COMP1002 Assignment

Report

Bharath Sukesh

19982634

May 28, 2020

# Abstract

This report presents a simple method to simulate a population represented as a network using bidirectional graphs, and an SIR-algorithm to model the spread of disease through this population. The population network mimics a population sample and the network exhibits behaviour similar to that of a real community. Simulations were run to investigate the impact of interventions, on the spread of disease. An analysis was made to determine which had more impact on the spread of disease –the virulence of an infection affecting transmission, against the measures taken to protect the population from the spread. I believed that interventions would play a more important role in the spread of disease as measures like social distancing are a powerful ‘community mechanisms’ to reduce the spread of disease, regardless of its virulence. It was shown that the frequency at which people recovered grew as more interventions were put in place, alongside a slight decrease in infection spread. Finally, some potential future work and improvements regarding the simulation algorithm are also provided.

# Background

## The problem at hand

In a population network, the number of infections occurring is heavily dependent on the virulence and interconnectivity of a population. This report focuses on simulating a population network and assessing the spread of a disease through the network with different parameters and interventions set in place. The network consists of a set of people that may have been in contact with one another. To achieve this, I chose to base my network using a graph. The entire reason for modelling disease, is so that predictions can be made for what may happen with similar epidemics in the future. We are able to learn from our past experiences with epidemics and be more prepared to respond to the dangers of disease spread, in a quicker manner. This is done by using a mathematical modelling technique called an ‘SIR’ model.

## Implementation

The simulations take inputs for the rates at which infection is spread (transmission rate); rate at which people can recover from said infection (if possible), and a rate for an infection causing death. The simulation consists of timesteps, with a function called update() being called, to move between timesteps. The network is traversed through using an algorithm similar to that of a breadth-first search traversal but tailored to our simulation. The infection algorithm works by firstly checking whether there exist some people that are infected. If there is not anybody connecting to the patient zero, the population is considered from infection. If there are people in contact with the patient zero, they have a chance of getting infected. This is managed by the sortingHat() method.

Note: If a person adjacent to an infected person, does not get infected, they may still potentially be exposed to the infection in a later timestep via a different adjacent infected person. This is realistic as you can contract a disease from any infected person.

Some parameters that will vary in the creation of a population network include:

* Number of people (Population size)
* Type of connection between people (and the amount of each connection present within a population).

What is tracked throughout the simulation?

* Number of Susceptible (S) people.
* Number of Infected (I) people.
* Number of Recovered (R) people.
* Number of (P) people passed away.

These are the main statistics that we will use as part of our investigation – to gauge which parameters cause the most impact on these statistics.

# Methodology

## Assumptions of the SIR model.

As is with any study, it’s important to acknowledge what assumptions we had made, regardless of how simple or complicated the testing phase is. We must interpret any data from this simulation with these assumptions in mind. For example, we can’t say anything definitive about the population from this model as it doesn’t account for factors that modify the population by means outside infection (For example births will modify the number of susceptible, but we cannot account for this).

1. Nobody can be *added* to the susceptible group. For this model, we are ignoring births, immigration and any other influencing factors that increase the number of susceptible since ‘*day zero*’.
2. The only way for someone to transition out of the susceptible group, is to become infected.
3. The disease can only be spread from person-to-person (contact transmission). This means that all other types of transmission, for example - vector (animals) or environmental transmission (food, water) is assumed not possible.
4. If a group of susceptible people interact with an infected person, all those people (among the group) have an equally random chance of acquiring the infection themselves  
   (Of course, the more exposed you are to infected people, the higher the chance of getting infected as you have more chances of getting infected).
5. The disease is assumed to be acute. Infected people will either recover or die. (Smith et al. 2001)

