from-covid-19-and-how-long-are-you-infectious

[3] The baseline in this case is slightly higher than the average 16.52 contacts estimated by Del Valle et al. [2007].

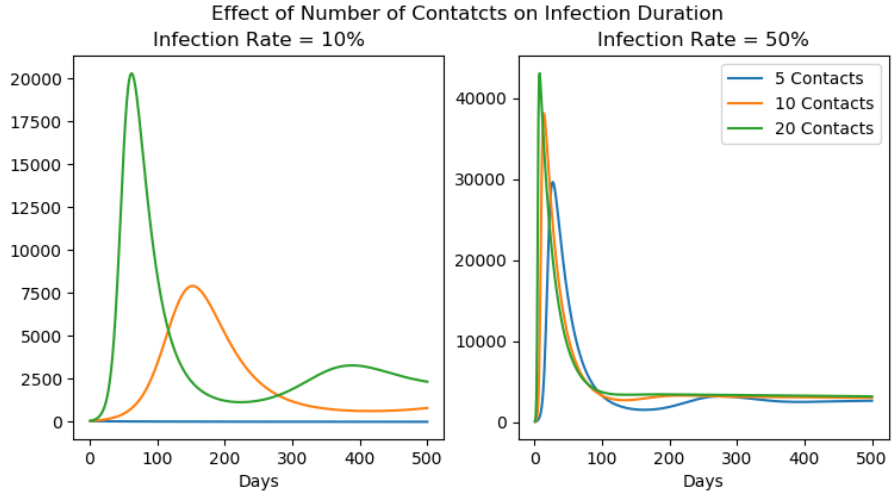
	10 Contacts			
	Infection Rate: 10%		Infection Rate: 50%	
P(No Infections on 500th Day)	0.0	(0.0, 0.0)	0.0	(0.0, 0.0)
Max. # Infected	8,034.99	(8024.29, 8045.70)	38,222.77	(38,216.11, 38,229.44)
Time of Max. # Infected	151.54	(151.15, 151.94)	13.57	(13.55, 13.58)
Max. # Immune	26,507.92	(26,496.27, 26,519.56)	40,047.19	(40,041.91, 40,052.47)
Time of Max. # Immune	246.64	(246.19, 247.08)	91.35	(91.15, 91.56)
# Infected on 500th Day	806.65	(800.07, 813.23)	3,016.92	(3,011.65, 3,022.19)
# Immune on 500th Day	16,026.99	(16,005.88, 16,048.09)	34,908.82	(34,901.58, 34,916.05)
# Susceptible on 500th Day	7,535.98	(7,486.75, 7,585.21)	2,231.91	(2,228.11, 2,235.72)
# Deaths on 500th Day	3,456.32	(3452.29, 3460.34)	7,543.71	(7,538.87, 7,548.54)
Time # Immune > # Infected	103.90	(102.93, 104.87)	27.28	(27.26, 27.29)

