EE451: SRE
SIRD Model with Network Topologies

Mithilesh Vaidya
Guide: Prof. Sharayu Moharir

Autumn 2020-21

1 Introduction

The SIRD model is an extension of the SIR model [10], which was developed for studying the spread of infectious diseases. It is a simple model yet has been proven to be very effective for studying the outbreak of various diseases. According to this model, each person is categorised into exactly **one of four** classes:

- Susceptible (S): Can get infected if the person is in contact with an infectious person.
- Infectious (I): Infected with the disease. Such a person can transmit the disease to susceptible individuals during contact.
- Recovered (R): Recovered from the disease. For simplicity, we do not consider reinfections.
- Dead (D): RIP

Most literature mainly focuses on SIR since D can be easily introduced as an additional independent transition from I to D without disturbing the rest of the equations.

1.1 Equations

The SIRD model consists of the following set of equations:

$$\begin{aligned} \frac{dS}{dt} &= \frac{-\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - (\gamma + \delta)I \\ \frac{dR}{dt} &= \gamma I \\ \frac{dD}{dt} &= \delta I \end{aligned} \tag{1}$$

- S, I, R and D indicate the number of nodes in each of the categories.
- $N = S + I + R$ refers to the total population, which is assumed to be constant (reasonable if D is not very high).
- β, γ and δ are the transmission rate, recovery rate and death rate respectively. They are constant throughout our analysis.
- Note that various versions of the SIR model can be found in literature which incorporate some more phenomenon e.g. including the natural birth rate and death rate. We stick to the simplest version for ease of analysis.

Observe:

- The equations for S and I are coupled, which complicates the mathematical analysis.
- β is the the probability that an infectious person will transmit the disease to a susceptible person during contact. This parameter depends on the disease e.g. COVID-19 will have a very high β. Assuming random mixing, $\frac{I}{N}$ is the probability of a susceptible person coming in contact with an infected person. Hence, $\frac{\beta SI}{N}$ is the average number of S to I transitions.

1.2 Dynamics

According to the system of equations, the disease will **always** die out i.e. $I^{\infty} = 0$ because:

- We have assumed a closed system i.e. no new nodes are added to our system.
- Recovered cannot get reinfected.

If any of the above assumptions are violated, the disease may persist by remaining in equilibrium with either new nodes which are added to the system or recovered nodes getting reinfected or both.

In literature, we find a threshold called the **Basic Reproduction Number**, denoted by R_0. It can be interpreted as the average number of susceptible people infected by a single individual over time. It is determined by the parameters β, γ and δ. Consider the 2 cases:
$R_0 > 1$: it leads to an **epidemic** i.e. a significant chunk of the population gets infected.
$R_0 \leqslant 1$: the disease **dies out** without affecting a large fraction of the population.

[14] contains a mathematical analysis for deriving R_0 from the basic SIR/SIRD equations. For our model, it can be shown that

$$R_0 = \frac{\beta}{\gamma + \delta}$$

We can get an intuitive feel for this expression by observing that β is the rate at which nodes get infected and $\gamma + \delta$ is the rate at which nodes exit the infected phase (by either recovery or death). If $R_0 > 1$, more nodes are entering the infected phase than nodes which are exiting $\implies$ there will be an explosion in the number of infected nodes. We can similarly reason for the case $R_0 < 1$.

1.3 Simulations

I tried out some basic simulations by integrating the SIRD equations (1) using *scipy.integrate*. The total population size (N) is 100 and we start out with only one infected node.The parameters β, γ and δ are varied and its effects are observed in figure 1.

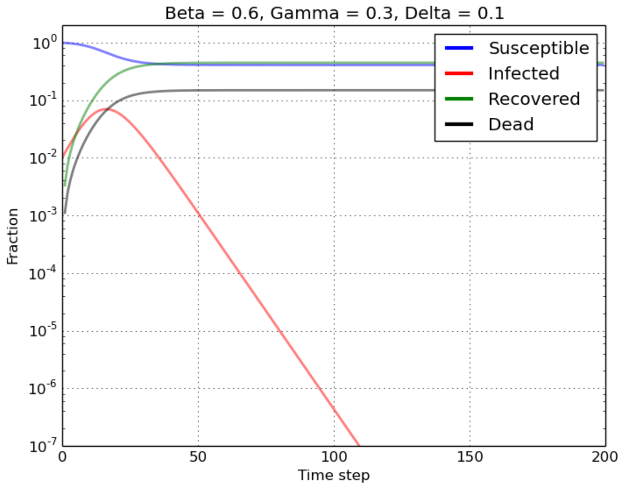
1.4 Estimation of parameters

- Estimating β, γ and δ is not trivial.
- [4] uses machine learning to estimate the parameters for the spread of COVID in Italy.
- A simple SIRD model is set up and the dynamics are computed.
- MSE between predicted and observed data is minimised using grid search and other ML techniques to find the best parameter set.