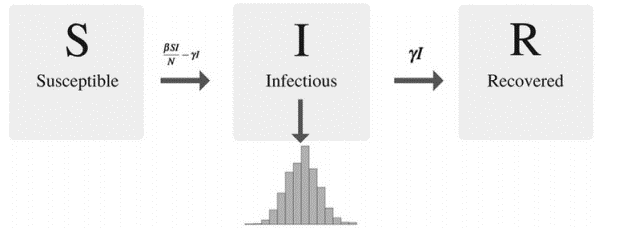


Figure 1.0  
(Daughton et al. 2017)

In an SIR model, individuals move between three categories —S (susceptible), I (infectious) and R (recovered). Movement between the categories is dependent on the rates defined at the beginning of the model. These rates define the virulence of the infection and how long the infectious period lasts. The number of infectious persons at any given time results in the epidemic curve familiar to many epidemiologists

## Investigation

I plan to have two independent tests with the same population size of 30. One with no transmissions selected, and the other with one specific type of intervention selected. I will keep the default transmission rate constant throughout both tests, so the only influencing factor is the interventions causing any change in data. There wouldn’t be much point recording data if it’s full of outliers, so to ensure an adequate level of reliability, I will be repeating these tests many times and comparing the average spread of infection graphically. Due to problems with modelling capabilities of my program, I slightly modified my program to output the population statistics at the end of each timestep, to the terminal so that I can record the results should the program fail to finish.

## Hypothesis

I believe that interventions play a much more substantial role in the spread of disease, than the virulence of the disease itself. The fewer connections you have between people, the less the chance of it spreading and so, I think the stricter the interventions, the safer the population is as a whole. This is to be investigated among the same population, by comparing the spread of infection but with different interventions in use.

## Execution

To run this experiment, I will be using healthSim in simulation mode, to isolate the timestep methods required for these tests. The test network and edge files are included in this submission with the names RandomNames30.csv and RandomEdges30.csv respectively and are required to be specified in the command-line arguments when running in simulation mode. Even if we were not limiting ourselves to simulation mode, two valid input files are necessary for performing the tests in interactive mode, i.e. without the two files, you cannot perform the test as there wouldn’t be a network and/or connections existing between people otherwise.

The command-line arguments format is as following:

|  |
| --- |
| java healthSim -s <network\_file> <edge\_file> <trans\_rate> <recov\_rate> <death\_rate> <int\_code> |

An example for running the tests with a transmission rate of 0.05, recovery rate of 0.001, and death rate of 0.0005 and for no interventions:

|  |
| --- |
| java healthSim -s RandomNames30.csv RandomEdges30.csv 0.05 0.001 0.0005 0 |

To run with interventions, simply change the last argument from a zero, into a 1,2 or 3. Simulation mode states what each intervention number means.

The network and edge files do not need to be in a .csv format, but they do however need to follow the following format (Referencing table from Documentation.docx).

|  |  |
| --- | --- |
| Type of file | Format |
| Network file | <Person name>**,**<Person age> |
| Edge file | <Person #1 name>**,**<Person #2 name> |

These String properties are split by a comma, with no spaces in between properties (Spaces are used only to separate first name and surnames, but not between two properties).

Once you enter valid files and rates, you will be greeted with a prompt instructing you how to timestep, which you should be able to do until satisfied. Upon the conclusion of a test, all relevant test statistics will be appended to a unique text file name saved to the same working directory as the program. This naming of this file will be - healthSimLog\_<date><time>.txt

## Performance

The program has a relatively big memory overhead. This is because of the number of linked lists used in the program, primarily for storing data and for the modelling/timestep process, however, the program in simulation mode does not seem to face any performance issues as it runs most functionality in less than a second, meaning that it is not usually a noticeable problem. In terms of time complexity, due to the time stepping process traversing through entire linked lists, this does slow down the program. This approach gives an expected time complexity of O(#of V), and O(#E + V) for removing and finding edges, and even though adding vertices and edges are usually O(1), since we perform checks to ensure we don’t have duplicate vertices/edges, the time complexity is equivalent to that of a find operation. Despite that, this is not noticeable for most datasets.

