



ASSIGNMENT 1

Incorporating Stochasticity Into the SIR-Framework Using Event-Driven Modelling and Networks

October 28, 2024

Tutor:
Hourican MSc, C.J.

Group:
A

Student:
Marvin Frommer, Frederieke Loth
15905756, 12016926

Lecturer:
Lees, dr. M.H.

Course:
Introduction to Computational
Science

1 Introduction

Modelling the spread of infectious diseases has the potential to provide us with valuable insights into how a disease spreads within a population. We can model this process under specific, controlled circumstances. These insights may provide a great advantage to help prevent any harmful outcome that an epidemic might bring. By analyzing computational models of disease-spread and running simulations under different experimental frames, we can investigate various scenarios in order to make informed decisions. This could mean investigating different parameters of the model, considering varying social structures or developing an effective vaccination strategy. Both of these approaches will be discussed in this report. The basic framework that underlies the simulations that we conducted is the SIR-framework, developed by William Ogilvy Kermack and Anderson Gray McKendrick[7]. The most basic form of this framework splits up any given population into three sub-components: Susceptible, infected and recovered individuals. The changes of each of these subpopulations over time are given by a set of three coupled ordinary differential equations. These are governed by three disease parameters: The infection rate β , the recovery rate γ and the ratio of these two parameters, defined as the basic reproduction number R_0 , given by

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

This value is crucial since it tells us the number of secondary infections a single infected person would cause when introduced in a fully susceptible population. In other words, it will tell us whether a disease will spread and potentially trigger an epidemic or not. Theoretically, a disease will spread if $R_0 > 1$. In some cases, we used a slightly extended version of the SIR-framework, adding a fourth parameter μ to the model. This value defines the death and birth rate within a population. By adding a natural birth and death rate to a population, oscillations in each population may occur as newly susceptible individuals are introduced while death occurs at every population stage. The natural birth and death rate is typically defined by a single parameter

therefore keeping the total population fixed. Adding this parameter changes our basic reproduction number to[7]

$$R_0 = \frac{\beta}{\gamma + \mu}.$$

Aside from differential equations, which are deterministic, the SIR-framework can also be used to model disease-spread while taking noise into consideration. Noise can be described as random occurrences that may deviate from an expected value. Within SIR Models, this could lead to a sudden increase in the infected population that is larger than the expected increase. We discuss two approaches to introduce noise into the model. First, we used a discrete-event method called Gillespie's First Reaction[5]. Following this, we employed a network-approach to simulate the spread of disease.

Networks are comprised primarily of nodes and edges[1]. When simulating a SIR model, nodes represent individuals and edges the connections between nodes. We can view this as interactions between individuals. Only when an interaction (edge) between these individuals (node) is present, can one individual infect another. One node can thus only infect another directly if they have an edge between them. The disease-spread on networks is simulated through three different network types, with varying topologies.

An Erdös-Renyi (ER) network is recognized by its random approach to generating edges between nodes. All edges occur with an equal probability (p) that is independent from the presence of other edges[3]. Another network we will explore is the Barabási-Albert (BA) network. The edges in this network are not added in a uniformly random manner. Instead, when a new node is added to the network, it has a certain probability to connect with an existing node. This probability is determined by the amount of edges the existing node already possesses. This is referred to as preferential attachment[1]. Finally, we will model SIR in a Watts-Strogatz (WS) network. In this type of network, nodes are initially arranged in a circle (ring topology), where each node is connected to its nearest neighbors. After the ring topology is set up, some of the edges are rewired randomly[10].

Overall, the SIR dynamics in our research are governed by the parameters as discussed above but may still result in different dynamics when looking at the spread of infections. This proposes the interest of our research as we explore how different models, methods and parameters affect the spread of infections within simulations. Understanding these models through computational methods could help formulate effective strategies to help prevent, understand and visualize the spread of infectious diseases.

2 Methods

2.1 Gillespie's First Reaction Method

For this report, the algorithm used to stochastically simulate the spread of an infectious disease was Gillespie's First Reaction (GSP). This discrete-event method calculates the time to the next event (infection, recovery, birth and death) based on predefined rates as well as a random number generator. More precisely, the general scheme of GSP goes as follows:

1. Define all possible events:
 - Identify all events that can occur in the system.
2. Define rates at which events occur:
 - Assign a rate to each event, which quantifies how likely each event is to happen based on the current state of the system.
3. Calculate the time until each event occurs:
 - For each event, compute the time until it occurs based on its rate. This is typically derived from the exponential distribution, reflecting the nature of a Poisson process.

4. Select the lowest event-time:

- Determine the event with the smallest calculated time until occurrence, indicating it is the next event expected to happen.

5. Execute the event:

- Carry out the selected event, which may involve updating the state of the system (e.g., changing the concentration of reactants/products).

6. Return to step 3:

- Repeat steps 3 through 5 at each iteration of the algorithm until a stopping criterion is met (e.g., a specified time has elapsed or all events have occurred)[5].

2.2 Stochastic and Deterministic SIR

Based on the algorithm described in section 2.1, we simulated the development of an infectious disease 100 times. From this, we calculated the average state of S , I and R for each time step. Consequently, we compared this result with the output of a deterministic demographic SIR model as described by the set of coupled ordinary differential equations

$$\frac{dS}{dt} = \mu - \beta SI - \mu S$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I$$

$$\frac{dR}{dt} = \gamma I - \mu R.$$

These equations were solved using the Runge-Kutta method of order 5(4) as follows:
Given the initial-value problem

$$y' = f(t, y), \quad y(t_0) = y_0,$$

- **Initialize:**

$$t_n = t_0 + nh, \quad n = 0, 1, 2, \dots, N$$

where h is the step size.

- **Compute intermediate values:**

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f\left(t_n + \frac{h}{4}, y_n + \frac{h}{4}k_1\right), \\ k_3 &= f\left(t_n + \frac{h}{4}, y_n + \frac{h}{4}k_2\right), \\ k_4 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_3\right), \\ k_5 &= f\left(t_n + \frac{3h}{4}, y_n + \frac{3h}{4}k_4\right), \\ k_6 &= f(t_n + h, y_n + hk_5). \end{aligned}$$

- **Update the value of y :**

$$y_{n+1} = y_n + \frac{h}{6} (k_1 + 4k_4 + k_5)$$

- **Error estimate** (optional):

$$R_n = \frac{h}{6} (-k_2 + 2k_3 + 2k_4 - k_5)$$

Where:

- t_n is the current time step,
- y_n is the current value of the solution,
- h is the step size,
- $k_1, k_2, k_3, k_4, k_5, k_6$ are intermediate slopes (derivatives) evaluated at specific points,
- R_n is the error estimate (optional) to measure the accuracy of the solution [2].

2.3 Computing the Mean for the Stochastic Simulation

To properly visualize the effects of stochasticity in the SIR model, we run the stochastic SIR model simulation 100 times and calculate the mean of each population for every time step. Since the time steps of the GSP-method are rarely equal, we had to compute fixed steps at which to determine the population sizes. All parameters across runs were set to be of the same value to properly compare them. We plotted both the mean and every run of the stochastic model to indicate the difference between each run. The deterministic model was plotted once.

2.4 Variance of Peak Infected Given R_0

Using GSP, we ran multiple stochastic simulations while varying the basic reproductive number R_0 . 50 different R_0 values were chosen and the simulation was run 10 times per value. The goal was to examine the impact of the disease parameter on the between-run variability of the peak number of infected. As a measure of variability, we chose variance, given by

$$\text{Var}(X) = \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

Where:

- $\text{Var}(X)$ or σ^2 : The variance of the random variable X .
- n : The number of observations in the dataset.
- x_i : The i -th observation in the dataset.
- μ : The mean (average) of the observations, calculated as $\mu = \frac{1}{n} \sum_{i=1}^n x_i$.

2.5 Negative Covariance Between S and I

Akin to what was done in section 2.3, we again ran 10 stochastic simulations across 50 different R_0 values. We now aimed to examine the impact of the reproductive number on the covariance between the susceptible and infected population. The measure of covariance is given by

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

where:

- X and Y are the two random variables.
- X_i represents each observed value of the variable X .
- Y_i represents each observed value of the variable Y .

- \bar{X} is the mean of the variable X , calculated as $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.
- \bar{Y} is the mean of the variable Y , calculated as $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.
- n is the total number of observations for each variable.[2]

2.6 Stochastic Resonance and Increased Transients

Stochastic Resonance occurs when oscillatory signals are enhanced due to the introduction of noise[4]. Adding these levels of noise may increase the expression of otherwise weak fluctuations or signals. To visualize the trends we plotted the deterministic SIR model against the GSP model for the same parameters. We would expect, where an equilibrium in the model was reached, to see a fluctuating wave that enhances the otherwise difficult to notice resonance induced by the natural birth and death rates.

Transients are temporary changes away from the equilibrium due to sudden changes in the parameters[4]. Transients may be understood by looking at GSP models where we see that a sequence of random events results in infections or recoveries that move away significantly from the equilibrium. You could view these sequences as temporary changes in the parameters.

2.7 Extinctions as a Function of R_0 and N

The impact of R_0 on the system's dynamics was investigated once more. This time, we ran the stochastic simulation 50 times for each of the 20 R_0 values. The goal was to see whether varying the disease parameter has an impact on the amount of stochastic extinctions that occur in the system. Stochastic extinction refers to the phenomena where a disease dies out before triggering an epidemic, even when $R_0 > 1$ [8]. The same was done using varying population sizes N to see their impact on the frequency of extinctions. The amount of different parameter values as well as runs per value remained the same as before. After investigating the impact of R_0 and N individually, we also looked at their interaction as a predictor for extinctions.