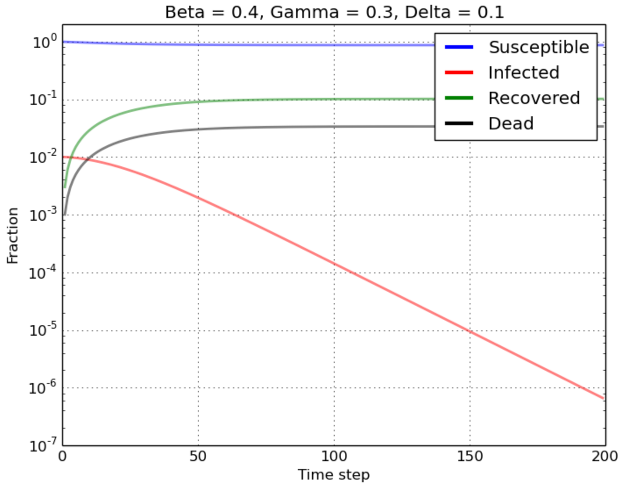
2 Network Models

By assuming $\frac{I}{N}$ as the probability of a S node coming in contact with a I node, we have assumed **random mixing** i.e. any person can interact with any other person. This is not true in real life.
Incorporating a network topology constrains the spread of disease and results in a more realistic modelling of the disease spread.
To incorporate this effect, we assume a graph with the following properties:

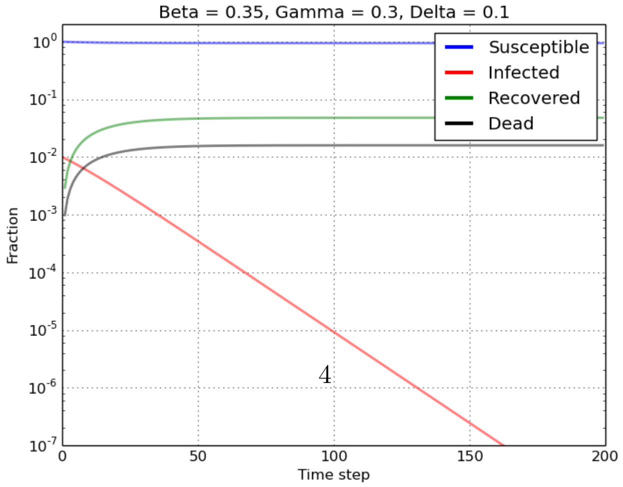
- Each person corresponds to a node/vertex of the graph.
- An edge between 2 nodes indicates that the two people are in contact and hence the disease can spread through this contact.
- Static: Although we come across new people every now and then, it will further complicate our analysis. Hence, we assume that the graph is static.



(a) $R_0 > 1$: We can observe a peak in infections



(b) $R_0 = 1$: Limiting case. Infections dies out



(c) $R_0 < 1$: Infection dies out quickly

Figure 1: Simulation results for various cases of R_0.

- Undirected: the infection can be transmitted in both directions.

Literature is replete with various topologies. Each topology models a particular property of the interactions observed in the real world. A comprehensive survey can be found in [9]. A few properties which need to be considered:

- Degree distribution: Degree refers to the number of neighbours of a node. The distribution of these degrees reflects the variability in how humans interact with each other in society.
- Distance: Average number of nodes between any 2 nodes. Human networks follow the **small world** property i.e. a vast majority of nodes are reachable in very few hops.
- Clustering: How dense the network is. Humans tend to have maximum interaction within a small group of nodes, which gives rise to clusters.

Refer to [11] for more properties of network models.

2.1 Examples of network topologies

An extensive survey of various topologies has been carried out in [9]. Given below are a few select topologies. Images from [9].