# Results

The trials for both tests were repeated four times to ensure reliability. The population size/sample size was kept consistent at 30 people; as was the recovery rate at 0.0001. The default transmission rate is kept at 0.025, with interventions decrementing that rate to their own specific degree. The tests are separated into two groups - one with no transmissions selected, and the other with different interventions selected. The raw data/results have been placed it its own folder in the project directory called “Report Test Logfiles.”

|  |  |
| --- | --- |
| **Key** | |
| Symbol: | Meaning: |
| T | Timestep number |
| S | Number of currently susceptible individuals |
| I | Number of currently infected individuals |
| R | Number of currently recovered individuals |

## No interventions set

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 26 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 |
| I | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 4 | 4 | 4 |
| R | 2 | 3 | 5 | 6 | 7 | 8 | 8 | 8 | 9 | 10 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 25 | 24 | 22 | 20 | 18 | 16 | 14 | 12 |  |  |
| I | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 7 |  |  |
| R | 2 | 3 | 3 | 5 | 7 | 9 | 10 | 11 |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 24 | 22 | 20 | 18 | 16 | 14 |  |  |  |  |
| I | 4 | 4 | 5 | 5 | 5 | 6 |  |  |  |  |
| R | 2 | 4 | 5 | 7 | 9 | 10 |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 26 | 24 | 22 | 20 | 18 | 15 | 13 | 11 |  |  |
| I | 2 | 4 | 5 | 5 | 5 | 6 | 7 | 8 |  |  |
| R | 2 | 2 | 3 | 5 | 7 | 9 | 10 | 11 |  |  |

## Intervention 1 only set (trans\_rate-0.0025);

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 26 | 25 | 24 | 23 | 22 | 21 | 20 |  |  |  |
| I | 2 | 2 | 3 | 3 | 4 | 4 | 5 |  |  |  |
| R | 2 | 3 | 3 | 4 | 4 | 5 | 5 |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 26 | 24 | 23 | 22 | 21 | 19 |  |  |  |  |
| I | 2 | 3 | 3 | 4 | 4 | 6 |  |  |  |  |
| R | 2 | 3 | 4 | 4 | 5 | 5 |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 25 | 24 | 23 | 22 | 21 | 20 |  |  |  |  |
| I | 3 | 3 | 3 | 3 | 4 | 5 |  |  |  |  |
| R | 2 | 3 | 4 | 5 | 5 | 5 |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 26 | 24 | 23 | 22 | 21 | 20 |  |  |  |  |
| I | 2 | 3 | 3 | 4 | 4 | 5 |  |  |  |  |
| R | 2 | 3 | 4 | 4 | 5 | 5 |  |  |  |  |

## Intervention 2 only set (trans\_rate-0.00005);

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 15 |
| I | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 5 | 7 |
| R | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 8 | 8 | 8 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 26 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 16 |  |
| I | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 | 6 |  |
| R | 2 | 4 | 4 | 5 | 6 | 6 | 7 | 8 | 8 |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 25 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 15 |  |
| I | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 7 |  |
| R | 2 | 4 | 5 | 6 | 6 | 7 | 8 | 8 | 8 |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 17 | 15 |  |
| I | 3 | 2 | 3 | 3 | 3 | 4 | 4 | 6 | 8 |  |
| R | 2 | 4 | 4 | 5 | 6 | 6 | 7 | 7 | 7 |  |

## Intervention 3 only set (trans\_rate-0.0005);

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 24 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 |
| I | 4 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 5 |
| R | 2 | 5 | 6 | 7 | 8 | 9 | 9 | 10 | 11 | 11 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 24 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 |
| I | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 5 |
| R | 2 | 5 | 6 | 7 | 8 | 9 | 10 | 10 | 11 | 11 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 24 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 |
| I | 4 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 5 |
| R | 2 | 5 | 7 | 7 | 8 | 9 | 10 | 11 | 11 | 11 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| S | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 |
| I | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |
| R | 4 | 5 | 5 | 6 | 7 | 8 | 8 | 9 | 10 | 11 |