2.8 SIR on Networks

Aside from using the deterministic model and GSP, we implemented SIR on a third type of model, namely networks. Mathematically known as graphs, these objects consist of numerous nodes that are connected to one another by a number of lines (edges)[1]. The connection patterns of networks depend on the type of network at hand. For this report, nodes in a network represented an individual in the population. Any given node could be in one of three states: susceptible, infected or recovered (also described as removed). The edges in a network tell us whether two individuals are in contact with each other. I.e. if two nodes are connected by an edge, we know that the two individuals represented by the nodes are in contact. For this report, we followed the basic, non-demographic SIR framework to simulate disease-spread. This implies that for each iteration, infected nodes had a probability of γ to change their state from infected to recovered. Moreover, every susceptible node connected to an infected node had the probability of β to transition to the infected state in the next iteration. We can assume that removed nodes remain removed for the duration of the simulation.

2.9 Generating Networks and Examining Network Statistics

Before simulating disease-spread on networks, we first looked at how different ways of generating a network can have an impact on its connectivity. Depending on what network type one is generating, there are different hyper-parameters to take into account. For each network type, we examined how varying one of said hyper-parameters can have an impact on its connectivity. To be more exact, the measurements of connectivity we used were:

- Average minimum and maximum degree (number of edges per node).

- Average clustering coefficient. For a single graph, the clustering coefficient defines how interconnected the neighbors of a given node are. It is determined by

$$C = \frac{1}{N} \sum_{i=1}^N C_i,$$

where:

- N is the total number of nodes in the graph,
- C_i is the clustering coefficient of node i , defined as:

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

where:

- * E_i is the number of edges between the k_i neighbors of node i ,
- * k_i is the degree of node i (the number of connections it has[1]).

2.10 Simulating Disease Spread on Networks

Each of the different network types is generated using different hyperparameters. Based on these hyperparameters For each network type one of said parameters as well as the basic reproduction number R_0 were varied in order to examine the effect on the maximum number of infected individuals.

2.11 Erdös-Renyi Network

One of the three networks we simulated disease-spread on is the Erdös-Renyi (ER) Network, also known as a random graph model. The structure of this network arises from the way it is generated, namely by adhering to the following scheme:

1. Define the total number of nodes N , representing the entities in the network.
2. Choose a connection probability p , where $0 \leq p \leq 1$. This value determines the likelihood that any two nodes will be connected.
3. For each pair of nodes ($Node_i, Node_j$):
 - (a) Generate a probability p_{ij} using a random number generator (RNG).
 - (b) If $p_{ij} \leq p$, create an edge between $Node_i$ and $Node_j$ [3].

Multiple networks of this kind were generated using varying values of p . Consequently, the average minimum and maximum degree as well as the average clustering coefficient were calculated for each given value of p . As a subsequent step, disease-spread was simulated for varying values of p as well as R_0 . Our goal was to examine the impact of different parameter values on the peak of the epidemic, i.e. the maximum number of infected.

2.12 Barabási-Albert Network

Another network we used is called Bárabasi-Albert (BA) network. Compared to the random graph model described in section 2.7, instead of generating edges randomly in a uniform manner, edges are preferentially attached. It is implied that in the process of generating a BA-network, new nodes are added iteratively as opposed to starting the generation process with all nodes already present at iteration 0. The general scheme to generate a BA-network is as follows:

1. Start with a small, fully connected network of m_0 nodes, where each node is connected to every other node.

2. For each subsequent iteration, add a new node (i) with an initial degree of 0, meaning it starts with no edges.
3. Define the number m of existing nodes a new node i should be connected to.
4. For each new node i , the probability p_{ij} of it being connected to each of m existing nodes j is given by:

$$p_{ij} = \frac{k_j}{\sum_\ell k_\ell}$$

where:

- k_j is the degree of the existing node j ,
 - $\sum_\ell k_\ell$ is the sum of the degrees of all nodes currently in the network.
5. Continue adding nodes, each time connecting the new node to m existing nodes according to the probability p_{ij} . Update the degrees and recalculate probabilities after each addition until the network reaches the desired size[1].

The network resulting from this method will have many nodes with a low degree and few nodes with a high degree. This degree distribution is common for networks that follow preferential attachment and add nodes iteratively, they are also referred to as scale-free networks[1]. The same network statistics as before were calculated. This time, we varied m , i.e. the maximum number of existing nodes that a newly added one can connect to. Again, disease-spread was simulated on this network according to the basic SIR-framework. Alongside R_0 , we varied the parameter m once more.

2.13 Watts-Strogatz Network

Lastly, we investigated the Watts-Strogatz (WS) network where edges are rewired randomly after setting up a ring topology where nodes are connected to their nearest neighbors. A Watts-Strogatz graph is generated in the following manner:

1. Generate a graph with a ring topology where each node is connected to its k nearest neighbors. Note that k is chosen to be even to ensure symmetry in the graph.
2. Define a rewiring probability p_r .
3. For each node in the network:
 - (a) Generate a random probability p .
 - (b) If $p < p_r$, randomly change one end of the edge to connect to a different, randomly chosen node[10].

Given the fact that many nodes are connected to their neighbors, which are also interconnected, this method results in a graph that forms multiple local clusters that contain many interconnected nodes. Hence, these networks are often referred to as small-world networks[10]. When examining the network statistics as well as the development of the infected population, we varied p_r , i.e. the probability of a node randomly changing one of its connections. Note that k , the number of neighbors any node is connected to, was held constant. The average maximum infected, R_0 , was also varied once more.

2.14 Dynamic Vaccination Strategy

Given a pre-defined network structure, we designed a vaccination strategy with the goal to significantly reduce the peak of the infection within the network. In this scenario, we followed the basic SIR-framework, meaning demography was excluded and recovered remain immune indefinitely. Given a total of 200 tests, as well as a specific budget of maximum tests and vaccinations to be administered at each time step. It should be noted that for this report, we assumed the

the network structure was known from the beginning. However, the disease status of individual nodes was unknown and could only be determined through testing. Individuals who were already deemed infected or recovered by previous testing were excluded from future testing. Moreover, only susceptible individuals could be vaccinated. Different vaccination strategies were deployed and one vaccination strategy was chosen to be compared to a null strategy. Varying test accuracies [0.25, 0.75, 1.0] were considered as well as daily vaccination budgets [1, 3, 5, 10]. It should be noted that some vaccinations may be administered to individuals who were wrongfully determined as susceptible due to test inaccuracies or have moved to a different status since testing. These individuals will not be moved to the 'recovered' population but will remain with their current status.

2.15 Null Strategy

In order to generate some baseline results, the first strategy was to randomly vaccinate individuals that have been deemed susceptible through testing beforehand.

2.16 Hub Vaccination Strategy

As a more systematic method of vaccinating the population, we first sorted all nodes by their number of edges in descending order. This way, we had the biggest hubs of the network available. Resulting from this, we were able to iteratively test and vaccinate them, greatly reducing the spreading potential of the largest hubs in the network.

2.17 Score Vaccination Strategy

To build upon the hub strategy, we employed numerous ways to preference nodes based on their position and connections. Apart from assigning scores to nodes based on their degrees (also known as adjacency[6]), we also assigned scores to nodes based on their closeness. Closeness indicates the path length that nodes have to other nodes[6]. Additionally, we assigned scores based on the betweenness of nodes which indicates how often a node is situated on the shortest path between other nodes[6]. These scores are scaled and added together to determine the nodes with the highest scores. Additionally, during the simulation, nodes that were determined to be near an infected node were given a score boost.

3 Results

3.1 Stochastic and Deterministic SIR

When comparing the deterministic and stochastic model from we notice an overall similar trend for the relationship between S, I and R. However, we may also note that the stochastic model includes many fluctuations around the path with temporary increases and decreases. Although the overall trend of both graphs is relatively similar, we can deduce that the stochastic simulation is less predictable.

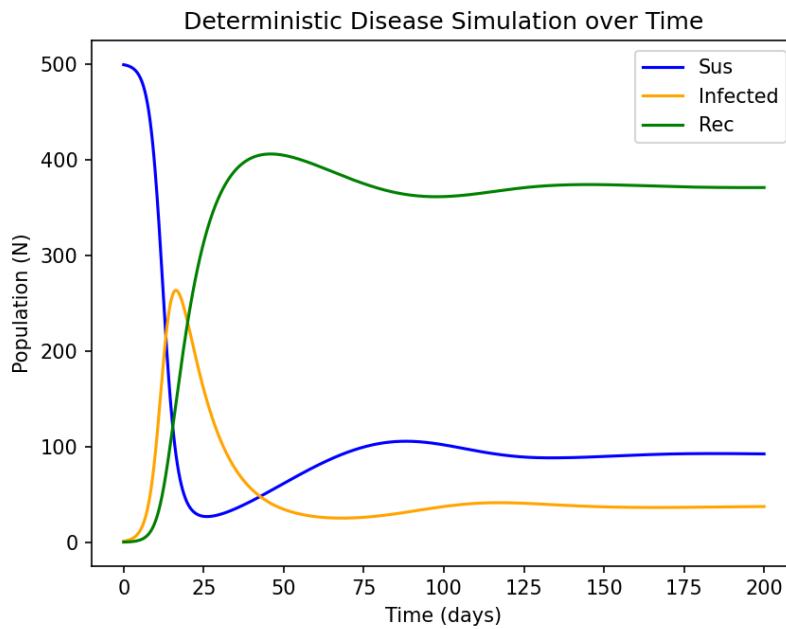


Figure 1: A single run of the deterministic SIR model simulation. Beta, gamma and mu parameters were set as 0.6, 0.1 and 0.01 respectively. The total population size was 500. A single infected individual was introduced into a fully susceptible population at t_0 .