2.1.1 Erdos-Renyi [6]

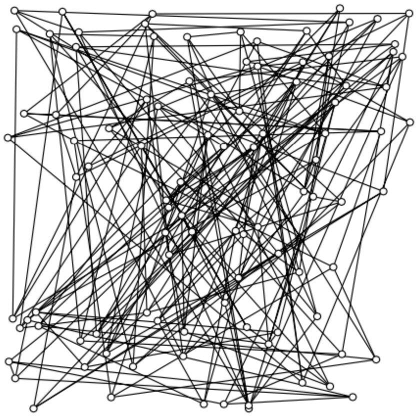


Figure 2: ER graph

- Any 2 nodes are connected with a probability p **independently** of other nodes.
- Degree distribution is binomial and it tends to Poisson for large number of nodes

- Average path between any two nodes is short - **representative** of human contact.
- Low clustering is observed, which is **not representative** of human contact as discussed previously.

2.1.2 Lattice

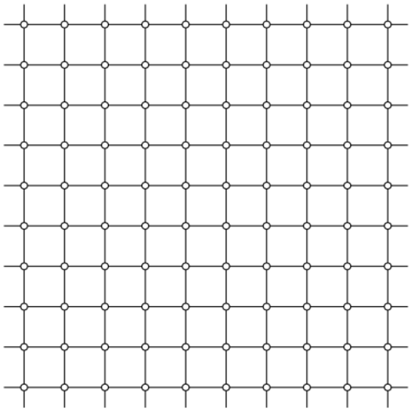


Figure 3: Lattice

- A lattice is a highly localised grid-like network.
- Average path length is high.
- Interactions are more localised than the ER graph and hence **clustering is high**.
- Ample literature on *Percolation Theory* can help in analysis. [3] uses a lattice as the underlying topology for studying SIR evolution.

2.1.3 Barabási–Albert [1]

- Distribution of nodes follows a power law:
  number of nodes with degree $k = N(k) \propto k^{-\gamma}$.
- It can be generated using **Preferential Attachment**:
  Higher the degree of a node $\implies$ Higher the chances of a new node forming an edge with it.
- The Internet can be modelled using such a topology.

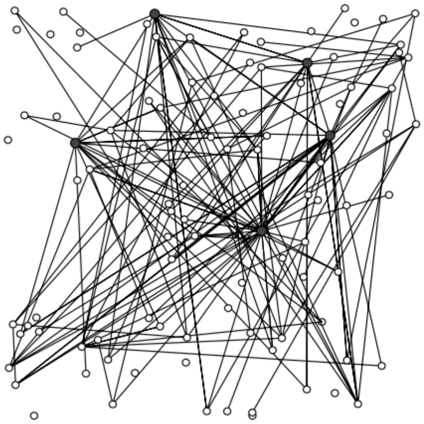


Figure 4: Barabási–Albert

2.1.4 Watts-Strogatz [13]

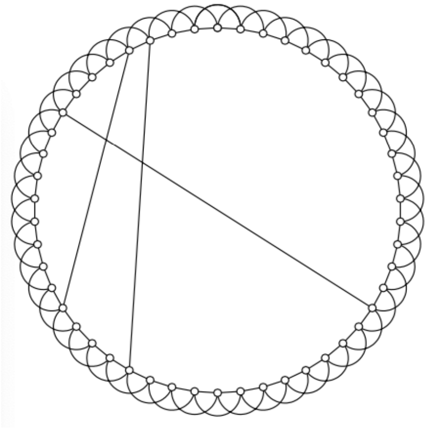


Figure 5: Watts-Strogatz

- A ring lattice gives rise to clustering.
- Random skip connections reduces the average distance between any two nodes.
- Since both the above properties are representative of human contact, I used this model for further analysis.

3 Simulations

3.1 Setup

The aforementioned SIRD equations are no longer valid since we need to also incorporate the network topology. A thorough mathematical analysis is very difficult (sometimes intractable) due to:

1. Coupling of the S and I equations.
2. Incorporation of the underlying topology.

To get a feel for the evolution of the model, I first simulated it in Python. A library called NetworkX [12] was used to to store the graph structure. Results were averaged out over multiple random seeds to reduce variance.
At each time step t, every node changes state according to the following state transition diagram:

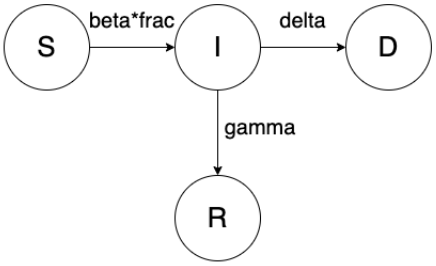


Figure 6: State machine for each node

Here, *frac* denotes the number of infected neighbours of the node and hence the network topology dictates this factor.