## Average (in 6 timesteps)

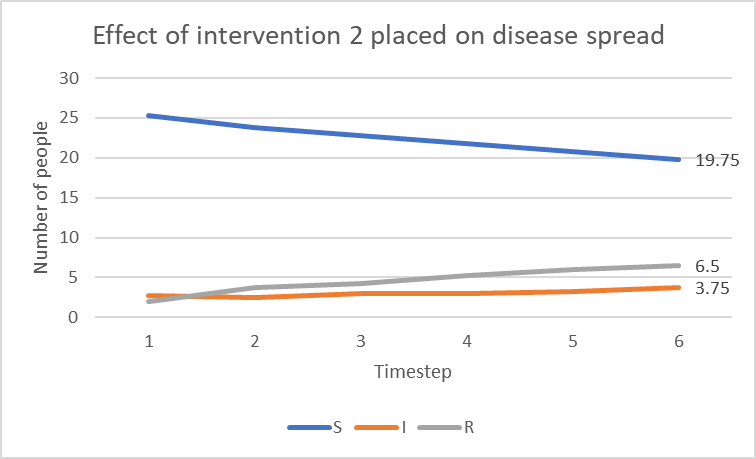
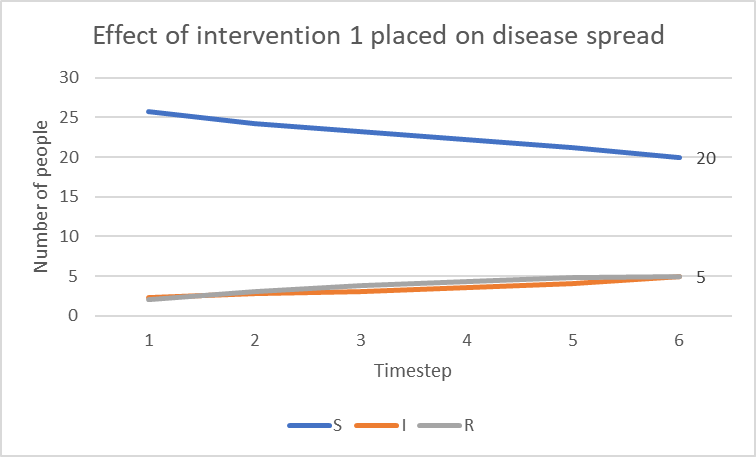
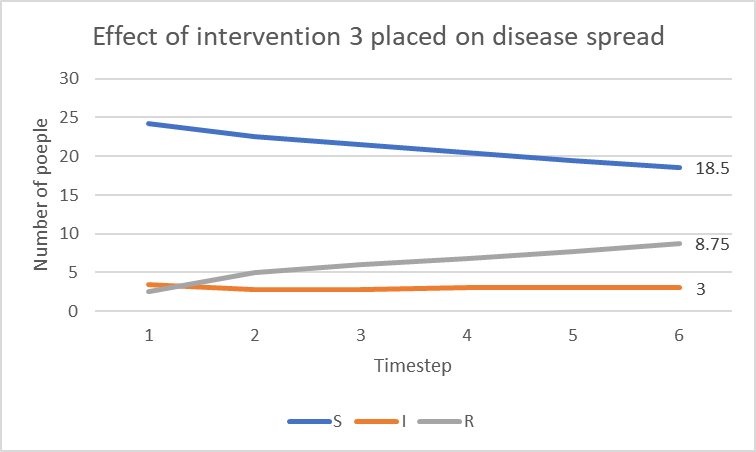
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Intervention 1 only | | | | | | |
| T | 1 | 2 | 3 | 4 | 5 | 6 |
| S | 25.75 | 24.25 | 23.25 | 22.25 | 21.25 | 20 |
| I | 2.25 | 2.75 | 3 | 3.5 | 4 | 5 |
| R | 2 | 3 | 3.75 | 4.25 | 4.75 | 5 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No intervention | | | | | | |
| T | 1 | 2 | 3 | 4 | 5 | 6 |
| S | 25.25 | 23.5 | 21.75 | 20 | 18.25 | 16.25 |
| I | 2.75 | 3.5 | 4.25 | 4.25 | 4.25 | 4.75 |
| R | 2 | 3 | 4 | 5.75 | 7.5 | 9 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Intervention 3 only | | | | | | |
| T | 1 | 2 | 3 | 4 | 5 | 6 |
| S | 24.25 | 22.5 | 21.5 | 20.5 | 19.5 | 18.5 |
| I | 3.5 | 2.75 | 2.75 | 3 | 3 | 3 |
| R | 2.5 | 5 | 6 | 6.75 | 7.75 | 8.75 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Intervention 2 only | | | | | | |
| T | 1 | 2 | 3 | 4 | 5 | 6 |
| S | 25.25 | 23.75 | 22.75 | 21.75 | 20.75 | 19.75 |
| I | 2.75 | 2.5 | 3 | 3 | 3.25 | 3.75 |
| R | 2 | 3.75 | 4.25 | 5.25 | 6 | 6.5 |