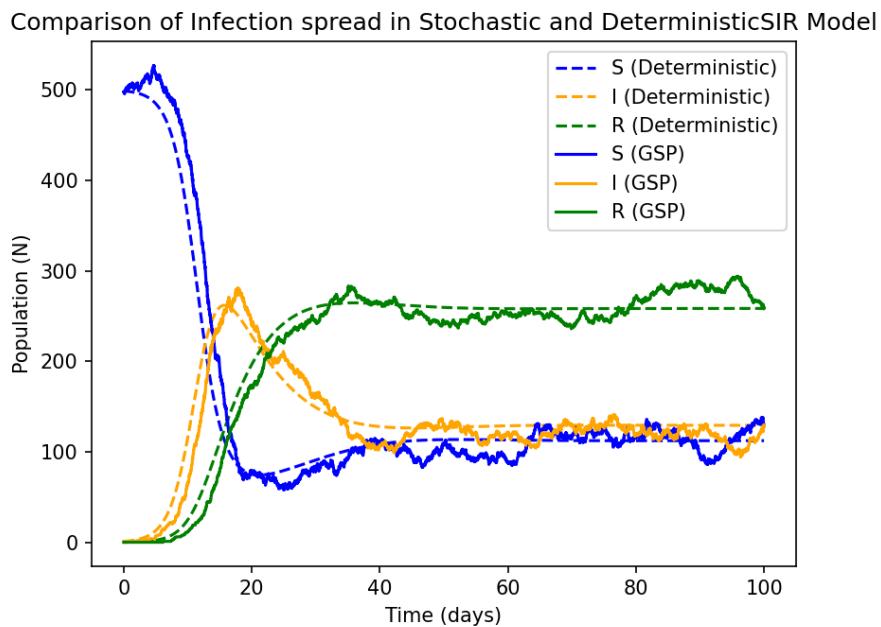


Figure 2: A single run of the stochastic SIR model simulation compared to a deterministic simulation. Beta, gamma and mu parameters were set as 0.6, 0.1 and 0.05 respectively. The total population size was 500. A single infected individual was introduced into a fully susceptible population at t_0 .

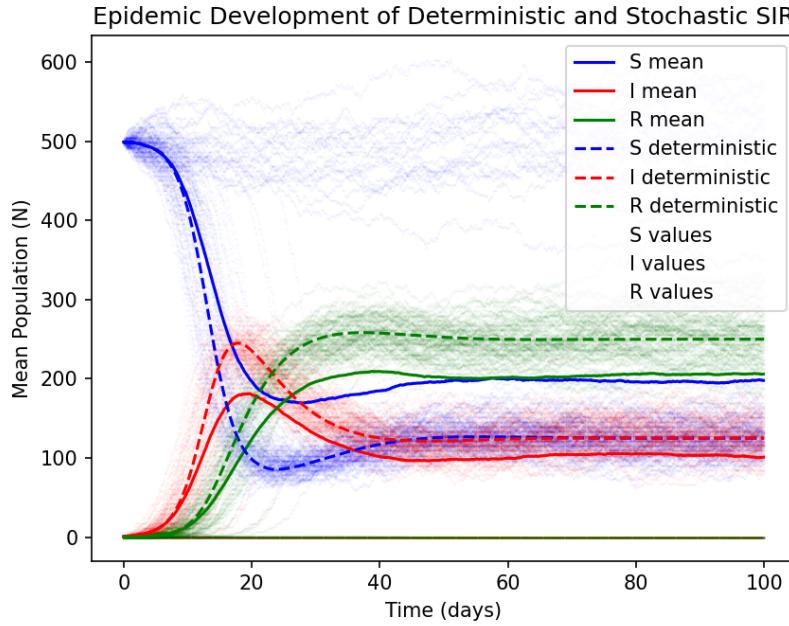


Figure 3: Multiple runs of the stochastic SIR model versus the deterministic model for the same set of parameters. The stochastic SIR model was run 100 times and plotted for each population category from which a mean was computed and plotted as well. The deterministic model was run once and plotted too. Beta, gamma and mu parameters were set as 0.6, 0.1 and 0.05 respectively. The initial population size was set at 500. One infected individual was introduced to a fully susceptible population at t_0 .

Additionally, we can notice the total population increases in some cases as the infected population reaches values higher than 500 in 3. When we compare the change in population sizes between figures 3 and 4 we can see that increasing the population sizes seems to decrease the stochasticity. Especially when we look at a larger number of runs for the stochastic model as shown in figure 3, we can notice how much it may vary from the deterministic model for the same set of parameters and initial conditions. We may notice that early deviation from the deterministic model may result in immediate or rapid extinction of the disease. We may also notice that the average values for the stochastic models are higher for the susceptible population, lower for the infected population and lower for the recovered population within these parameters. This lower average is likely due to early extinction possibilities within the stochastic model.

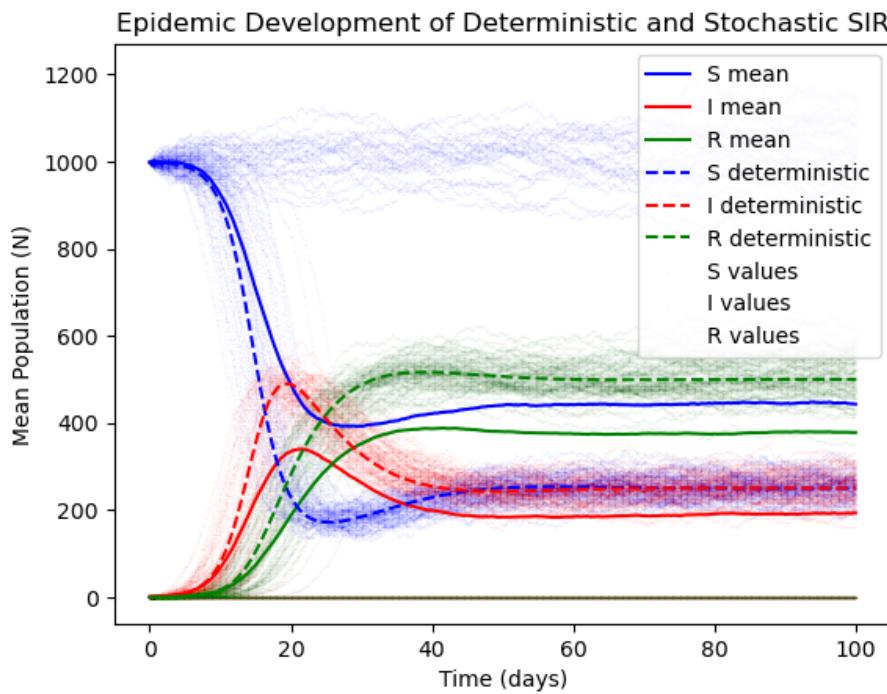


Figure 4: Multiple runs of the stochastic SIR model versus the deterministic model for the same set of parameters. The stochastic SIR model was run 100 times and plotted for each population category from which a mean was computed and plotted as well. The deterministic model was run once and plotted too. Beta, gamma and mu parameters were set as 0.6, 0.1 and 0.05 respectively. The initial population size was set at 1000. One infected individual was introduced to a fully susceptible population at t_0 .

When values for R_0 approach 1, we notice that the mean infected of the population is relatively similar to the deterministic mean and we also do not see large deviations from this mean as we did with a larger value for R_0 .

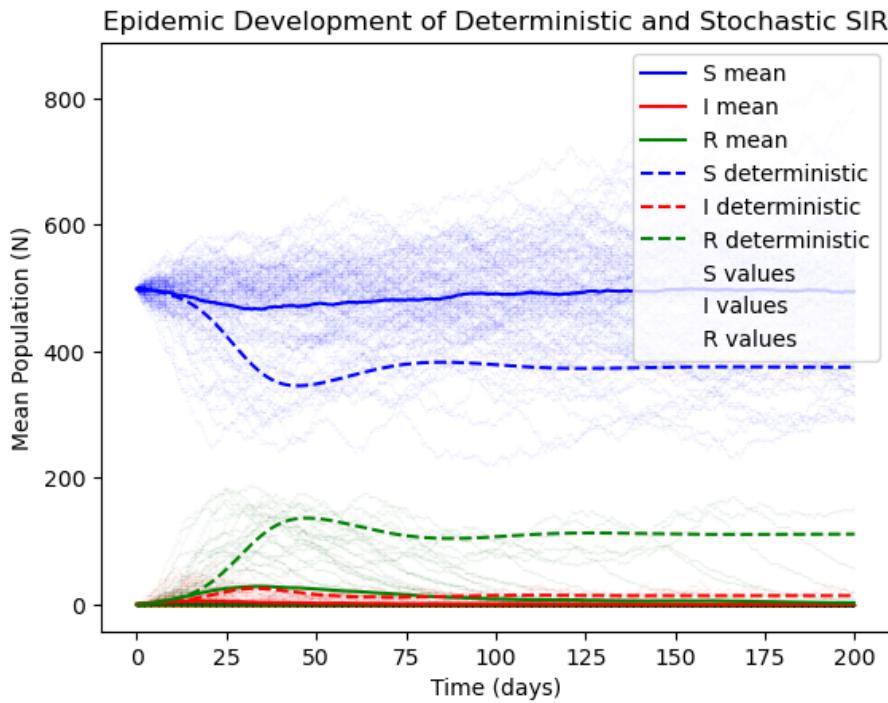


Figure 5: Multiple runs of the stochastic SIR model versus the deterministic model for the same set of parameters. The stochastic SIR model was run 100 times and plotted for each population category from which a mean was computed and plotted as well. The deterministic model was run once and plotted too. Beta, gamma and mu parameters were set as 0.6, 0.5 and 0.05 respectively. The initial population size was set at 500. One infected individual was introduced to a fully susceptible population at t_0 .

3.2 Inter-Run Variance given different R_0 -values

Looking at how the inter-run variance changes for a given R_0 , we found that as R_0 increases, the maximum amount of infected individuals seems to vary more across multiple runs.