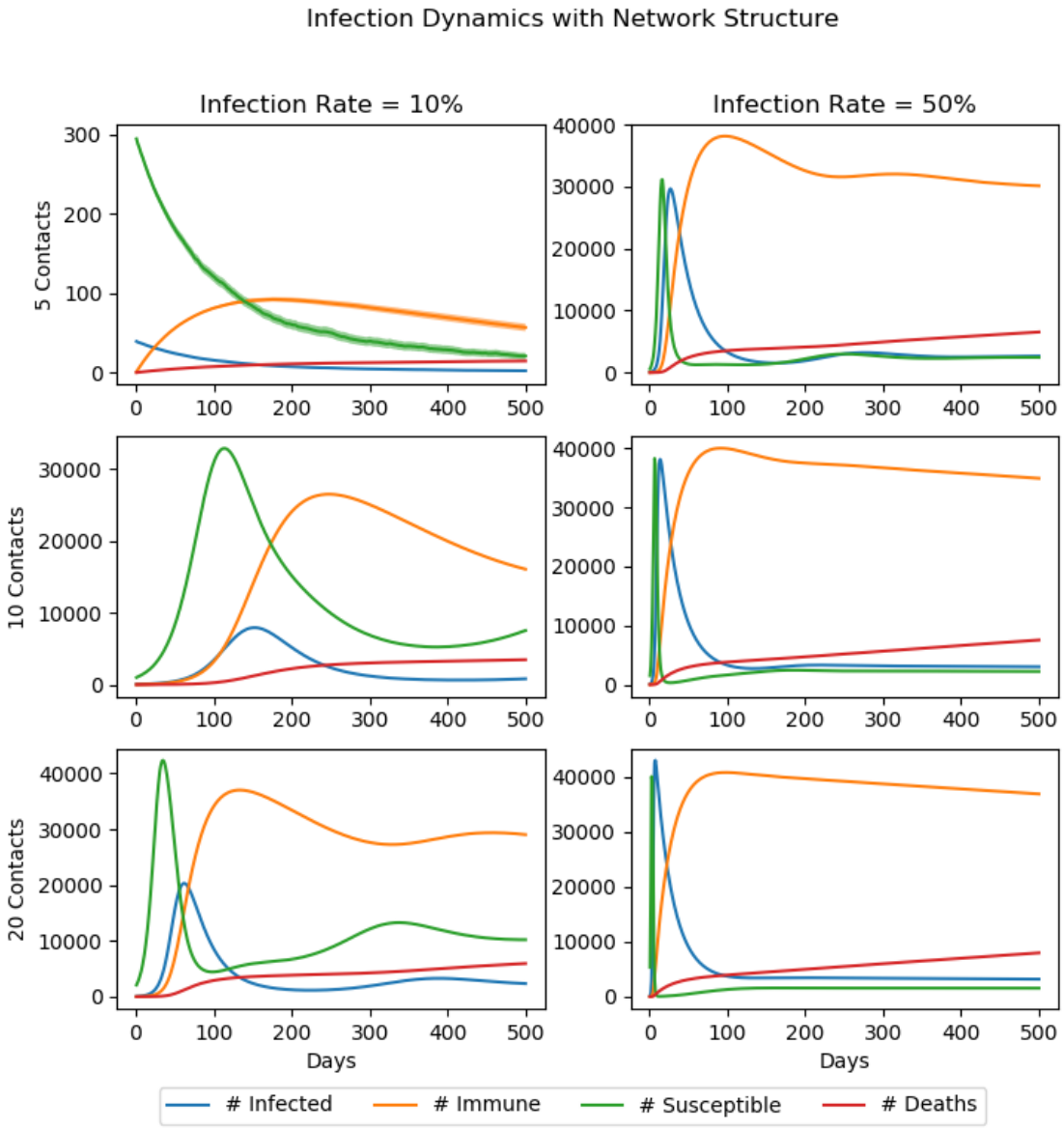
Table 2: Experimental Results with n = 1,000 and 5 Contacts

	20 Contacts			
	Infection Rate: 10%		Infection Rate: 50%	
P(No Infections on 500th Day)	0.0	(0.0, 0.0)	0.0	(0.0, 0.0)
Max. # Infected	20,397.22	(20,387.21, 20,407.22)	43,347.78	(43,342.59, 43,352.97)
Time of Max. # Infected	61.46	(61.36, 61.55)	6.99	(6.98, 7.00)
Max. # Immune	37,038.44	(37,032.39, 37,044.49)	40,876.32	(40,871.37, 40,881.28)
Time of Max. # Immune	132.95	(132.77, 133.12)	98.39	(98.07,98.71)
# Infected on 500th Day	2,335.29	(2,328.78, 2,341.81)	3,194.65	(3,190.39, 3,198.91)
# Immune on 500th Day	29,018.81	(29,006.09, 29,031.53)	36,926.73	(36,920.71, 36,932.76)
# Susceptible on 500th Day	10,202.69	(10,190.29, 10,215.10)	1,542.72	(1,539.93, 1,545.50)
# Deaths on 500th Day	5,942.01	(5,937.47, 5,946.56)	7,979.25	(7,974.12, 7,984.37)
Time # Immune > # Infected	67.19	(67.11, 67.28)	22.20	(22.19, 22.21)