3.2 Results

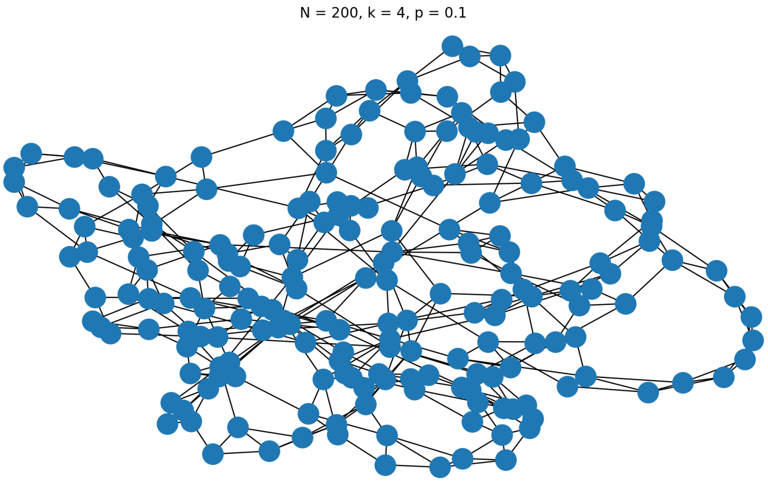

N = 200, k = 4, p = 0.1

Figure 7: The Watts-Strogatz topology which was used for the simulations. N: population, k: number of neighbours, p: probability of rewiring i.e. change a neighbourhood edge to a random edge

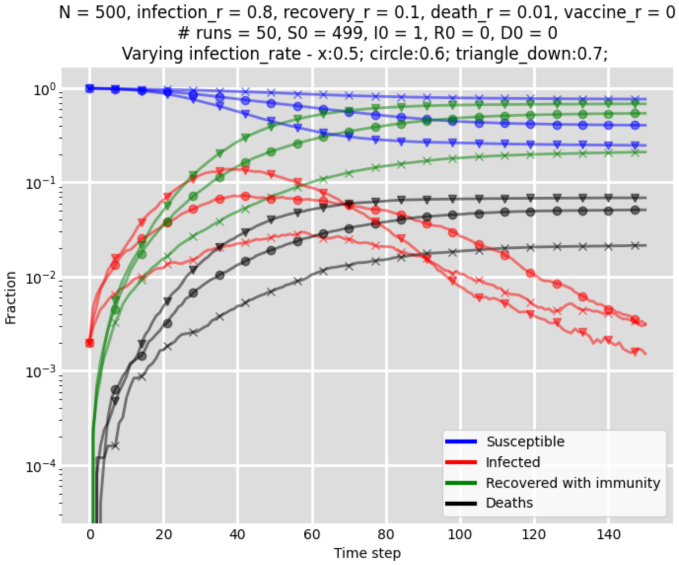

N = 500, infection_r = 0.8, recovery_r = 0.1, death_r = 0.01, vaccine_r = 0
# runs = 50, S0 = 499, I0 = 1, R0 = 0, D0 = 0
Varying infection_rate - x:0.5; circle:0.6; triangle_down:0.7;

Figure 8: The infection rate is varied in this experiment. As expected, higher the infection rate, higher the # infected nodes, higher the deaths

[8] is an extensive study of COVID-19 in the city of Mumbai. They simulated a complicated model with various features such as different death rates for different age groups, different transmission rates in different environments, etc. Policy recommendations for lifting the lockdown are proposed purely based on the simulation results.

4 Towards a Mathematical Analysis

In [5, 7], the network topology is incorporated and thresholds are derived in terms of the adjacency matrix and it's properties.

[2] uses a branching process to study the spread with a *household* model as the underlying network topology. In a household model, each node is a part of a cluster (which is akin to a house, where there is interaction between all nodes). Additionally, there are some random skip connections from one household to the other, which models other social interactions. Methods for calculating the probability of a major outbreak, given few initial infectives, and the expected proportion of the population which is ultimately infected by a major outbreak are discussed.

