

BACHELOR INFORMATICA



UNIVERSITY OF AMSTERDAM

# Different SIR models and their relevance

Pjotr Piet & Maarten Stork

November 12, 2024

**Supervisor(s):**

**Signed:**

## Abstract

The recent COVID-19 pandemic highlighted the critical need for understanding disease transmission dynamics within populations. This study employs the Susceptible-Infected-Recovered (SIR) model to investigate the complexities of disease spread through stochastic and deterministic approaches, particularly focusing on the Gillespie algorithm as a stochastic modeling tool. The research examines the effects of various network topologies—specifically Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER)—on infection dynamics. Additionally, it evaluates the efficacy of targeted versus random vaccination strategies within a real-world contact network derived from sociopatterns data.

Through comparative simulations, the findings reveal that network structure significantly influences the spread of infectious diseases, with heterogeneous networks facilitating more rapid transmission. The targeted vaccination strategy demonstrated superior effectiveness in reducing infection rates compared to random vaccination, particularly under varying testing accuracies and resource allocations. This work contributes to the understanding of epidemic modeling and highlights the importance of strategic public health interventions, offering insights into improving disease management in future outbreaks.

---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Background theory . . . . .	5
1.1.1	Gillespie Algorithm . . . . .	5
1.1.2	Transients and Stochastic Resonance . . . . .	7
1.1.3	SIR Model and Its Variations . . . . .	7
1.1.4	Different Types of Networks . . . . .	9
1.1.5	Network Statistics . . . . .	10
1.1.6	Vaccination Strategies . . . . .	11
1.2	Related work . . . . .	12
1.2.1	Stochastic SIR model including demography; birth and death rate . . . . .	12
1.2.2	SIR & Networks . . . . .	13
1.3	Research Questions . . . . .	13
1.3.1	Gillespie-Related Research Question . . . . .	13
1.3.2	Network-Related Research Question . . . . .	14
1.4	Report structure . . . . .	14
<b>2</b>	<b>Method</b>	<b>16</b>
2.1	Gillespie Algorithm . . . . .	16
2.2	SIR model with demography . . . . .	17
2.3	Experiments for Gillespie-Related Research Question . . . . .	18
2.3.1	Implementation and comparison of Gillespie's algorithm . . . . .	18
2.3.2	Investigate Simulation Variability and Negative Co-variance . . . . .	19
2.3.3	Stochastic resonance and increased transients . . . . .	20
2.3.4	Extinction events and Critical Community Size . . . . .	21
2.4	SIR Model on Networks . . . . .	21
2.5	Different Types of Networks . . . . .	23
2.6	Vaccination Strategies . . . . .	23
<b>3</b>	<b>Results</b>	<b>25</b>
3.1	Gillespie . . . . .	25
3.1.1	Implementation Gillespie . . . . .	25
3.1.2	Investigate Simulation Variability and Negative Co-variance . . . . .	28
3.2	Stochastic resonance and increased transients . . . . .	31
3.2.1	Stochastic resonance general . . . . .	31
3.2.2	Model parameters and resonance and examples . . . . .	31
3.2.3	Increased transients . . . . .	34
3.2.4	Extinction events and Critical Community Size . . . . .	34
3.3	SIR Model on Networks . . . . .	34
3.3.1	SIR Model on BA/WS/ER Networks . . . . .	34
3.3.2	SIR Model on Real-Life Network . . . . .	35
3.4	Network Statistics . . . . .	37
3.5	Vaccination Strategies . . . . .	38

<b>4</b>	<b>Discussion</b>	<b>42</b>
4.1	Implementation of Gillespie . . . . .	42
4.2	Stochastic resonance and increased transients . . . . .	42
4.3	Extinction events and Critical Community Size . . . . .	43
4.4	SIR Model on Networks . . . . .	43
4.4.1	SIR Model on BA/WS/ER Networks . . . . .	43
4.4.2	SIR Model on Real-Life Network . . . . .	44
4.5	Network Statistics . . . . .	44
4.6	Vaccination Strategies . . . . .	45
<b>5</b>	<b>Conclusion</b>	<b>47</b>
5.1	Conclusion Gillespie-Related Research Question . . . . .	47
5.2	Conclusion Network-Related Research Question . . . . .	47

---

## CHAPTER 1

# Introduction

---

The recent COVID-19 pandemic has shown how important it is to understand how diseases spread in a population. Traditional models like the Susceptible-Infectious-Recovered (SIR) model are useful but often miss the randomness that plays a big role in real-world outbreaks, especially in smaller populations. This report aims to address that by implementing a stochastic version of the SIR model using Gillespie's algorithm, a method commonly used to simulate systems where randomness is key.

Along with exploring the random nature of disease spread, we also look at deterministic SIR models applied to specific network structures. By considering how individuals in a population are connected (for example, through social networks), we aim to see how the structure of these connections affects the spread of disease. This helps us make models that are more realistic and better reflect real-world situations.

The report focuses on several important features of Gillespie's algorithm, such as its ability to capture fluctuations, temporary changes in behaviour, and the concept of stochastic resonance. We also explore specific phenomena like the negative relationship between infected and recovered individuals, as well as extinction events—both of which are not fully captured by simple deterministic models. By comparing the stochastic model with deterministic SIR models on networks, this report aims to give a clearer understanding of how randomness, network structure, and population behaviour influence the spread of disease.

### 1.1 Background theory

This section provides an overview of the foundational theories relevant to our study. The key concepts and frameworks that underpin the analysis and discussions presented in the subsequent chapters will be explored.

#### 1.1.1 Gillespie Algorithm

In his 1976 paper, *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions* [15], Daniel T. Gillespie introduced a powerful stochastic algorithm to model the time evolution of chemical systems where the number of molecules is small enough that random fluctuations become significant. This approach provided an alternative to deterministic models, which rely on ordinary differential equations (ODEs) and assume continuous and well-mixed populations of reactants. For many biological and chemical systems, especially those at the molecular level, deterministic models fail to capture the inherent randomness in reaction events. Gillespie's method addresses this by treating chemical reactions as discrete and random processes.

#### 1.1.1.1 Key Assumptions and Framework

Gillespie's algorithm operates on a few fundamental assumptions:

**Well-Mixed System:** The algorithm assumes that the system is spatially homogeneous, meaning all molecules are equally likely to interact with each other. This simplifies the treatment of reaction probabilities and times.

**Stochastic Reaction Rates:** Each chemical reaction occurs randomly, and the probability of a specific reaction happening depends on the number of reactant molecules and the reaction rate constant. The algorithm treats the time between reactions as a random variable following an exponential distribution, which reflects the memoryless nature of molecular collisions and reaction events.

**Discrete Molecular Events:** The system is modelled as discrete molecular populations, where the exact number of molecules of each species is tracked over time.