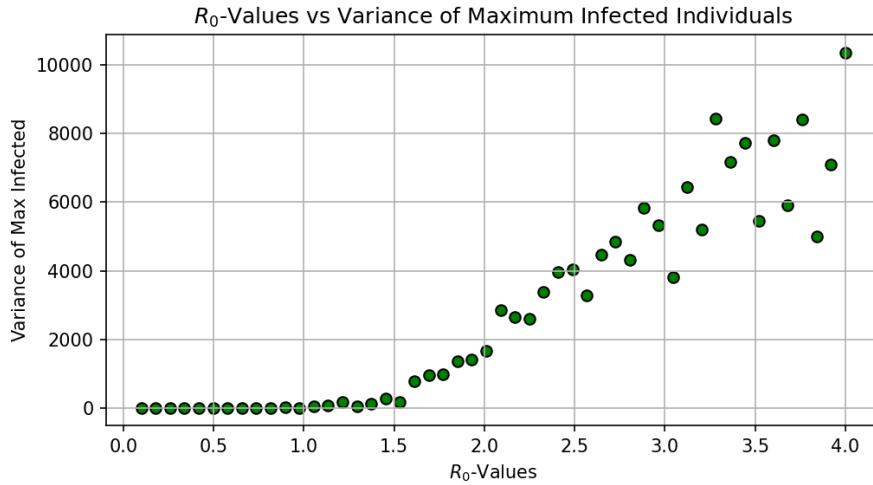


Figure 6: Between-run variance of stochastic SIR-model given different R_0 values. A total of 50 values were chosen at an equal spacing ranging from 0.1 to 4. Beta was computed from a range of gamma values from 0.1 to 0.9. Per R_0 - value, the simulation was run 20 times. The total population size was 500. A single infected individual was introduced into a fully susceptible population at t_0 .

The results shown in figure 6, might arise from the fact that for small R_0 values, the speed with which the infected population (I) changes is relatively slow, leading to less ambiguous outcomes. However, as R_0 increases, the fast changing dynamics of I as well as stochasticity can lead to a wide range of different peaks in the epidemic. The outcome of the epidemic might especially be determined by early events, which, in light of a large R_0 , can vary greatly.

3.3 Covariance between S and I given different R_0 - values.

When looking at the strength of the covariance between the infected and susceptible population, we found results that are inverse compared to what we found in section 3.2. More precisely, we found that as R_0 increases, the covariance between S and I becomes more and more negative.

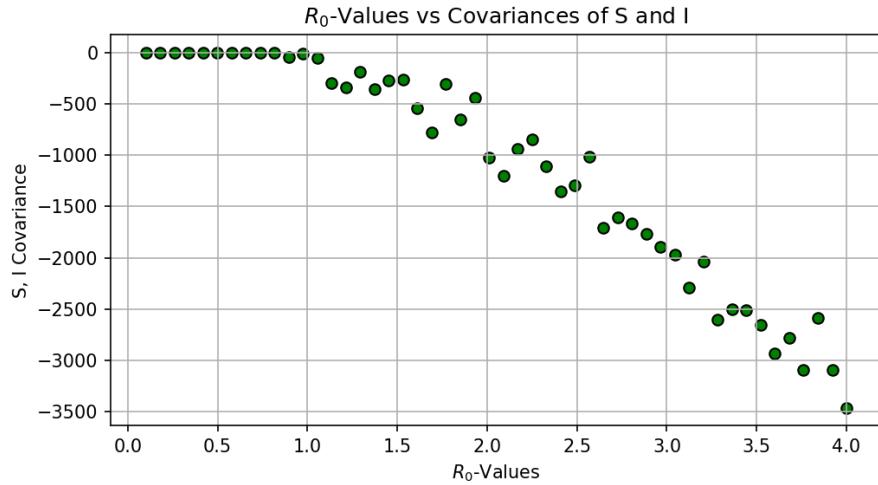


Figure 7: Covariance between S and I given different R_0 -values. A total of 50 values were chosen at an equal spacing ranging from 0.1 to 4. Beta was computed from a range of gamma values from 0.1 to 0.9. Per R_0 - value, the simulation was run 20 times. The total population size was 500. A single infected individual was introduced into a fully susceptible population at t_0 .

This result seems intuitive, since a higher infection pressure will lead to a more rapid decrease in S , while I increases at virtually the same rate. This leads to a sharper inverse relationship between the two.

3.4 Stochastic Resonance and Increased Transients

When we consider the deterministic SIR model, we may notice slight oscillations around the equilibrium due to the natural birth and death rates. When we follow any of the stochastic simulations in figure 8 we can see that they may temporarily increase before decreasing again. This may be due to enhancements of these deterministic oscillations.

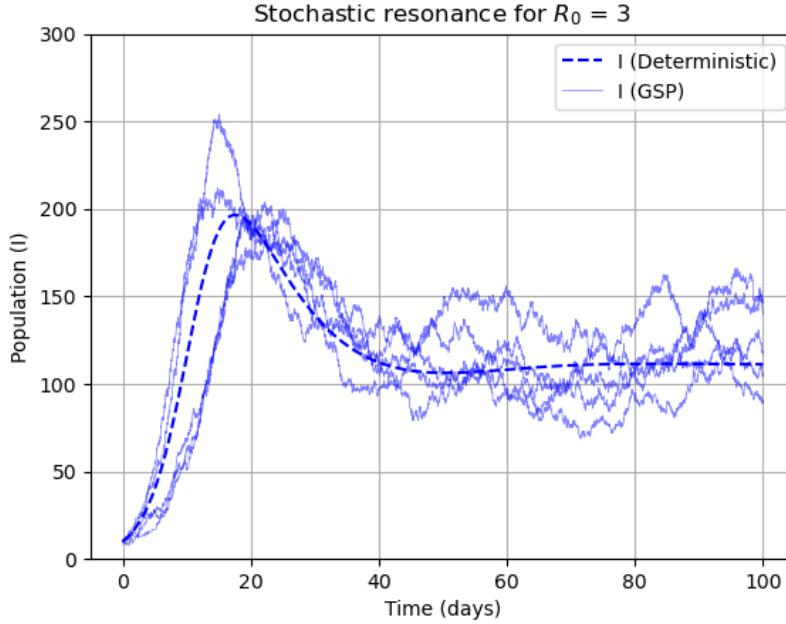


Figure 8: The effect of stochastic resonance in a Deterministic SIR model for $R_0 = 3$. Values for gamma and mu were set as 0.1 and 0.05 respectively while beta was computed from these parameters. The population size N was set as 500 with 10 infected individuals introduced to a fully susceptible population. The stochastic model was repeated five times while the deterministic model was run once.

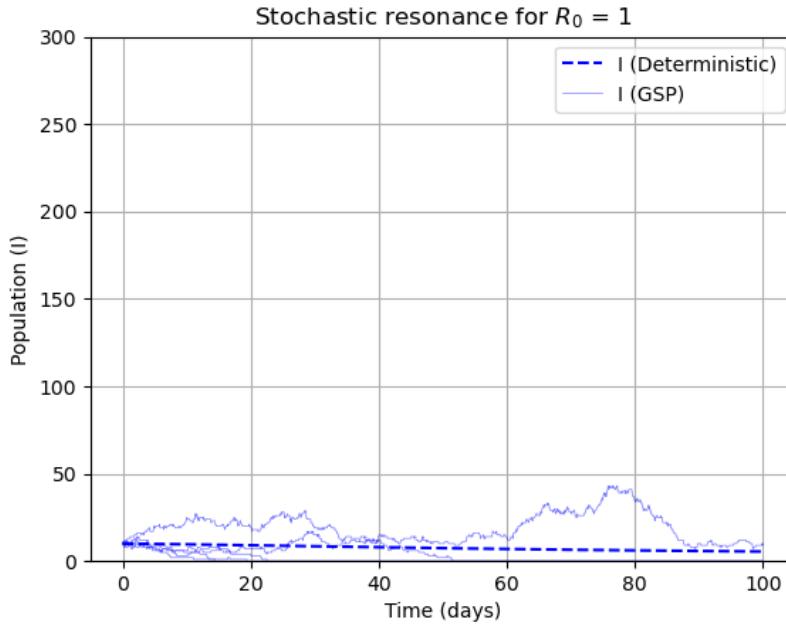


Figure 9: The effect of stochastic resonance in a Deterministic SIR model for $R_0 = 1$. Values for gamma and mu were set as 0.1 and 0.05 respectively while beta was computed from these parameters. The population size N was set as 500 with 10 infected individuals introduced to a fully susceptible population. The stochastic model was repeated five times while the deterministic model was run once.

Additionally, when considering the endemic model of figure 9 we may notice these amplified oscillations as well. However, when we consider a SIR model with a basic reproductive number smaller than 1, it may become difficult to notice any oscillatory behavior. This may be due to how we acquire the data for the GSP SIR model. In this model, the population is always considered as integers and can therefore not be expressed as floats, when the steps are relatively big compared to the oscillatory behavior, it may become hard to notice if there is any. This behavior would likely be more visible for lower values of R_0 if the total population and initial infected population is increased.

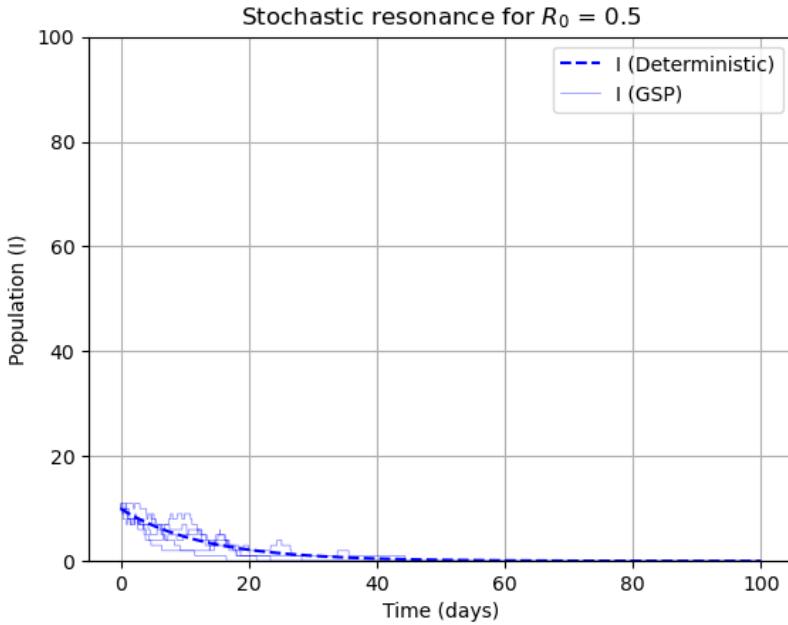


Figure 10: The effect of stochastic resonance in a Deterministic SIR model for $R_0 = 10.5$. Values for gamma and mu were set as 0.1 and 0.05 respectively while beta was computed from these parameters. The population size N was set as 500 with 10 infected individuals introduced to a fully susceptible population. The stochastic model was repeated five times while the deterministic model was run once.