Note that both analysis deal with the asymptotic behaviour of the system i.e. the total number of infected nodes as $t \to \infty$.

However, the actual dynamics are very crucial since they can help in planning a response e.g. hospitals will be overworked if the peak number of infected nodes exceeds the capacity.

I tried my hand at calculating a closed-form expression for the exact dynamics. I could come up with 2 approaches:

4.1 Markov Chains

Assume a SIR model for simplicity purposes.

1. We can construct a Markov Chain with 3^N nodes where N is the total number of nodes. Each state corresponds to a unique configuration of the world.
   e.g. Let $N = 2$. Node 1 can be either S, I or R and similarly for node 2. The cross product gives us 9 total possible configurations of the world i.e. SS, SI, SR, IS, II, IR, RS, RI, RR. Name these states as $s_0, s_1, ..., s_8$.

2. We can extract the state of each node from the state number s_n using modular arithmetic. The function $S_i(s_n)$ returns the state of node i when the world is in state s_n. Let 0 denote S, 1 denote I and 2 denote R. For N nodes,

$$
\begin{aligned}
S_N(s_n) &= s_n \% 3 \\
S_{N-1}(s_n) &= (s_n/3^1)\% 3 \\
S_{N-2}(s_n) &= (s_n/3^2)\% 3 \\
&\vdots \\
S_1(s_n) &= (s_n/3^{N-1})\% 3
\end{aligned}
$$

Here, / denotes integer division.

3. We can then construct the one-step transition matrix T by using the state of each node, parameters β, γ and δ and the adjacency matrix of the network A. For each transition, consider each node and extract the initial and final state. Then calculate the transition probability using the state machine in figure 6. We can simply take the product of the transition probabilities of each node as the Markov Chain state transition probability.
4. Obtain t-step transition probability matrix T^t.

We now have distribution of world over all states at time t. Hence, we can calculate the expected number of nodes in each category at each time step.

Limitations:

- There is no **straightforward closed form expression**.
- It does not scale with the number of nodes due to the exponential number of nodes in the Markov Chain.

One property which we haven't exploited is symmetry. Instead of modelling each configuration as a different state of the Markov Chain, we can come up with some other formulation which is not so fine-grained. We are not particularly interested in the state of each node at each time step; all we care about is the total number of nodes in each of the 3 (or 4, if we consider SIRD) categories. Hence, any formulation which directly deals with the number of nodes in each time step might make it more scalable. Symmetry refers to the fact that no node is special and in average case analysis, we can collapse many states of the Markov Chain into one single state.

4.2 Linear Algebra

Let S^t be a random vector of dimension N.
$S_i^t = 1$ if node i is susceptible at time t and 0 otherwise.
Similarly, we have I^t, R^t and D^t for time t. Therefore, exactly one of

$S_i^t, I_i^t, R_i^t, D_i^t$ is 1 and the remaining three are 0.

A few definitions:

- A is the adjacency matrix of size NxN if there are N nodes in the graph.
- $\mathbb{B}(\overline{v})$ is a Bernoulli operator which operates on each element $\overline{v}_i$ of a d-dimensional vector $\overline{v}$ in the following way:

  $$\mathbb{B}(\overline{v}) = [Ber(\overline{v}_1), Ber(\overline{v}_2), ..., Ber(\overline{v}_d)]^T$$

  Here, $Ber(a)$ is a Bernoulli random variable with parameter a.
- $\otimes$ denote the element-wise product/Hadamard product.
- Define $K^t = \mathbb{B}\left(\frac{\beta A I^t}{k}\right)$
- If the degree of each node is a constant k (which is the case with the Watts-Strogatz model), the i^{th} element of the vector $\frac{AI^t}{k}$ refers to the fraction of infected nodes in the neighbourhood of node i. This is the crucial term which incorporates the network topology into our system of equations. If degree of each node is not the same, we can calculate it as: $A\mathbb{1}$ where $\mathbb{1}$ is a N-dimensional vector of all ones.

Consider the following system of equations:

$$\begin{aligned}
S^{t+1} &= S^t \otimes \left(1 - K^t\right) \\
I^{t+1} &= I^t + S^t \otimes K^t - \mathbb{B}\left(\gamma I^t + \delta I^t\right) \\
R^{t+1} &= R^t + \mathbb{B}\left(\gamma I^t\right) \\
D^{t+1} &= D^t + \mathbb{B}\left(\delta I^t\right)
\end{aligned}$$

Base cases:
$I^0 = [1, 0, ..., 0]^T$ i.e. without loss of generality, we assume that only the first node is infected.
$S^0 = [0, 1, ..., 1]^T$ i.e. all nodes except the first node are susceptible.
$R^0 = D^0 = [0, 0, ..., 0]^T$ i.e. no node is in the recovered/dead state.