#### 1.1.1.2 The Two Main Steps of Gillespie's Algorithm

**Reaction Propensity and Waiting Time** For each possible reaction, Gillespie introduced the concept of *propensity functions*, which define the likelihood that a specific reaction will occur in the next infinitesimal time step. The propensity  $a_i$  for a reaction  $i$  depends on the number of reactant molecules available and the reaction rate constant. The algorithm generates the waiting time  $\tau$  until the next reaction using an exponential distribution:

$$\tau = \frac{-\ln(r_1)}{a_1} \quad (1.1)$$

where  $r_1$  is a uniformly distributed random number between 0 and 1, and  $a_0$  is the sum of all reaction propensities.

**Reaction Selection** After calculating the waiting time until the next reaction, the algorithm selects which reaction will occur. The probability that a given reaction  $i$  occurs is proportional to its propensity  $a_i$ . The reaction to occur is chosen by generating another random number  $r_2$  and using it to select a reaction based on cumulative propensity.

Once a reaction is selected, the system's state (i.e., the number of molecules) is updated according to the stoichiometry of the reaction, and the process repeats.

#### 1.1.1.3 Advantages of Gillespie's Method

Gillespie's algorithm offers several advantages for modelling stochastic systems, particularly when populations of molecules are small, and random fluctuations have a significant impact on system behaviour. These advantages include:

**Capturing Intrinsic Noise:** The algorithm accounts for the intrinsic randomness in reaction events, making it especially useful for systems where stochastic effects dominate (e.g., gene regulation or viral infections).

**Exact Timing of Events:** Unlike deterministic models, which average out fluctuations by using continuous equations, Gillespie's method simulates the exact times at which reactions occur. This makes it highly accurate for small systems.

**Versatility:** Gillespie's method can be applied to a wide range of systems involving multiple reactions, and the framework is flexible enough to handle coupled reactions and complex systems with many interacting species.

#### 1.1.1.4 Applications in Epidemiology

Although originally developed for chemical reactions, the principles of Gillespie's algorithm have been extended to other fields, including epidemiology. In the context of disease modelling,

each "reaction" corresponds to an event in the disease transmission process (e.g., an individual becoming infected or recovering). The algorithm's ability to capture the stochastic nature of these events makes it particularly suited for simulating disease outbreaks in small populations, where randomness plays a major role in the spread of infection.

### 1.1.2 Transients and Stochastic Resonance

In dynamical systems, transient dynamics refer to the initial behaviour of a system as it approaches a steady state or equilibrium. These transients are essential for understanding how a system responds to perturbations, especially in biological systems where short-term behaviour can significantly differ from long-term dynamics.

Stochastic resonance is a phenomenon where the presence of noise enhances the response of a system to weak periodic signals, rather than simply adding randomness. In biological contexts, stochastic resonance has been observed in population dynamics, where environmental noise can amplify weak signals, enabling organisms or populations to "resonate" with specific frequencies and respond more predictably than in purely deterministic models.

The Gillespie algorithm, a widely used stochastic simulation method, is particularly suitable for modelling these phenomena in systems where reactions occur probabilistically. The algorithm's discrete nature captures the inherent randomness of transients and allows for the simulation of stochastic resonance in biological systems. In particular, the Gillespie algorithm enables simulations of fluctuations that reveal how noise can drive systems through transient states and potentially induce resonance effects that would not appear in a purely deterministic model.

For an in-depth exploration of transient dynamics and stochastic resonance using the Gillespie algorithm, see [9] for insights on noise-driven systems in chemical and biological contexts.

### 1.1.3 SIR Model and Its Variations

This section describes the base SIR model and a slight addition to the base model which includes demography. In these separate sections the differential equations for both systems are given as well as a small explanation on the  $R_0$  formulas for both models.

#### 1.1.3.1 Base SIR model

The classic SIR represents a set of ordinary differential equations to simulate a simple epidemic in a population. In this model,  $S$ ,  $I$ , and  $R$  represent the portions of the population that are susceptible, infected, and recovered, respectively. The parameter  $\beta$  is case-specific and represents the transmission rate of the disease. Additionally, the model assumes a constant total population, implying that the sum of all portions remains constant:  $S + I + R = \text{const}$ . For this report, the population is normalized to 1, such that  $S + I + R = 1$ . Lastly, the  $\gamma$  represents the recovery rate. This indicates how quickly the infected portion transitions to the recovered portion.

Combining these expressions we get the following set of differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

A key metric in the SIR model is the basic reproduction number,  $R_0$ , which represents the average number of secondary infections caused by a single infected individual in a fully susceptible population. It is given by the formula:

$$R_0 = \frac{\beta}{\gamma} \quad (1.2)$$

If  $R_0 > 1$ , the infection is expected to spread in the population, while  $R_0 < 1$  suggests that the infection will likely die out. More specifically, this measure is essential to determine the state of the model. When  $R_0 < 1$  we can say that the state of the disease will stay endemic. On the other hand, if the  $R_0 > 1$  we can say that the state of the disease will reach an epidemic state at some point in the simulation.

#### 1.1.3.2 SIR model with demography

The first variant of the basic SIR model described in this report introduces demographic dynamics. This is achieved by incorporating birth and death rates, represented by the parameter  $\mu$ . This parameter adds inflow to the susceptible group  $S$  and outflow from all other groups. Adding these demographic factors to the basic SIR model mentioned in the previous section 1.1.3.1, results in the following system of equations:

$$\begin{aligned}\frac{dS}{dt} &= \mu - \beta SI - \mu S \\ \frac{dI}{dt} &= \beta SI - (\mu + \gamma)I \\ \frac{dR}{dt} &= \gamma I - \mu R\end{aligned}$$

Here,  $\mu$  represents both the birth rate, adding to the susceptible population, and the death rate, reducing each population group. Despite these flows, the overall population size remains constant because individuals removed through death are replaced by births in the susceptible group.

With demographic factors, the basic reproduction number  $R_0$  is adjusted to reflect the rates of both recovery and natural turnover in the population. It is given by:

$$R_0 = \frac{\beta}{\gamma + \mu} \quad (1.3)$$

Here again, if  $R_0 > 1$ , the infection is expected to spread in the population, while  $R_0 < 1$  suggests that the infection will likely die out. More specifically, this measure is essential to determine the state of the model. When  $R_0 < 1$  we can say that the state of the disease will stay endemic. On the other hand, if the  $R_0 > 1$  we can say that the state of the disease will reach an epidemic state at some point in the simulation.

This demographic variant provides a more realistic framework for studying diseases over long periods, as it accounts for natural population changes through birth and death, enabling more accurate predictions in dynamic populations.

#### 1.1.3.3 SIR Model on Network

In traditional SIR models, the assumption is that the population is homogeneously mixed, meaning that every individual has an equal chance of interacting with every other individual [18]. This simplifies the model but doesn't reflect the more structured and localized interactions found in real-world networks. Implementing the SIR model on networks introduces a layer of realism by simulating disease spread through specific pathways, based on the network's topology [20]. In this approach, individuals are represented as nodes, and interactions between them are modelled as edges [23]. This structure captures local transmission dynamics, where only individuals connected by an edge can potentially transmit the infection. By accounting for heterogeneity in interactions, such as contact patterns or proximity, the network-based model better reflects

scenarios like schools, workplaces, or communities, where interaction patterns vary significantly [17].