To consider the increasing transients, we refer back to figure 3 as well where we may see increased transients due to stochasticity, leading to temporary shifts from an equilibrium as we might see in the highest peak in figure 8 where a temporary change slowly reverts back to an equilibrium. In more extreme cases, these may lead to a shifted equilibrium as the model approaches values that are incompatible with the expected equilibrium. This may occur due to disease extinction or a relatively small susceptible population after which the model may settle into the new equilibrium until another transient may occur. This can be seen in the differences between the specific stochastic runs in figure 3.

3.5 Extinction Events

Looking at what parameter values are most likely to cause frequent extinctions, it seems there is a noticeable negative correlation between R_0 and the total number of extinctions across runs.

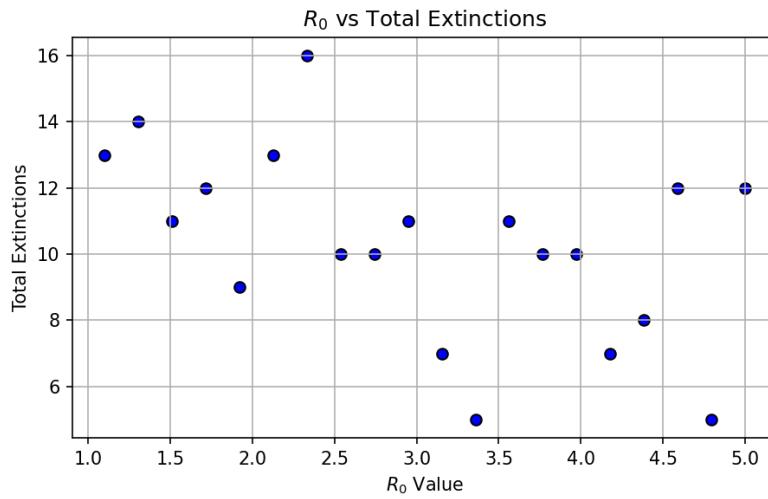


Figure 11: Total number of extinction across multiple runs for varying R_0 -values. A total of 20 values were chosen at an equal spacing, ranging from 1.1 to 5. This was computed using $\gamma = 0.1$ and the corresponding β -values. Per R_0 - value, the simulation was run 50 times. The total population size was 500. A single infected individual was introduced into a fully susceptible population at t_0 . This demographic SIR-model includes a natural birth and death rate $\mu = 0.0001$. Each simulation was run for 100 time steps.

The transmission force is higher when a large R_0 is selected, thus infections grow more rapidly. This would explain the results in the figure, as the chance of extinctions at the early stages of the epidemic decreases when I is still small but increases relatively fast.

Looking at how varying N impacts the total number of extinctions, we see a similar trend as before with extinctions decreasing as a function of an increasing total population N .

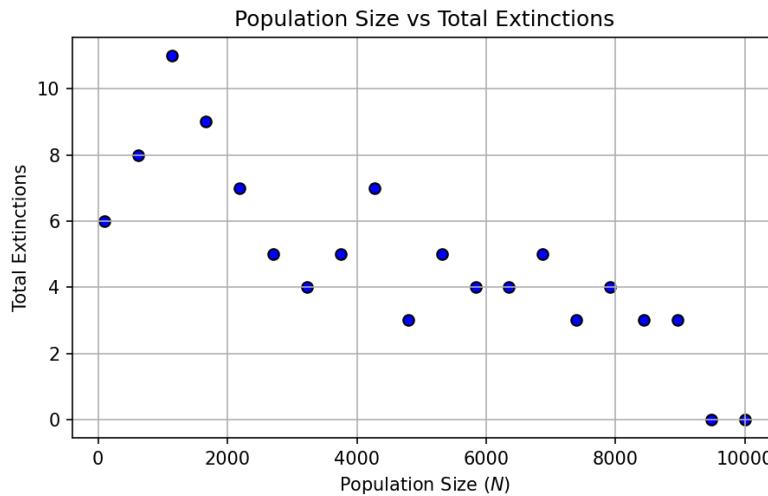


Figure 12: Total number of extinction across multiple runs for varying N -values. A total of 20 values were chosen at an equal spacing, ranging from 1000 to 10000. Sir-parameters were fixed at $\beta = 0.4$, $\gamma = 0.1$ and $\mu = 0.0001$. Per N - value, the simulation was run 50 times. A single infected individual was introduced into a fully susceptible population at t_0 . Each simulation was run for 100 time steps.

When a single infected individual is introduced into a relatively large population, the likelihood of the individual infecting another before recovering decreases as N increases. We can note in

the results that stochastic fluctuations have a smaller chance of causing premature extinctions as N grows.

When looking at extinctions as a function of N and R_0 simultaneously, it seems that N has a stronger impact compared to the basic reproduction number.

Extinction Counts Across Population Size and R_0

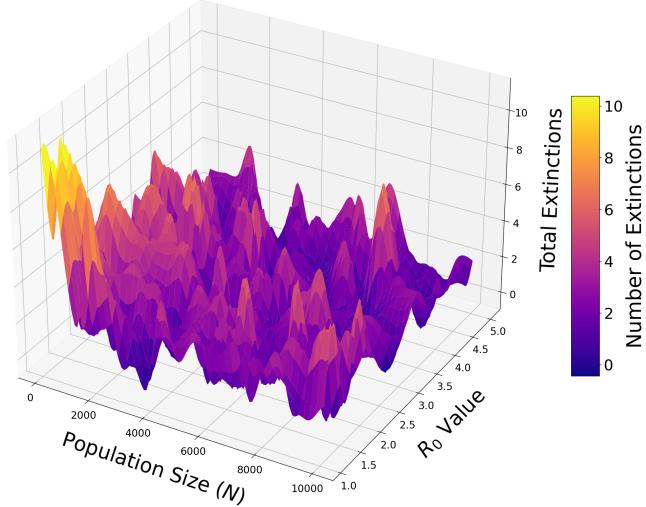


Figure 13: Total extinction counts across varying population sizes N and basic reproduction numbers R_0 . A total of 20 R_0 values were selected with equal spacing, ranging from 1.1 to 5, computed using $\gamma = 0.1$ and corresponding β -values. Each R_0 value underwent 20 simulation runs with population sizes ranging from 1000 to 10000. The model was initialized by introducing a single infected individual into a fully susceptible population at t_0 . This demographic SIR model accounts for natural birth and death rates, with $\mu = 0.0001$. Each simulation was run for 100 time steps.

These results indicate that a low N contribute more greatly to the model's stochasticity than a low basic reproduction number R_0 , being the stronger predictor of high extinction rates.

3.6 Generating Erdős-Renyi Networks and Simulating Disease-Spread

When generating multiple ER-graphs while varying the connection probability p we found that as p increases the average minimum and maximum degree also increases. More precisely both of these values increase at roughly the same rate.

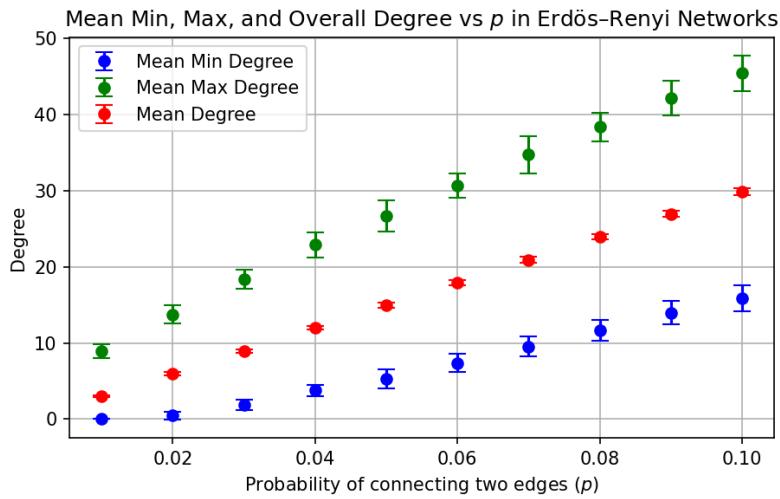


Figure 14: Average minimum, maximum and overall average number of nodes of ER-graphs given different connection probabilities p . 10 values were selected for p , with evenly spaced values ranging from 0.01 to 0.1. Error bars indicate the standard deviation around the mean. Generated networks have a total of 300 nodes. Statistics are based on 50 graphs per p -value.