The results are evidently not complete. This is due to the limitations of the program used to model. To try gain some sort of reliable results and to make conclusions from the data, I repeated the test four times, and took the average of each type of test. To maintain consistency between the datasets, I only used data up to 6 timesteps as this was the minimum timestep reached among all tests. Despite this, it was not easy to observe a difference between the averages graphically.



Whilst the data does not provide us with anything conclusive, we can still observe an increase in the frequency at which recovered, as well as a slight decrease in transmission (as shown by the end number of infections) when interventions were put in place. This supports my theory that interventions do in fact play a pivotal role in keeping the community safe from the spread of disease.

If I were to get the simulation code working, I would have expected to come to similar conclusions, as the introduction of interventions would reduce the rate of infection each to their own degree (it is dependent on how much it decrements the transmission rate), and with the lack of interventions, there would be a greater peak of infections during the modelling period.

# Conclusion

The data matched up to my hypothesis of interventions playing more of an impact in the spread of disease, than virulence of itself. The less you cut yourself off from each other, the safer you are from contracting the disease. We assume that the sample size is representative of a real population, however, due to the reliability of the data, we cannot say for certain that the parameters we used and tests that were conducted are representative of a real population. This meant that we can’t use this model to make any conclusive statements about the spread of infection.

If we were to look at an alternate SIR model test (Smith et al. 2001), it can be understood that the best way to fight the disease is to reduce interaction among the population regardless of its virulence; which supports my hypothesis.

Future work that could be done to improve this simulation and analysis, could be:

* Using another algorithm to allow for the number of susceptible people to change. This is more realistic as events such as births, and immigration are common, and will alter the population size (unless travel is restricted of course).
* Making the algorithm acknowledge artificial immunity if the disease at hand is a re-emergence of a disease that has occurred in the past. Those that are vaccinated to it, will be considered immune. To build on this, the model could acknowledge the concept of herd immunity and its influence on keeping the population safer.

Most of these proposals for future work involve needing to increase the applicability of the program by increasing the complexity of the model, whether it be conducting better analysis of the current data, using different algorithms to model infection, and implementing more graph statistics to measure the flow of infection better.

# References

Daughton, Ashlynn, Nicholas Generous, Reid Priedhorsky, and Alina Deshpande. 2017. “An Approach to and Web-Based Tool for Infectious Disease Outbreak Intervention Analysis.” Scientific Reports 7 (April): 46076. <https://doi.org/10.1038/srep46076>.

Smith, David, and Lang Moore. 2001. “The SIR Model for Spread of Disease - Introduction | Mathematical Association of America.” The SIR Model for Spread of Disease. Mathematical Association of America. Accessed May 28, 2020. <https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-introduction>.