For a node  $i$  to become infected, the probability depends on its neighbors. The probability that a susceptible node  $i$  will be infected depends on the state of its neighboring nodes and is given by:

$$P(\text{infected})_i = 1 - \prod_{j \in \text{neighbors}(i)} (1 - \beta A_{ij} \delta_{Ij})$$

where  $A_{ij}$  is the adjacency matrix, indicating whether nodes  $i$  and  $j$  are connected (1 if connected, 0 if not), and  $\delta_{Ij}$  is an indicator that takes the value 1 if node  $j$  is infected, and 0 otherwise [6]. This formula shows how the infection spreads only through connected nodes, making the network structure critical in determining transmission dynamics.

In network-based models, the basic reproduction number,  $R_0$ , also adapts to the network structure [6]. Unlike in homogeneous models, where  $R_0$  is a fixed average, in network models, it depends on both the average degree of nodes and the probability of transmission along edges. An approximate formula for  $R_0$  in network models is given by:

$$R_0 = \frac{\beta \langle k \rangle}{\gamma}$$

where:

- $\beta$  is the transmission rate, representing the probability of infection per contact.
- $\langle k \rangle$  is the average degree of the network, reflecting the average number of contacts per individual.
- $\gamma$  is the recovery rate.

This formula indicates that  $R_0$  is directly influenced by the network's connectivity: a higher average degree means a greater likelihood for an infected node to transmit the infection, whereas a higher recovery rate decreases  $R_0$ . In heterogeneous networks, like scale-free networks with hubs,  $R_0$  may be skewed due to highly connected nodes acting as super-spreaders [11]. As a result, in some cases,  $R_0$  is considered a distribution rather than a single value, to capture the varied transmission dynamics influenced by network structure [18].

The effectiveness of interventions, such as vaccination or quarantine, can also be more accurately modeled on networks [7]. In a well-mixed population, such strategies are often applied uniformly. In a network-based model, targeted interventions become possible, such as isolating highly connected individuals to break transmission chains or vaccinating clusters that are at higher risk of exposure due to their connectivity [8].

### 1.1.4 Different Types of Networks

In the study of complex networks, three predominant types are often discussed: Small World (WS), Scale-Free (BA), and Random (ER). Each of these networks has distinct characteristics and underlying theoretical foundations that make them suitable for different types of applications and analyses.

#### 1.1.4.1 Small World Networks (WS)

Small World networks, proposed by Watts and Strogatz in their 1998 paper [28], are characterized by relatively short paths between any two nodes, combined with a high degree of clustering. This type of network is typically constructed by starting with a regular lattice and randomly rewiring links with a small probability [28]. The resultant networks exhibit small average path lengths (like random networks) and high clustering coefficients (like regular lattices). Examples of Small World networks can be found in social networks, neural networks, and power grids, where the phenomenon of six degrees of separation often applies [19].

#### 1.1.4.2 Scale-Free Networks (BA)

Scale-Free networks, identified by Barabási and Albert [5], are defined by a power-law distribution in their degree connectivity. This means that a few nodes (hubs) have a very high number of connections, while the majority have very few. These networks evolve through mechanisms of growth and preferential attachment, where new nodes are more likely to connect to already well-connected nodes. This feature is common in many real-world networks, including the World Wide Web, citation networks, and some biological networks, reflecting robustness against random failures but vulnerability to targeted attacks Albert2000.

#### 1.1.4.3 Random Networks (ER)

Random Networks, conceptualized by Erdős and Rényi [10, 11], involve constructing a network where each pair of nodes is connected with a constant probability  $p$ . The primary characteristics of such networks are their lack of clustering and a binomial degree distribution, which simplifies to a Poisson distribution for large networks [11, 7]. These properties make Random Networks ideal for theoretical modeling and analytical tractability, although they are less likely to capture the complex structures found in real-world networks [1].

While all three types of networks are foundational in network theory, they differ significantly in their structural properties and implications for dynamics on the networks [21]. Small World networks are highly clustered with short path lengths, making them efficient for information transfer [28]. Scale-Free networks are dominated by hub nodes, leading to resilience against random node failures but susceptibility to attacks on major nodes [2]. Random Networks, with their simple connectivity model, serve as a baseline in network science, offering insights into the effects of randomness alone on network properties [11]. These differences have been noted in *Table 1.1*.

Network Type	Example Image	Key Characteristics	Degree Distribution	Example Applications
<i>Small World (WS)</i>		High clustering, short average path length	Varies, generally more uniform	Social networks, neural networks, power grids
<i>Scale-Free (BA)</i>		Few hubs with many connections, many nodes with few connections	Power-law	World Wide Web, citation networks, biological networks
<i>Random (ER)</i>		Random connections, low clustering	Poisson for large networks	Theoretical models, baseline comparisons in network science

Table 1.1: Comparison of Small World, Scale-Free, and Random Networks

#### 1.1.5 Network Statistics

In analyzing network structures, several critical statistics are used to capture the fundamental characteristics of connectivity, centrality, and clustering. These metrics provide insights into the topological properties of networks, which are essential for understanding how network structure can influence processes such as disease spread.

**Average Path Length:** The average path length is the mean of the shortest paths between all pairs of nodes within a network. It measures how efficiently information or influence can travel across the network. Mathematically, the average path length  $L$  of a network  $G$  with  $N$  nodes is

defined as:

$$L = \frac{1}{\frac{N(N-1)}{2}} \sum_{i < j} d_{ij}$$

where  $d_{ij}$  is the shortest path distance between nodes  $i$  and  $j$ . A shorter average path length can imply quicker dissemination of information or, in our case, faster disease transmission [22].

**Diameter:** The diameter of a network is the longest shortest path between any two nodes, providing a measure of the network’s “size” in terms of reachability. Formally, the diameter  $D$  of a network  $G$  is:

$$D = \max_{i,j} d_{ij}$$

where  $d_{ij}$  is the shortest path between nodes  $i$  and  $j$ . Networks with a small diameter facilitate rapid interactions or transmissions across distant parts of the network [7].

**Clustering Coefficient:** The clustering coefficient quantifies the tendency of nodes to form tightly-knit clusters or groups. For a given node  $i$ , the clustering coefficient  $C_i$  is defined as:

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

where  $e_i$  is the number of edges between the neighbors of  $i$ , and  $k_i$  is the degree (number of neighbors) of node  $i$ . The average clustering coefficient  $C$  for the network is:

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

A high clustering coefficient suggests strong local interconnectedness within the network [28].

**Betweenness Centrality:** Betweenness centrality measures the extent to which nodes lie on paths between other nodes. For a given node  $v$ , the betweenness centrality  $B(v)$  is calculated as:

$$B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where  $\sigma_{st}$  is the total number of shortest paths between nodes  $s$  and  $t$ , and  $\sigma_{st}(v)$  is the number of those paths that pass through  $v$ . High betweenness nodes can act as “bridges” within the network, facilitating disease transmission between otherwise unconnected regions [14].