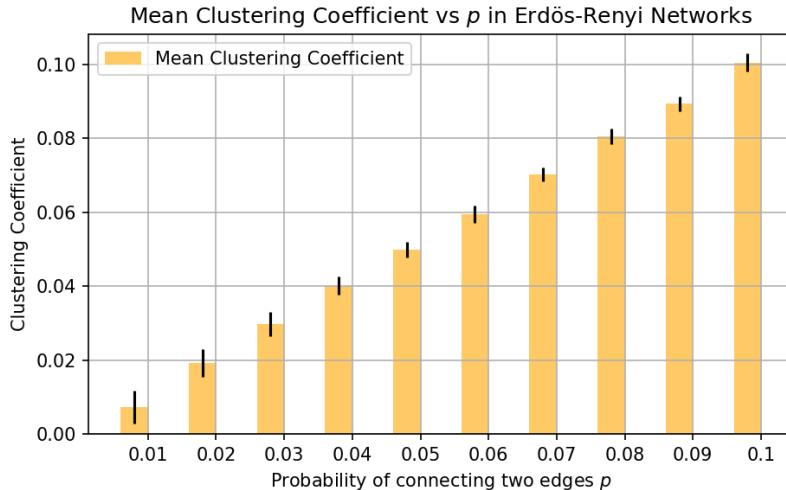


Figure 15: Average clustering coefficient of ER-graphs given different connection probabilities p . 10 values were selected for p , ranging from 0.01 to 0.1, evenly spaced. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. Statistics are based on 50 graphs per p -value.

From figure 14 we see that the general average and average minimum and maximum increase linearly with p . This indicates that there is no preferential attachment when adding new edges. This is in correlation with the uniformly and randomly generated edges in this network type. We can state the same for the clustering coefficient in figure 15, which increases linearly with p . When simulating the SIR spread given different R_0 values on the networks generated above, we find that p also has a noticeable positive impact on the disease-spread, even when R_0 is smaller than 1.

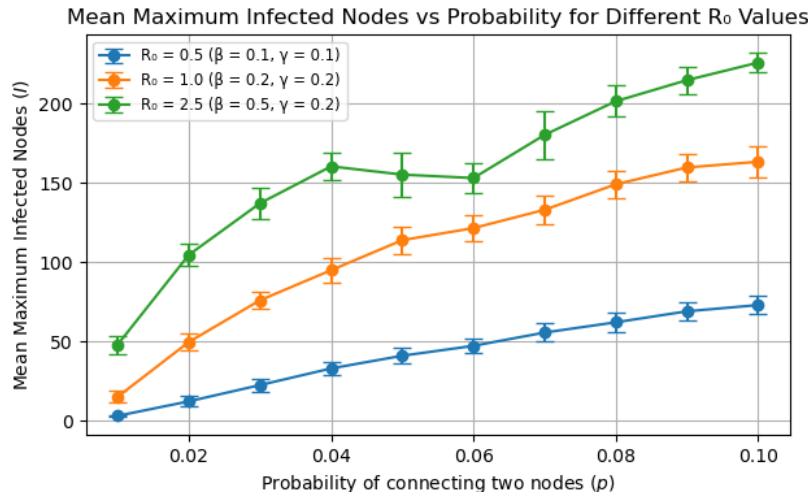


Figure 16: Average maximum infected given different p -values as well as R_0 -values. 10 values were selected for p , ranging from 0.01 to 0.1, evenly spaced. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. The initial fraction of infected was 0.01. Disease-spread was simulated 50-times for 50 iterations for each given p -value.

From the results in figure 16 we can ascertain that a larger p leads to higher overall connectivity in the network. This indicates that an increase in p results in a higher likelihood of infected nodes being connected to multiple susceptible nodes, which would lead to an accelerated increase in infected nodes as the infection can spread to more nodes per iteration.

3.7 Generating Barabási-Albert Networks and Simulating Disease-Spread

Going forward, we again generated multiple graphs. This time, we used the Barabási-Albert technique, varying m , i.e. the maximum number of existing nodes a new node will be connected to. When looking at the aforementioned network statistics, we find again that, as the network parameter (m) increases the average minimum, maximum and overall degree of the network increases. This time, however, the maximum number of edges increases with a significant faster rate than the minimum number of edges.

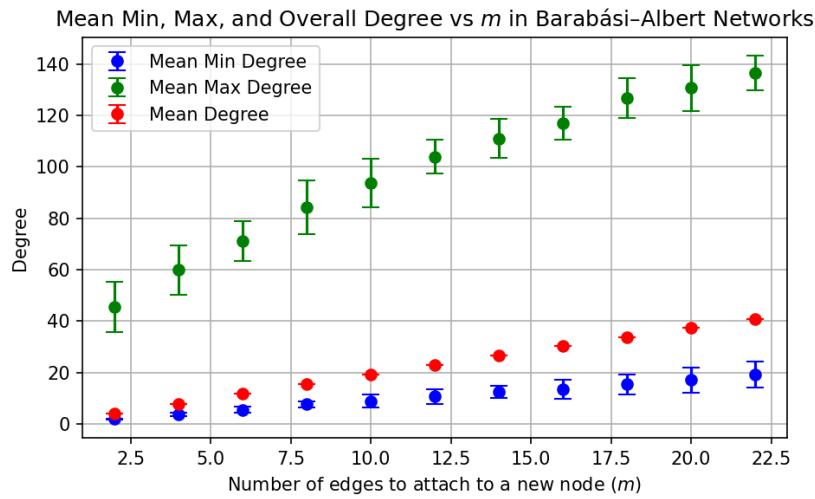


Figure 17: Average minimum, maximum and overall average number of nodes of BA-graphs given different values for m . 12 evenly spaced values were selected for m , ranging from 2 to 22. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. Statistics are based on 50 graphs per m -value.

From figure 17 we see that the likelihood of a node being connected to a hub increases with m . While nodes connecting to hubs is also being amplified by preferential attachment. This means that already existing hubs become larger, resulting in the relatively sharp increase in maximum degrees as compared to minimum degrees. We can notice this in the figure from the increasing difference between the maximum mean degree and the minimum mean degree.

When looking at the average clustering coefficient, we again see it increase as a function of m .

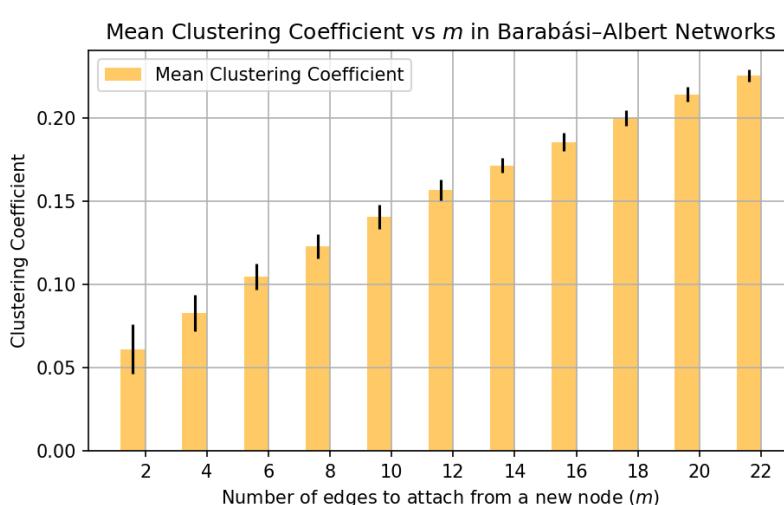


Figure 18: Average clustering coefficient of BA-graphs given different m -values. 12 evenly spaced values were selected for m , ranging from 2 to 22. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. Statistics are based on 50 graphs per m -value.

The upward trend in figure 18 indicates that the clustering coefficient increases as m increases, which could again be explained by preferential attachment. This is in correlation with the BA Network where high-degree nodes are likely to increase their degree indefinitely resulting in increasing the number of neighboring nodes. Nodes that are connected to hubs are likely to con-

nect to other nodes that are also connected to said hub, creating local environments with greater clustering.

Looking at disease-spread on the network 19, we can see that the parameter m has a significant impact on the spreading potential of a disease. Even when $R_0 < 1$, which would indicate that the recovery rate surpasses that of the infection, the parameter m exercises significant influence on the spreading of the infection. This continuously highlights the importance of hubs. In a population network these hubs may represent people who are in contact with many different individuals throughout the day.

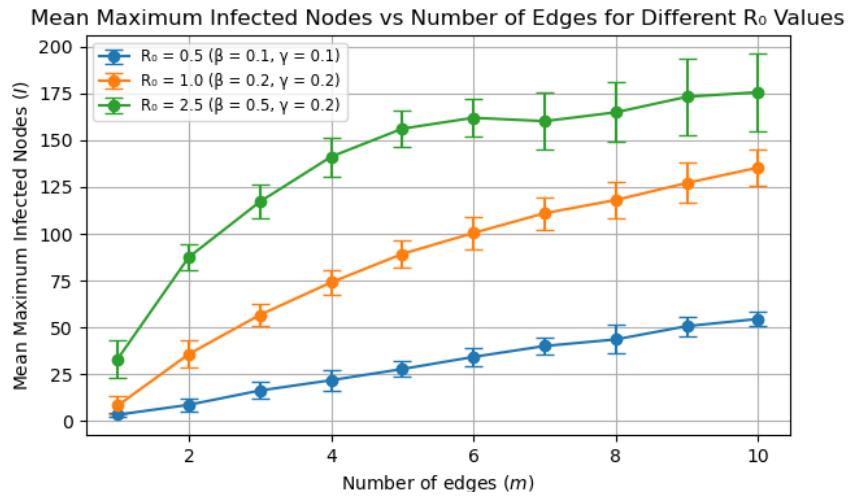


Figure 19: Average maximum infected given different m -values as well as R_0 -values. 10 evenly spaced values were selected for p , ranging from 0.01 to 0.1. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. The initial fraction of infected was 0.01. Disease-spread was simulated 50-times for 50 iterations for each given m -value.

3.8 Generating Watts-Strogatz Networks and Simulating Disease-Spread

Looking at the final network type, multiple graphs were generated by varying a specific hyper-parameter in the process. This time, the parameter at hand was the rewiring probability p_r . As mentioned previously, this value determines how likely a node will randomly rewire one of its connections to a different random node.