Some properties of the system which make it difficult to obtain a closed-form expression:

1. Matrix and vector products.
2. The element-wise multiplication operator.
3. Recursive formulation.
4. The Bernoulli operator which makes it a non-deterministic system.

Since we only care about the average case analysis, we could somehow use expectation to eliminate the probabilistic $\mathbb{B}(.)$ operator.

5 Conclusion

We discussed the famous SIR model which can be used to model the outbreak of a disease. Then, the need for an underlying network topology was discussed and a few famous models along with their properties were briefly mentioned. Finally, we tried our hand at obtaining the exact dynamics of the system. However, the task of obtaining a **closed-form** expression for the **exact dynamics** in terms of the adjacency matrix for any general graph proved to be too ambitious.
Future work can build on the two approaches by considering various special cases:

- A special graph structure with desired properties which can simplify the equations.
- Instead of the exact dynamics at each time step, one could solve the equations for obtaining only the global optimum e.g. estimating the peak of infected nodes.

Code for all the experiments can be found here.

6 Acknowledgements

I would like to thank Prof. Sharayu Moharir for her guidance.
Special thanks to my classmate Syomantak Choudhary for his inputs.

References

[1] Réka Albert and Albert-László Barabási. "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1 (2002), p. 47.

[2] Frank Ball, David Sirl, and Pieter Trapman. "Threshold behaviour and final outcome of an epidemic on a random network with household structure". In: *Advances in Applied Probability* 41.3 (2009), pp. 765–796.

[3] Michael Boots and Akira Sasaki. "'Small worlds' and the evolution of virulence: infection occurs locally and at a distance". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266.1432 (1999), pp. 1933–1938.

[4] Giuseppe C. Calafiore, Carlo Novara, and Corrado Possieri. *A Modified SIR Model for the COVID-19 Contagion in Italy*. 2020. arXiv: `2003.14391 [physics.soc-ph]`.

[5] Moez Draief, Ayalvadi Ganesh, and Laurent Massoulié. "Thresholds for Virus Spread on Networks". In: *Proceedings of the 1st International Conference on Performance Evaluation Methodolgies and Tools*. valuetools '06. Pisa, Italy: Association for Computing Machinery, 2006, 51–es. ISBN: 1595935045. DOI: 10.1145/1190095.1190160. URL: https://doi.org/10.1145/1190095.1190160.

[6] P ERDdS and A R&wi. "On random graphs I". In: *Publ. math. debrecen* 6.290-297 (1959), p. 18.

[7] A. Ganesh, L. Massoulie, and D. Towsley. "The effect of network topology on the spread of epidemics". In: *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*. Vol. 2. 2005, 1455–1466 vol. 2. DOI: 10.1109/INFCOM.2005.1498374.

[8] Prahladh Harsha et al. *COVID-19 Epidemic in Mumbai: Projections, full economic opening, and containment zones versus contact tracing and testing: An Update*. 2020. arXiv: 2011.02032 [physics.soc-ph].

[9] Matt J Keeling and Ken TD Eames. "Networks and epidemic models". In: *Journal of the Royal Society Interface* 2.4 (2005), pp. 295–307.

[10] William Ogilvy Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115.772 (1927), pp. 700–721.

[11] Lauren Ancel Meyers, Ben Kerr, and Katia Koelle. *Network perspectives on infectious disease dynamics*. 2011.

[12] *NetworkX*. URL: https://networkx.org.

[13] Duncan J Watts and Steven H Strogatz. "Collective dynamics of 'small-world'networks". In: *nature* 393.6684 (1998), pp. 440–442.

[14] Gul Zaman, Yong Han Kang, and Il Hyo Jung. "Stability analysis and optimal vaccination of an SIR epidemic model". In: *Biosystems* 93.3 (2008), pp. 240 –249. ISSN: 0303-2647. DOI: https://doi.org/10.1016/j.biosystems.2008.05.004. URL: http://www.sciencedirect.com/science/article/pii/S0303264708001044.