**Closeness Centrality:** Closeness centrality calculates the average inverse distance from a node to all other nodes, reflecting how quickly a node can access the rest of the network. For a given node  $v$ , the closeness centrality  $C(v)$  is:

$$C(v) = \frac{N-1}{\sum_{i \neq v} d_{vi}}$$

where  $d_{vi}$  is the shortest path distance between  $v$  and  $i$ . Nodes with high closeness centrality can access other nodes quickly, making them efficient hubs for disease transmission [25].

Together, these network statistics provide a comprehensive toolkit for analyzing and understanding network structures, laying the groundwork for assessing their impact on disease spread dynamics.

### 1.1.6 Vaccination Strategies

Vaccination strategies within network-based SIR models are designed to assess their effectiveness in controlling disease spread by removing susceptible individuals from the infection pool. In this study, we compare two primary strategies: random and targeted vaccination.

#### 1.1.6.1 Random Vaccination Strategy

The random vaccination strategy serves as a baseline approach, applying vaccinations without prioritization. This method assumes an equal likelihood of any individual being vaccinated, regardless of their connectivity or role in the network structure. Random vaccination has historically been used in population-level immunization efforts, particularly when resources are limited or specific data on social contacts and network structure is unavailable [3, 13]. While less efficient than targeted methods, random vaccination provides a foundational comparison and insight into the minimum expected effectiveness in reducing disease spread.

Testing accuracy in this approach represents the probability of correctly identifying susceptible individuals, with an accuracy level of 1.0 indicating perfect identification. Varying this accuracy level allows exploration of how testing reliability influences the success of a non-prioritized approach.

#### 1.1.6.2 Targeted Vaccination Strategy

In contrast, the targeted vaccination strategy prioritizes individuals based on their degree, or number of connections, within the network. High-degree nodes—often called hubs—are more influential in disease transmission due to their greater number of contacts. This strategy has roots in targeted immunization campaigns historically used for diseases like smallpox, where health officials focused on vaccinating or isolating high-risk contacts to curb the spread [12, 24].

Targeted vaccination assumes that focusing on these highly connected individuals, or their immediate neighbors, can more effectively limit transmission by reducing potential “super-spreader” nodes. Studies on network immunization have shown that targeting central nodes can significantly reduce epidemic spread in various types of networks, highlighting its relevance in public health efforts where network data is available [2, 21].

## 1.2 Related work

In this sub-chapter, we on the academic works that serve as the foundation for the methodologies and theoretical approaches underpinning the research in this paper. While other sources were additionally valuable, their contributions were more focused or specific, and as such, they will be cited in the general reference list rather than discussed in detail here.

### 1.2.1 Stochastic SIR model including demography; birth and death rate

The following paper have highlighted the effectiveness of the Gillespie algorithm in epidemiological modelling. A notable contribution is the Temporal Gillespie Algorithm, introduced in [27]. This algorithm allows for efficient simulations of contagion dynamics on time-varying networks, significantly improving upon traditional methods. By accurately capturing the changing interactions among individuals over time, this approach is crucial for understanding disease spread in realistic scenarios. Their work demonstrates the algorithm’s applicability to both Poissonian and non-Markovian epidemic models, providing valuable insights into the time-dependent nature of disease transmission.

In our study, we build on these foundational concepts by implementing a variant of the Gillespie algorithm tailored for a rule-based SIR model framework. This integration allows us to examine the impact of dynamic interactions and varying transmission parameters on epidemic outcomes. The flexibility shown in the Temporal Gillespie Algorithm provides a valuable reference for enhancing the robustness of our simulations, particularly when considering the influence of time-varying networks on disease spread.

Another significant work is the Generalised Gillespie Algorithms for Simulations in a Rule-Based Epidemiological Model Framework [16], which expands the Gillespie algorithm’s application to complex epidemiological models that consider various factors such as age distribution and social interactions. This research enriches our understanding of how intricate contagion scenarios can

be simulated, which is directly relevant to our work. By incorporating these insights, we aim to analyse the interplay between network structures and disease dynamics more effectively, aligning with the goal of improving epidemic predictions in our model.

### 1.2.2 SIR & Networks

The following papers provide theoretical frameworks, practical insights, and foundational principles necessary for designing and analysing the network models and vaccination strategies explored through this paper.

(1) In *Spread of epidemic disease on networks*, M.E.J. Newman (2002) provides a detailed examination of how diseases spread across different network topologies. This foundational paper is critical to our research as it outlines the roles of various network structures in influencing epidemic dynamics. Newman's work is especially relevant in guiding our selection and analysis of the Barabasi Albert, Watts-Strogatz, and Erdos-Renyi networks. His insights into the thresholds and dynamics of disease spread across these networks form a theoretical backbone for our experimental simulations [20].

(2) *The effect of opinion clustering on disease outbreaks* by Salathé and Bonhoeffer (2008) explores the dynamics of how behavioral patterns, represented as opinion clustering, affect the spread of diseases and the efficacy of vaccination strategies. This paper provides a perspective relevant to our design of dynamic vaccination campaigns, where individual node behaviors significantly impact overall strategy effectiveness. The analogies drawn between opinion clustering and node behavior in networks offer a nuanced approach to designing our vaccination strategies within the simulated networks [26].

(3) *When individual behaviour matters: homogeneous and network models in epidemiology* by Bansal, Grenfell, and Meyers (2007) delves into the significant impact of individual node behavior within network models on the outcomes of epidemiological interventions like vaccinations. This study is pivotal for our research as it directly relates to our experiments with dynamic vaccination strategies. Bansal et al.'s analysis enhances our understanding of how individual decisions within a network can alter the course of disease spread and the effectiveness of controlled vaccination efforts [4].

These three texts collectively provide a framework for understanding the complexities of network-based disease modelling and the deployment of vaccinations in said models. Their theoretical and practical insights are crucial in shaping both the experimental design and the analytical framework of our exploration of the integration of SIR models with various network topologies.

## 1.3 Research Questions

This sub-chapter introduces the two main research questions of our study. While the topics we are investigating are closely related and interconnected, they are distinct enough to warrant these two separate research questions. This approach enables a clearer and more focused exploration of each topic. Here, we outline these questions and provide concise hypotheses for each, guiding the direction of our experiments and analysis.

### 1.3.1 Gillespie-Related Research Question

The first part of this report is guided by the following research question:

*How does the Gillespie method compare to standard ODE models, both including demography, and what phenomena emerge when applying the Gillespie method?*

This research question seeks to compare two methodologies for modeling the spread of infectious diseases using the SIR (Susceptible-Infected-Recovered) framework. One approach employs deterministic ordinary differential equations (ODEs) to predict the dynamics of the epidemic based on fixed infection and recovery rates. In contrast, the Gillespie algorithm offers a stochastic

perspective, incorporating randomness to yield more realistic simulations that can reflect the complexities of disease transmission in real life.

We hypothesize that the Gillespie method will reveal more intricate dynamics than deterministic models, particularly concerning the variability inherent in stochastic simulations. We will explore specific phenomena arising from the Gillespie method, including the analysis of variability such as negative covariance and standard deviation. These measurements will enhance our understanding of the unpredictability of disease spread, showcasing how stochastic elements can influence epidemic outcomes.