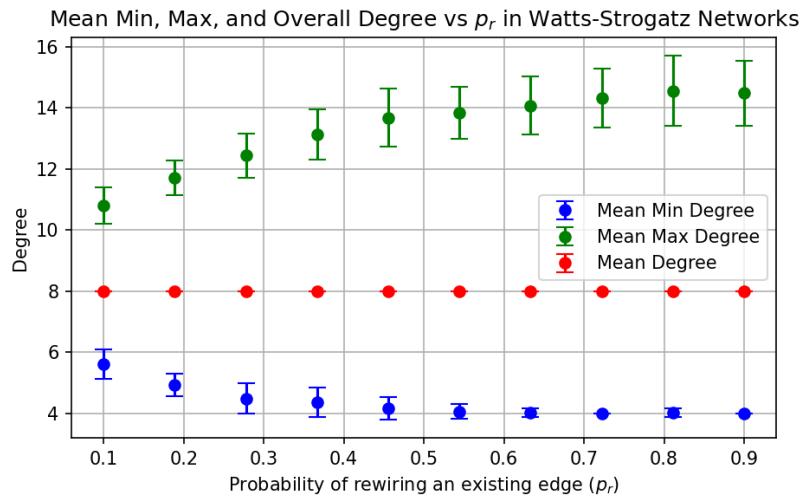


Figure 20: Average minimum, maximum and overall average number of nodes of WS-graphs given different values for p_r . 10 evenly-spaced values were selected for p_r , ranging from 0.1 to 0.9. A static value $k = 8$ is used to define the number of edges. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. Statistics are based on 50 graphs per p_r -value.

From figure 20 we can notice that as p_r increases, the average maximum degree increases, while the average minimum degree decreases. Furthermore, the overall average degree seems to remain at the number of neighbors we defined at initialization ($k = 8$). As a large p_r indicates an increased amount of rewiring of nodes over time, some nodes might be randomly assigned more nodes than average. The same can be found for the inverse with the mean minimum degree. This directly correlates to how the Network is defined. Since the network has a predetermined number of edges ($k = 8$), the total number of edges will always remain the same, so extremes like the maximum and minimum will cancel each other out in order to remain equal overall. When examining the clustering coefficient, it becomes evident that this value decreases as a function of increasing p_r .

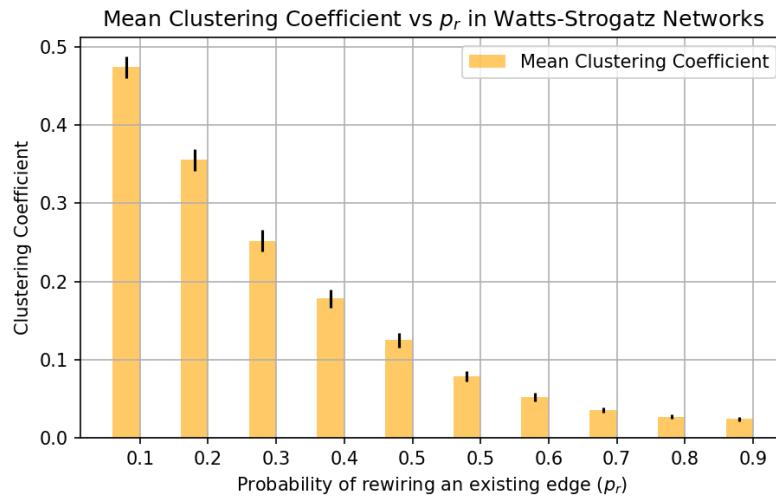


Figure 21: Average clustering coefficient of WS-graphs given different p_r -values. 10 evenly-spaced values were selected for p_r , ranging from 0.1 to 0.9. A static value $k = 8$ is used to define the number of neighbors per node. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. Statistics are based on 50 graphs per p_r -value.

The results shown in 21 seems intuitive given the structure of WS-Networks. Since, at iteration 0, each node is connected to its k - nearest neighbors, clustering is already high at that point in time. Given a high rewiring probability, these clusters are more likely to be broken up into less interconnected structures, leading to a lower clustering coefficient.

Looking at the joint impact of p_r and R_0 on the system, it seems that p_r has a rather small, positive impact on the spreading potential of a given disease, regardless of R_0 .

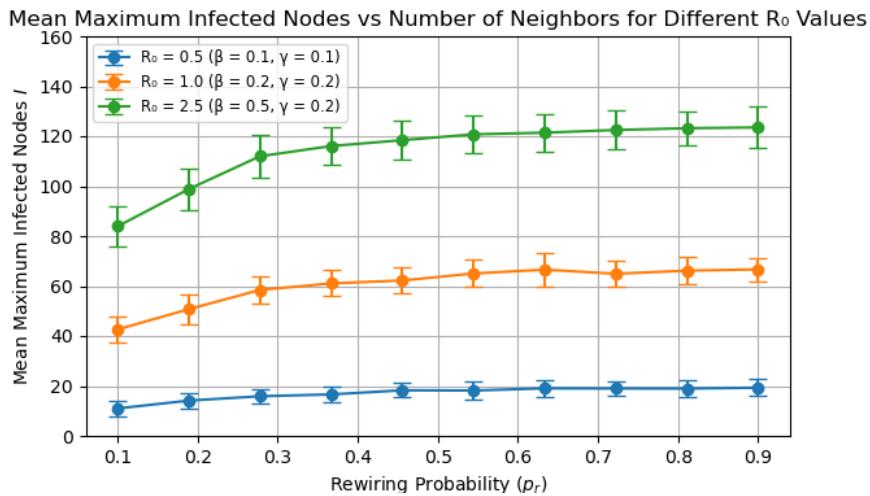


Figure 22: Average maximum infected given different p_r -values as well as R_0 -values. 10 evenly spaced values were selected for p_r , ranging from 0.1 to 0.9. Error bars indicate the standard deviation around the mean. Generated networks had a total number of 300 nodes. The initial fraction of infected was 0.01. Disease-spread was simulated 50-times for 50 iterations for each given p_r -value.

One explanation for this might be the fact that, even when p_r is high, the local-clustering structure typical for WS-networks is preserved. This means that, within a cluster, spreading occurs

significantly more frequently within said clusters compared to the global dynamics. This "robustness" to rewiring might aid keeping the infection dynamics similar across p_r -values.

3.9 Investigating Vaccination Strategies

Comparing the vaccination strategies discussed in sections 2.13 to 2.15, we found that both of our proposed systematic strategies significantly reduce the infection peak as compared to the null-strategy. However, when comparing both of our developed strategies with each other, there seems to be no significant difference between the two.

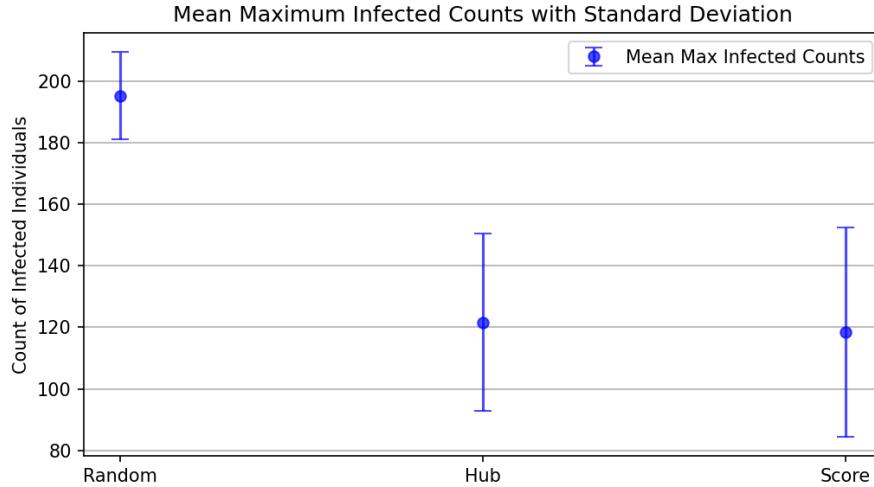


Figure 23: Comparison of the effect of three vaccination strategies on the average maximum infected count. A β -value of 0.05 and γ -value of 0.01 was used, resulting in $R_0 = 5$. The number of infected at $t = 0$ was 5. Statistics arise from 30 runs of each strategy. Each simulation was run for 50 iterations.

Looking at figure 23, it seems that passing additional metrics such as closeness and betweenness does not increase vaccination effectiveness significantly when compared to merely focusing on nodes with the highest degrees (hubs). This result could be considered surprising given the fact that the "score" strategy results in a higher average of maximum recovered, while the infection peak is the same across both methods.

When performing the strategies multiple times, we see in figure 23 that the overall standard deviation for the score campaign is larger than that of the hub campaign, resulting in both a solution with a lower peak of infected and a larger peak of infected than those for the hub campaign. Whilst a lower peak may indicate success, a high standard deviation also hints towards more unreliability in the score campaign compared to the hub campaign.