Table 3: Experimental Results with n = 1,000 and 20 Contacts

In order to generate the plots in Figures 3.2 and 3.2, the simulations are discretized by taking the average of each counter over each day (if there is no change the the counter takes the same value as the previous day's average). The plots are the the averages of each daily average across the 1,000 simulations, with the shaded regions representing 95% confidence intervals (in most cases these are too tight to actually visualize).





The most notable feature of these results is that social distancing is very effective at reducing the spread of infection. In Table 1 we see that the model configuration with an average five contacts and a 10% infection rate results in a large positive probability (0.868) that the infection ends in the first 500 days. This is reflected in Figure 3.2, where we see the infection failing to take hold for this parameterization so that the number of infections decreases towards zero from the very start. Figures 3.2 and 3.2 also clearly display the "flattening of the curve" phenomenon that social distancing is meant to accomplish (this is also reflected in Tables 1, 2, and 3, where decreasing the number of contacts pushes out the time that the number of infections reaches its peak and lowers that peak). Despite this, the number of infected agents on the 500^{th} day is decreasing in the average number of contacts, regardless of the infection rate. Looking at the results in Tables 1, 2, and 3 we see that social distancing also has the effect decreasing the number of agents that

have died by the 500^{th} day, regardless of the infection rate.

One surprising result from the 50% infection rate case is that even though decreasing the number of contacts decreases the maximum number infected (29,723, 38,223, and 43,348 in the five, ten, and twenty contact cases respectively), and pushes out the time of the infection peak (27, 14, and 7 days respectively), it has much less of a pronounced effect on the maximum number of agents gaining immunity (38,171, 40,047, and 40,876 respectively).

The other interesting behavior we can see emerging is the occurrence of resurgences, e.g. in the five contact, 50% infection rate and twenty contact, 10% infection rate cases. The fact that there is recurrence in the no social distancing case is especially interesting, as it tells us that even though the pandemic peaks earlier, the drop in infections after the peak is not sharp enough to stop the infection from spreading further, which contradicts the logic of trying to achieve herd immunity as quickly as possible by refusing to adopt social distancing (e.g. the policy pursued by Sweden and the UK with COVID-19).

4 Conclusion

This project has shown how simulating a pandemic with explicit social structure can help us understand the spread of infectious diseases through a population. Understanding the mechanics of how a disease spreads is important for predicting how effective an intervention like social distancing will actually be. The results from this simulation show that in some cases, very aggressive social distancing can keep a disease from spreading at all. To better understand these dynamics going forward, it the next step would be validating the random graph structure chosen against actual social structures. Experimenting with other graph models of representing social interactions would give us insight into how robust the results in this paper are to the specific choice of a configuration model with a Poisson degree distribution.

References

Sara Del Valle, James Hyman, Herbert Hethcote, and SG Eubank. Mixing patterns between age groups in social networks. *Social Networks*, 29:539–554, 10 2007. doi: 10.1016/j.socnet.2007.04.005.

William Ogilvy Kermack and Anderson Gray McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A — Mathematical, Physical and Engineering Sciences*, 772 (115):700–721, August 1 1927. https://doi.org/10.1098/rspa.1927.0118.

Yun Ling, Shui-Bao Xu, Yi-Xiao Lin, Di Tian, Zhao-Qin Zhu, Fa-Hui Dai, Fan Wu, Zhi-Gang Song, Wei Huang, Jun Chen, et al. Persistence and clearance of viral rna in 2019 novel coronavirus disease rehabilitation patients. *Chinese medical journal*, 2020.

.