Additionally, we will investigate the phenomenon of stochastic resonance, where randomness enhances the system's response to weak signals, potentially affecting the behavior of disease transmission under specific conditions. This part of the research will also encompass experiments aimed at understanding extinction, or the conditions under which an outbreak dies out. By varying parameters like infection and recovery rates, we can examine the likelihood of persistence versus decline in outbreaks. Overall, this analysis will provide valuable insights into the dynamics of disease spread and the role of chance in epidemics, contributing to a comprehensive understanding of both deterministic and stochastic models in epidemic modeling.

### 1.3.2 Network-Related Research Question

This second part of this study is guided by the overarching research question:

*How do different network topologies influence the spread of infectious diseases modeled using SIR simulations, and what impact does a targeted vaccination strategy have on a real-world network?*

This question seeks to understand the role of certain network structures (i.e. Barabási-Albert, Watts-Strogatz, and Erdős-Rényi models) in shaping the dynamics of disease spread within SIR-based simulations. It also examines the effects of implementing a targeted vaccination strategy on a real-world contact network. This dual focus allows us to explore how variations in network topology can influence epidemic outcomes and assess the effectiveness of a targeted intervention within a realistic setting.

We hypothesize that networks with heterogeneous connectivity, such as the Barabási-Albert model, will facilitate faster and broader disease spread compared to more homogeneous networks like the Watts-Strogatz and Erdős-Rényi models. Additionally, we anticipate that a targeted vaccination strategy implemented on a real-world network will significantly curb this spread as opposed to more random strategies, demonstrating the effectiveness of strategic interventions in disease control.

The insights gained through answering this research question aim to provide both a deeper theoretical understanding of network influences on disease spread and practical implications for optimizing vaccination strategies in managing public health.

## 1.4 Report structure

This report is organized to provide a clear and logical exploration of the research questions posed. We begin with a section dedicated to background information (in this *current* chapter, where we establish the context and theoretical underpinnings of our study. This section is essential for understanding the complexities of the topics and frameworks that our research builds upon.

Following the background, we delve into the methodology section. Here, we detail the experimental and analytical procedures used to gather and analyse data. This section outlines the approach taken to address each research question, ensuring that the methods align with our stated objectives and hypotheses.

After methodology, the report presents the results of our experiments. This section is structured to clearly convey the data collected, emphasizing significant findings and patterns observed during the research process.

The discussion section follows, where we interpret the results in the context of our initial hypotheses and the broader field of study. This part of the report evaluates the implications of our findings, discusses potential limitations of our approach, and suggests areas for future research.

Finally, we conclude the report by summarizing the key outcomes and their implications. The conclusion reaffirms the value of the research and reflects on the contributions made to the existing body of knowledge.

Overall, this structured approach ensures that the report is accessible and informative, providing a comprehensive overview of the research from conception to conclusion.

---

## CHAPTER 2

# Method

---

This section will explain the specific methodology used to implement the aspects of this report. This also includes specific functions and python packages for reproducibility.

### 2.1 Gillespie Algorithm

As mentioned before in section 1.1.1, the Gillespie method is a stochastic algorithm to implement certain system dynamics including in this case the SIR model. For this report this algorithm has been implemented in Python and consists of the following six steps:

1. Initialize the system by setting the initial state of the system, including the number of susceptible, infected, and recovered individuals, as well as the reaction rates.
2. Calculate the propensity functions for each possible reaction. These functions represent the likelihood of each reaction occurring based on the current state of the system.
3. Compute the total propensity, which is the sum of all the propensity functions. This value will be used to determine the time until the next reaction.
4. Generate a random number  $r_1$  uniformly distributed between 0 and 1 to determine the time until the next reaction occurs using the formula:

$$\Delta t = \frac{1}{\sum a_i} \ln \left( \frac{1}{r_1} \right)$$

5. Generate a second random number  $r_2$  to select which reaction will occur. This is done by finding the smallest  $j$  such that:

$$\sum_{k=1}^j a_k \geq r_2 \cdot \sum_i a_i$$

6. Update the state of the system based on the selected reaction and the calculated time step. Repeat steps 2 to 6 until the desired simulation time is reached or a stopping condition is met.

Note that system updates consist only of moving individuals from one population to another. This is a discrete event which means that the values of all the different sub-populations will always be an integer value.

Before actually running the Gillespie algorithm, we first need to establish the initial conditions for the populations of susceptible (S), infected (I), and recovered (R) individuals. The standard initial values are set to  $S = 990$ ,  $I = 10$ , and  $R = 0$ . In addition to these initial conditions, we

must also define the rates that govern the dynamics of the model. The standard values for these rates are  $\beta = 0.3$  (the transmission rate),  $\gamma = 0.1$  (the recovery rate), and  $\mu = 0.01$  (the rate of natural death).

With the initial conditions and rates established, we can proceed to implement and execute the entire Gillespie algorithm, allowing us to simulate the dynamics of the SIR model effectively. The following pseudocode gives a good representation of the specific algorithm used to implement the Gillespie method:

---

**Algorithm 1** Gillespie Algorithm

---

```

1: function GILLESPIE(S, I, R, beta, gamma, mu, max_time)
2:    $t \leftarrow 0$ 
3:   Initialize output list
4:    $N_{total} \leftarrow S + I + R$ 
5:   while  $t < max\_time$  do
6:     Append  $[t, S, I, R]$  to output
7:     Calculate rates based on  $S, I, R, \beta, \gamma, \mu$ 
8:      $total\_rate \leftarrow$  Sum of all rates
9:     if  $total\_rate = 0$  then
10:      Break
11:    end if
12:     $time\_step \leftarrow$  RANDOM_TIME_STEP( $total\_rate$ )
13:     $t \leftarrow t + time\_step$ 
14:    Select a random event based on the rates
15:    Update  $S, I, R$  based on the selected event
16:   end while
17:   Append  $[t, S, I, R]$  to output
18:   return output
19: end function

```

---

## 2.2 SIR model with demography

As mentioned before in section 1.3.1, we will compare the Gillespie method with demography to the standard SIR model with demography. Therefore we also implemented the deterministic SIR model as a system of ordinary differential equations (ODEs) that includes birth and death rates. This deterministic approach provides a smoother, average behaviour for the epidemic dynamics over time.

For solving the system of differential equations, we use the `odeint` function from the `scipy.integrate` package in Python. The `odeint` function performs numerical integration by taking the system of differential equations, initial conditions, time array, and model parameters as inputs. It returns the values of  $S$ ,  $I$ , and  $R$  populations over time, allowing us to observe the overall trend of the epidemic without random fluctuations.