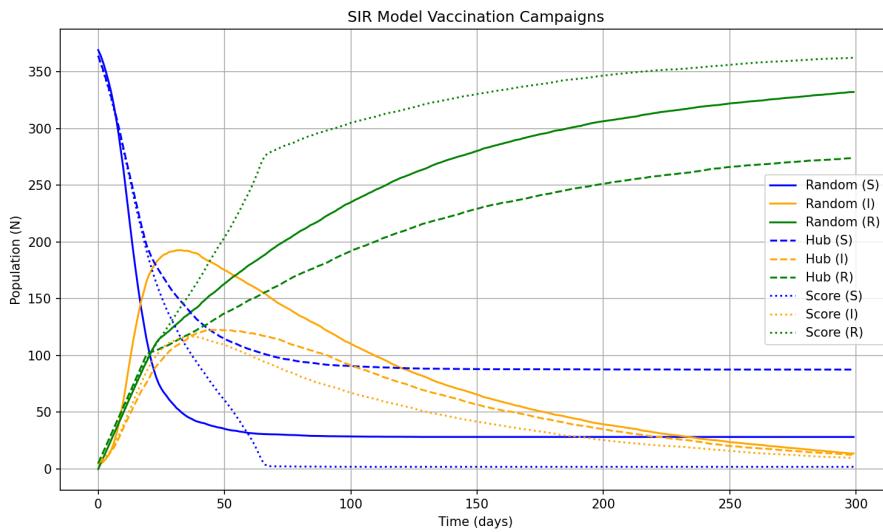
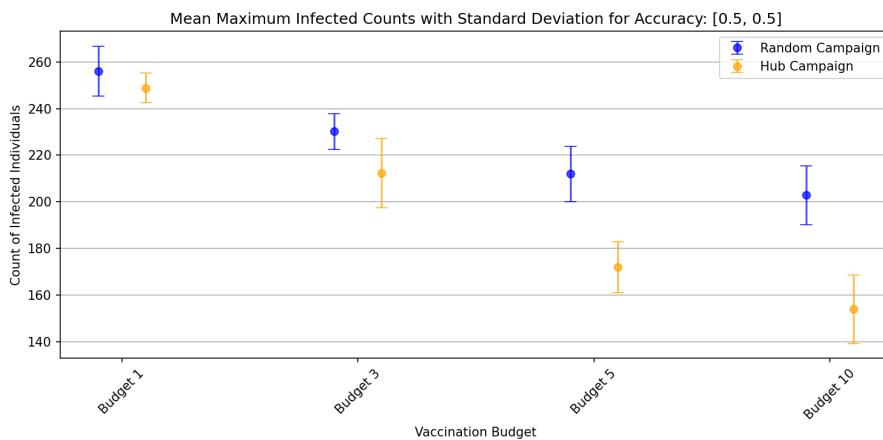


Figure 24: SIR-development given three vaccination strategies on the average maximum infected count. A β -value of 0.05 and γ -value of 0.01 was used, resulting in $R_0 = 5$. The number of infected at $t = 0$ was 5. Statistics arise from 30 runs of each strategy. Each simulation was run for 300 iterations.

The results from figures 23 and 24 indicate that the "hub" strategy is as effective as the "score" method, while vaccinating less individuals. This implies that, resource-wise, the "hub" strategy might be more efficient. However, we can also notice that the amount of susceptible individuals in the score strategy is significantly lower. This could prove useful when sudden spikes in infections occur by creating herd immunity.

Comparing the "hub" strategy against the null strategy given different test accuracies as well as daily vaccination budgets, we find the expected result that the "hub" strategy is superior in each case, however, it should be noted that the difference is minimal when the daily testing budget is 1. Which might be due to the fact that the spreading-force is too strong to be damped by merely vaccinating a single person a day.



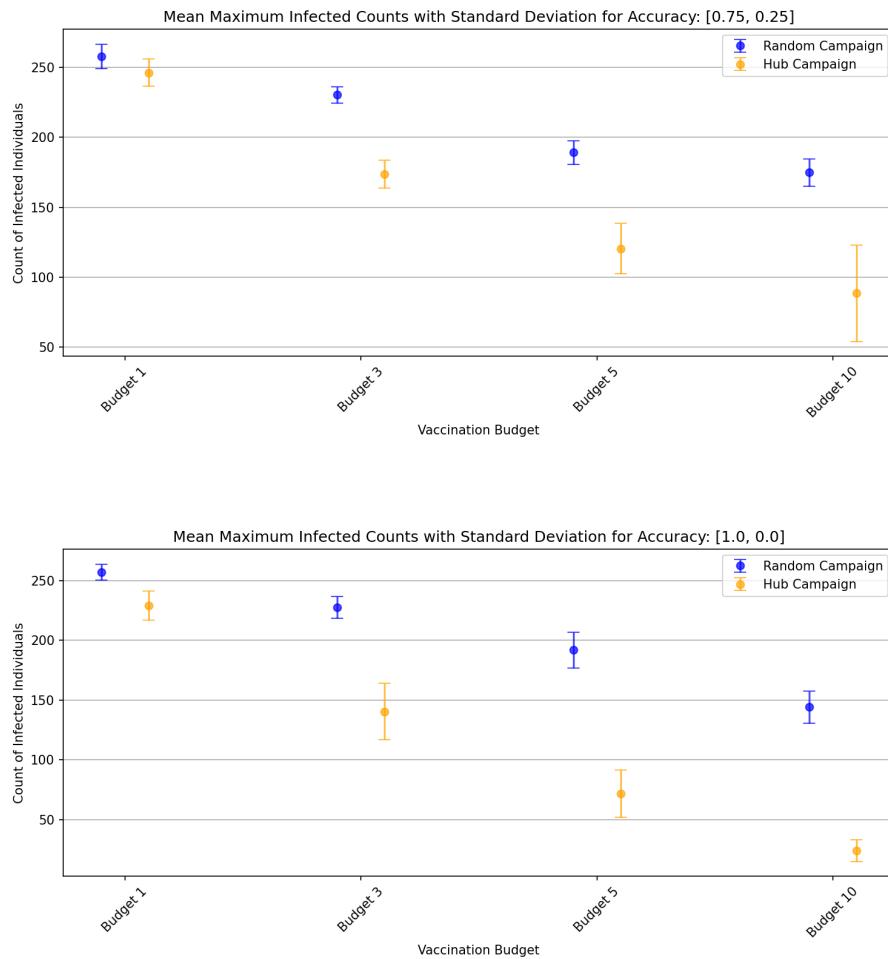


Figure 25: Mean number of maximum infected given null- and hub vaccination strategies. A β -value of 0.05 and γ -value of 0.01 was used, resulting in $R_0 = 5$. The number of infected at $t = 0$ was 5. Statistics arise from 30 runs of each strategy. Each simulation was run for 300 iterations. Testing accuracies ranged from 0 to 1 in steps of 0.25. Daily vaccination budgets ranged from 1 to 10 per time step.

4 Conclusion

The aim of this report is to examine the spread of diseases governed by the SIR-framework in a stochastic manner. This was first done using Gillespie's First Reaction method and comparing it with the deterministic SIR model. We noticed that stochasticity can greatly impact of the infection spread of the disease. In some cases the disease may even go extinct early on in the model even though the deterministic model approached an equilibrium. Stochastic resonance in the model seemed to amplify oscillations around the equilibrium while increased transients could lead to early extinction or a shifted equilibrium. When looking at certain outcomes that are due to stochasticity in said model, we found that R_0 and N can have a significant impact on these outcomes. More precisely, as both of these values grow larger, the number of stochastic extinctions decrease. This relates to a decrease in deviations from the stochastic mean for the populations when increasing N . Here, N seems to be more dominant when it comes to predicting the number of extinctions. Looking solely at R_0 we found that as this parameter increases, the between-run variance of the GSP becomes larger, due to the occurrence of various infection peaks of different amplitudes. Moreover, the more drastic infection when given a large R_0 leads to a more negative covariance between the susceptible and infected population.

Looking at network types and their structural differences, it seems that in Erdős-Renyi networks,

a higher parameter p can significantly increase the infection-peak, even when given smaller R_0 -values. The same holds for the parameter m in Barabási-Albert networks. This is due to the formation of large hubs with high spreading potential. When examining Watts-Strogatz network and the corresponding r_p hyperparameter. It seems that the epidemic develops rather independently from this value. Out of all the networks and their hyperparameters, our results suggest that the spreading potential is the largest in an Erdős-Renyi network with a large p . When investigating different vaccination strategies, we found that vaccinating individuals based on the number of connections they have to others, is an effective and cost-efficient strategy as compared to randomly vaccinating individuals. We also found that adding network statistics such as betweenness and closeness adds little benefit to the "hub" vaccination strategy. Future research on stochastic disease-spread modeling could focus on different approaches, such as discrete-time instead of discrete-event models. The advantage here would be the ability to incorporate observational noise instead of process noise. This might be helpful when accounting for measurement error, which is unavoidable when dealing with large populations across cities or countries. Considering networks, epidemic spread throughout networks could focus on different network-topologies that were not discussed in this report. One topology that comes to mind arises from a so-called hub-and-spoke network. The key feature of said network is a single central node (hub) connected to many peripheral nodes that are not connected to one another (spokes). This network is commonly used in modelling transport systems. Hence, it might provide useful insights into the spreading dynamics of infectious disease within cities with displaying large public-transport infrastructures[9].

References

- [1] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks". In: *Science* 286.5439 (1999), pp. 509–512. DOI: 10.1126/science.286.5439.509.
- [2] John C Butcher. "Runge-Kutta methods for ordinary differential equations and the numerical solution of spacecraft trajectory problems". In: *Celestial Mechanics and Dynamical Astronomy* 102.1 (2008), pp. 233–249.
- [3] P. Erdős and A. Rényi. "On Random Graphs I". In: *Publicationes Mathematicae Debrecen* 6 (1959), pp. 290–297.
- [4] Luca Gammaitoni et al. "Stochastic resonance". In: *Reviews of Modern Physics* 70.1 (1998), pp. 223–287. DOI: 10.1103/RevModPhys.70.223.
- [5] Daniel T. Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434. DOI: 10.1016/0021-9991(76)90041-3.
- [6] Sergio Gómez. "Centrality in Networks: Finding the Most Important Nodes". In: *URV DEIM* (2019). URL: https://webs-deim.urv.cat/~sergio.gomez/papers/Gomez-Centrality_in_networks-Finding_the_most_important_nodes.pdf.
- [7] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press, 2008, p. 366.
- [8] Russell Lande. "Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes". In: *The American Naturalist* 142.6 (1993), pp. 911–927. DOI: 10.1086/285580.
- [9] Morton E. O'Kelly. "The Location of Interacting Hub Facilities". In: *Transportation Science* 20.2 (1986), pp. 92–106. DOI: 10.1287/trsc.20.2.92.
- [10] Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* 393.6684 (1998), pp. 440–442. DOI: 10.1038/30918.