The pseudocode for this SIR model follows the exact mathematical definition as mentioned in section 1.1.3.2. The following pseudocode also repeats this for convenience:

---

**Algorithm 2** SIR Model with Demography

---

```

1: function SIR_MODEL(S, I, R, beta, gamma, mu)
2:    $dS/dt = \mu - (\beta \cdot S \cdot I) - (\mu \cdot S)$ 
3:    $dI/dt = (\beta \cdot S \cdot I) - (\gamma \cdot I) - (\mu \cdot I)$ 
4:    $dR/dt = (\gamma \cdot I) - (\mu \cdot R)$ 
5:   return  $[dS/dt, dI/dt, dR/dt]$ 
6: end function

```

---

In this model: -  $dS/dt$  calculates the rate of change in the susceptible population based on new

births, infections, and natural deaths.  $-dI/dt$  represents the change in the infected population due to new infections, recoveries, and deaths.  $-dR/dt$  is the change in the recovered population due to new recoveries and deaths.

This deterministic model serves as a baseline, showing how the population sizes of susceptible, infected, and recovered groups evolve over time without stochastic variation. By comparing the results of this model with those from the Gillespie algorithm, we can observe how randomness and discrete events impact the dynamics of the epidemic.

Lastly, to run the SIR model with demography, we use initial conditions similar to those in the Gillespie implementation. This ensures that the comparison between the deterministic and stochastic methods is fair and based on identical model parameters. Specifically, we initialize the values for  $S$ ,  $I$ , and  $R$  as 990/1000, 10/1000, and 0, respectively, to represent the initial susceptible, infected, and recovered populations as proportions of the total population.

The model parameters for the infection rate  $\beta$ , recovery rate  $\gamma$ , and birth/death rate  $\mu$  are set to 0.3, 0.1, and 0.01, respectively. These values provide a basis for comparing the dynamics of both methods, allowing us to fairly compare the impact of stochasticity in the Gillespie algorithm compared to the baseline SIR model.

## 2.3 Experiments for Gillespie-Related Research Question

This section includes the explanation, motivation and parameter settings for the experiments that we implemented to answer the first research question of this report.

### 2.3.1 Implementation and comparison of Gillespie's algorithm

In this experiment, we compare the behaviour of the deterministic SIR model with demography and the stochastic Gillespie model with demography to assess how each method represents epidemic dynamics under similar parameter conditions. Both models share identical values for the infection rate ( $\beta$ ), recovery rate ( $\gamma$ ), and birth/death rate ( $\mu$ ), allowing us to examine differences arising solely from the deterministic versus stochastic implementations.

#### 2.3.1.1 Endemic and Epidemic Experiment

Additionally, we also use two different parameter sets to show the difference between endemic ( $R_0 < 1$ ) and epidemic ( $R_0 > 1$ ) scenarios by running the Gillespie model with the following parameters:

- Epidemic scenario:  $\beta = 0.3$ ,  $\gamma = 0.1$ , and  $\mu = 0.01$ , resulting in  $R_0 > 1$ . - Endemic scenario:  $\beta = 0.05$ ,  $\gamma = 0.1$ , and  $\mu = 0.01$ , resulting in  $R_0 < 1$ .

Note that the starting values for  $S$ ,  $I$ , and  $R$  remain the same as in the first experiment:  $(S, I, R) = (990, 10, 0)$ . These two parameter sets help us observe how changes in infection rate affect population dynamics. In the epidemic scenario ( $R_0 > 1$ ), the infection rate  $\beta$  is high enough to sustain infection growth, leading to a larger outbreak. In contrast, the endemic scenario ( $R_0 < 1$ ) has a lower infection rate, causing the infection levels to gradually decline until the disease dies out. Running both parameter sets in the Gillespie model clearly shows these differences, demonstrating the model's response to  $R_0$  and its flexibility for various epidemiological situations.

#### 2.3.1.2 Simple Comparison

The initial conditions differ slightly to reflect the way each model handles population sizes. For the Gillespie model, which operates on absolute counts, we set the initial populations of susceptible, infected, and recovered individuals to 990, 10, and 0, respectively, ensuring that the total population remains consistent at 1000. For the deterministic SIR model, we use proportional values to represent the population:  $S = 0.99$ ,  $I = 0.01$ , and  $R = 0.0$ , resulting in  $S + I + R = 1$ . This scaling allows the deterministic model to reflect relative population fractions rather than absolute numbers.

The choice of a population size of 1000 for the Gillespie model strikes a balance between capturing stochastic fluctuations and maintaining comparable trends with the deterministic SIR model. A larger population would reduce the stochastic noise in Gillespie's method, making it closely resemble the deterministic behaviour. However, with a population of 1000, we achieve noticeable stochastic variability, while the overall trajectories of the susceptible, infected, and recovered populations remain similar to those in the deterministic model. This population size allows us to observe the inherent randomness of the Gillespie model without diverging too drastically from the expected epidemic dynamics.

#### 2.3.1.3 N runs comparison

Lastly, to enable a meaningful side-by-side comparison between these two results, we scale the output of the deterministic SIR model to match the scale of the Gillespie model. This is achieved by multiplying the deterministic results by 1000, so that the total population size aligns with that of the Gillespie model. We then plot the results from both models side by side, making it easier to visually compare the dynamics, observe the effects of stochastic variation, and analyse differences in population trends over time.

#### 2.3.1.4 Noise Level Control

The choice of a population size of 1000 for the Gillespie model balances the two sides of the stochastic model. It maintains comparable trends with the deterministic model while also inducing visible noise in the trends. A larger population would reduce the stochastic noise in Gillespie's method, making it closely resemble the deterministic behaviour. However, with a population of 1000, we achieve noticeable stochastic variability, while the overall trajectories of the susceptible, infected, and recovered populations remain similar to those in the deterministic model. This population size allows us to observe the inherent randomness of the Gillespie model without diverging too drastically from the expected epidemic dynamics.

### 2.3.2 Investigate Simulation Variability and Negative Co-variance

In this code, an investigation is conducted into the effects of varying model parameters on the stochastic dynamics of an SIR epidemiological model. The primary focus of the experiment is to analyse how changes in the infection rate ( $\beta$ ), recovery rate ( $\gamma$ ), and birth/death rate ( $\mu$ ) influence the variability of the stochastic simulations, particularly in terms of variance between runs and the negative covariance between the susceptible and infected populations. The comparison between the stochastic outcomes and the deterministic model serves to highlight the differences in behaviour between these two approaches to modelling disease dynamics.

Initially, the code sets up initial conditions for both deterministic and stochastic simulations. The populations of susceptible, infected, and recovered individuals are defined proportionally for the deterministic model and in absolute numbers for the Gillespie simulation. A scale factor is applied to ensure that both models operate on comparable population sizes. The experiment includes a series of different parameter sets, allowing for the examination of a range of infection, recovery, and birth/death rates. Each set is run multiple times (50 stochastic simulations) to capture the inherent variability associated with the Gillespie method.

The simulations involve running the deterministic SIR model using the `odeint` function from SciPy, while the stochastic simulations employ a Gillespie algorithm. For each parameter set, results from both models are compared. The mean outcomes from the stochastic simulations are calculated and interpolated for better comparison with the deterministic model's output. The experiment further quantifies the variability by computing the variances for susceptible and infected populations across all stochastic runs. Additionally, the covariance between these two groups is analysed to understand the dynamics of disease spread better, particularly how increases in one population may correlate with decreases in the other.

Finally, the results are plotted to visualize the differences between the deterministic and stochastic models. The first subplot illustrates the populations of susceptible, infected, and recovered

individuals over time for both approaches, while the second subplot depicts the variance in the susceptible and infected populations across the stochastic simulations. This comprehensive analysis provides insights into how parameter variations affect the epidemic dynamics, emphasizing the importance of stochastic processes in modelling disease spread and the implications for public health interventions.

### 2.3.3 Stochastic resonance and increased transients

#### 2.3.3.1 Stochastic resonance general

This experiment utilizes a Susceptible-Infected-Recovered (SIR) model to analyse disease spread, comparing deterministic and stochastic approaches. The model parameters are set as follows: transmission rate ( $\beta = 0.2$ ), recovery rate ( $\gamma = 0.1$ ), and natural death rate ( $\mu = 0.01$ ). The initial populations are defined as  $S_0 = 0.99$ ,  $I_0 = 0.01$ , and  $R_0 = 0.0$ .

The deterministic equilibrium is calculated to identify stable populations of susceptible ( $S$ ), infected ( $I$ ), and recovered ( $R$ ) individuals. Subsequently, 50 stochastic simulations are performed using the Gillespie algorithm, which incorporates randomness into the disease spread. The outcomes of these simulations are interpolated for comparison with the deterministic results. This analysis helps to explore variability in infection dynamics and evaluates how accurately the deterministic model represents real-world disease transmission.

#### 2.3.3.2 Model parameters and resonance and examples

This experiment investigates the effect of varying key parameters on the variability of the infected population in a stochastic SIR model, using the Gillespie algorithm to capture stochastic dynamics. The infection rate ( $\beta$ ), recovery rate ( $\mu$ ), and population scaling factor ( $N_{\text{scale}}$ ) are varied independently, with 60 stochastic simulations run for each parameter setting. The ranges for each parameter are as follows:

- Infection rate,  $\beta$ : 0.2 to 0.5
- Recovery rate,  $\mu$ : 0.005 to 0.05
- Population scaling factor,  $N_{\text{scale}}$ : 0.5 to 2.0

For each varied parameter, the variance and standard deviation of the infected population are calculated over time, starting from a specified point ( $t_{\text{start}} = 60$ ). Two types of plots are generated:

1. Sample trajectories for the infected population across selected parameter values to illustrate individual stochastic paths.
2. Average variance of the infected population with standard deviation error bands to visualize overall variability and the influence of parameter changes.

These results provide insights into how each parameter impacts the epidemic variability in a stochastic setting.

#### 2.3.3.3 Increased transients

This experiment examines how changes in the infection rate parameter,  $\beta$ , affect transient peaks in the infected population using a stochastic SIR model with the Gillespie algorithm. The transients (deviations from equilibrium) are recorded for each  $\beta$  value, and their averages are plotted to understand  $\beta$ 's impact on infection dynamics.

## Parameter Settings

- **Population Scale Factor:** 1000

- **Initial Populations (Susceptible, Infected, Recovered):**  $S_0 = 0.99 \times \text{scale\_factor}$ ,  $I_0 = 0.01 \times \text{scale\_factor}$ ,  $R_0 = 0$
- **Max Time:** 100
- **Simulations per  $\beta$ :** 10
- **Fixed Parameters:**  $\gamma = 0.1$ ,  $\mu = 0.01$
- **Varying  $\beta$  Values:** 0.1 to 0.6

### 2.3.4 Extinction events and Critical Community Size

We implement the SIR model using a set of differential equations representing the rates of change for susceptible (S), infected (I), and recovered (R) individuals within a closed population of size  $N$ . The model is governed by the parameters  $\beta$  (transmission rate) and  $\gamma$  (recovery rate), with  $R_0 = \frac{\beta}{\gamma}$ .

We explore the dynamics under different  $R_0$  values to observe how they affect the trajectory of the disease and the likelihood of extinction. The initial conditions include a single infected individual, with the rest of the population susceptible.

To effectively communicate the results of our stochastic SIR model simulations, we will present a series of graphs that illustrate the relationship between the basic reproduction number  $R_0$ , population size  $N$ , and the probability of disease extinction within a closed population.

Each graph will plot population size on the x-axis on a logarithmic scale, reflecting the range from small communities to large, dense populations. The y-axis will represent the probability of the disease going extinct within the simulation period, providing a clear visualization of how extinction likelihood changes with community size. Multiple curves will be displayed on each graph, with each curve corresponding to a different  $R_0$  value. This will allow for direct comparison across different scenarios of disease contagiousness. The values for  $R_0$  will range from values less than 1 (where disease is expected to die out quickly) to values significantly greater than 1 (where disease is typically more resilient and spreads more aggressively). The graph has the following features:

- **Color Coding:** Each  $R_0$  curve will be color-coded to enhance readability and ease of comparison.
- **Markers:** Specific markers will be used to denote data points for each  $R_0$  value, helping to distinguish between the curves visually.
- **Legend:** A legend will be included to identify the  $R_0$  values associated with each curve, ensuring that the graph is informative and self-explanatory.
- **Grid Lines:** Grid lines will be added to improve the accuracy of reading the exact values of extinction probabilities and population sizes from the graph.

These graphs will serve as a pivotal tool for understanding how the size of a population and the basic reproductive number interact to influence the dynamics of disease spread and extinction. Specifically, they will help in identifying critical thresholds for  $R_0$  and  $N$  that might dictate different public health strategies aimed at disease eradication or management in closed populations.

## 2.4 SIR Model on Networks

In our methodology, we integrated the traditional SIR model within a network-based framework to simulate the dynamics of infectious disease spread across both real-world and synthetic networks. This approach allowed us to capture the complexities of real-world interactions more accurately than the homogeneously mixed assumptions typically used in simpler SIR models, while also enabling a comparison of disease dynamics across different network structures.

We began by acquiring sociopatterns data (provided to us by the *University of Amsterdam*), which records the non-weighted interactions within a community, presented in a CSV file format. Utilizing Python's `pandas` library, we loaded this data to parse individual interactions

and construct a graph using `NetworkX`. This graph represented individuals as nodes and their interactions as edges, directly reflecting the actual interaction patterns documented in the data.

In addition to the real-data network, we generated three synthetic network types for comparative analysis: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER) networks. These networks, created with `NetworkX` using specific parameters (discussed in *Section 1.1.4*), enabled us to observe how varying network topologies could influence disease spread.

For the implementation of the SIR model on both the real-data and synthetic networks, we initialized all nodes in the susceptible state except for a few randomly selected nodes, which were set as infected to simulate the outbreak's inception. The progression of the model was then handled through a series of discrete time steps, during which each node's state could change according to the following rules:

- **Infection process:** A susceptible node would become infected if it was connected to an infected node, with the probability of infection governed by the parameter  $\beta$ , representing the infection rate.
- **Recovery process:** Infected nodes had a probability  $\gamma$  of recovering during each time step, representing the recovery rate.

The core of our simulation involved iterating through these steps and updating each node's state based on its interactions and the model's rules. To ensure a robust representation of the dynamics, we averaged the results over 100 simulation runs for each network type, as well as for the real-data network. This averaging allowed us to capture the true spread dynamics more accurately and minimized the influence of any single anomalous run. To illustrate the model's behavior, we provide two forms of visualization: network graphs and an SIR curve. The network graphs display snapshots of the disease spread, offering a visual understanding of how infection propagates across each network type over time. In these graphs, each node is color-coded based on its state: green for susceptible, red for infected, and blue for recovered. These visualizations help in capturing the spatial distribution of infections and highlight any clustering effects influenced by network structure.

For clarity, we focus on five key stages in the progression of the disease, each represented by a graph snapshot:

- **Initialization:** The first time step, where only a few randomly selected nodes are infected, representing the outbreak's inception.
- **Growth:** An early stage where the infected count is rising, indicating the disease's spread through the network.
- **Peak:** The time step at which the number of infected nodes reaches its maximum, marking the height of the outbreak.
- **Dying Down:** After the peak, when the infected count is noticeably decreasing as more nodes transition to the recovered state.
- **Extinction:** The final stage, when there are no infected nodes left, signifying the outbreak's resolution.

In addition to these snapshots, the SIR curve provides an analytical view of the epidemic progression by plotting the number of susceptible, infected, and recovered individuals over time. This combination of visual and quantitative analysis allows us to track both the spatial and temporal dynamics of disease spread across different network structures.

By combining these visual and analytical perspectives, we aimed to achieve a comprehensive understanding of the epidemic's progression on both real and synthetic networked populations, allowing us to assess the effects of different network structures and initial conditions. This framework highlights the role of network topology in influencing disease dynamics and offers insights into potential intervention strategies based on network characteristics.

## 2.5 Different Types of Networks

We implemented the three types of networks (previously discussed in *Chapter 1.1.4*) — Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER) — using Python’s `NetworkX` library. For each network type, we conducted an experiment by generating 100 instances of networks with  $n = 100$  nodes. This sample size allowed us to observe and analyze the variability within each network structure and assess how these differences impact disease dynamics.

To extract key characteristics from each network, we designed a custom function, `calculate_network_statistics`. This function processes each network instance to compute and store a set of metrics (previously discussed in *Chapter 1.1.5*): Average Path Length, Diameter, Clustering Coefficient, Average Betweenness Centrality, and Average Closeness Centrality.

The function follows these steps for each instance:

- **Connectivity Check:** It first checks if the network is fully connected, which affects the calculation of some metrics (such as average path length and diameter). If the network is not connected, the function identifies the largest connected component, allowing those metrics to be calculated within a subgraph representative of the network’s primary structure.
- **Metric Calculations:**
  - **Average Path Length and Diameter:** For connected networks, these metrics are computed directly; for unconnected networks, they are calculated on the largest component. The function uses `NetworkX`’s built-in functions to efficiently handle these calculations.
  - **Clustering Coefficient:** This global metric is calculated using `NetworkX`’s `average_clustering` function, giving insight into the network’s tendency to form node clusters.
  - **Betweenness and Closeness Centralities:** The function computes these values for all nodes, then averages them across the network to yield summary statistics. `NetworkX`’s centrality functions facilitate these calculations, aggregating individual centralities to form network-level averages.

By iterating over 100 instances per network type and applying this function, we generated a comprehensive dataset of network statistics for each model type. This dataset captures structural tendencies and variability within each network model.

To further analyze the impact of network structure on key metrics, we plotted the average values of these statistics against different population sizes: 100, 200, 300, and 400. This procedure involved generating the networks for each size, calculating the relevant statistics, and then visualizing the results.

The histograms help illustrate how network characteristics change with population size, providing a basis for understanding the dynamics of each network type without delving into specific findings at this stage. This methodological approach establishes a reliable framework for analyzing the role of different network structures in subsequent SIR model simulations.

## 2.6 Vaccination Strategies

In this study, we designed an experiment to evaluate and compare targeted and random vaccination strategies within an SIR model framework on the Sociopatterns dataset (provided to us by *University of Amsterdam*), a real-world contact network. Our objective was to assess the efficacy of different vaccination approaches in controlling disease spread by selectively vaccinating susceptible individuals based on varying prioritization criteria.

We imported this unweighted edgelist, with nodes symbolizing individuals and edges representing interactions, representing a community’s interaction patterns as a network, into `NetworkX`, treat-

ing each contact as equally likely to transmit disease. Disease spread was initiated by randomly selecting five infected nodes at the start of the outbreak.

To evaluate the dynamics under different transmission conditions, we simulated two scenarios with distinct parameter values for infection ( $\beta$ ) and recovery ( $\gamma$ ) rates. The network structure consists of 374 nodes and 1,265 edges, giving an average degree  $\langle k \rangle$  of:

$$\langle k \rangle = \frac{2 \times \text{edges}}{\text{nodes}} = \frac{2 \times 1265}{374} \approx 6.76$$

Given this average degree, we selected parameters for each scenario as follows:

- **Epidemic Scenario:** For this scenario, we set  $\beta = 0.1$  and  $\gamma = 0.02$ . Calculating  $R_0$  using the formula  $R_0 = \frac{\beta \langle k \rangle}{\gamma}$ :

$$R_0 = \frac{0.1 \times 6.76}{0.02} \approx 33.8$$

This high  $R_0$  suggests a strong likelihood of rapid and widespread infection, modeling a scenario where the disease can propagate extensively across the network.

- **Non-Epidemic Scenario:** Here, we set  $\beta = 0.01$  and  $\gamma = 0.1$ . Using the same formula for  $R_0$ :

$$R_0 = \frac{0.01 \times 6.76}{0.1} \approx 0.676$$

With  $R_0 < 1$ , this configuration models a scenario where the disease is unlikely to spread widely, simulating a situation where infections naturally die out due to higher recovery relative to transmission.

In these experiments, we considered the following key parameters:

- **Testing Budget:** A total of 200 tests was available, with a variable number of tests applied per iteration.
- **Vaccination Criteria:** Vaccinations were limited to susceptible individuals, who immediately moved to the removed state upon vaccination.
- **Testing Accuracy:** We conducted experiments with varying testing accuracy levels at 1.0, 0.75, and 0.5.
- **Vaccination Budget:** We adjusted the per-iteration vaccination budget to {1, 3, 5, 10}, allowing for comparison between resource-constrained and resource-abundant scenarios.

The progression of disease spread under each vaccination strategy was visualized through SIR graphs and state plots over time. Each experiment was run for 50 time steps, with the number of susceptible, infected, and recovered nodes recorded at each step. In the epidemic scenario, we expected a significant increase in infections early on, with peaks at higher infection counts, allowing us to evaluate the effectiveness of targeted and random vaccinations in reducing this peak. In the non-epidemic scenario, we anticipated a gradual die-off of infections, providing a baseline to assess how each vaccination strategy performs in containing low-level disease spread.

To ensure robustness, results were averaged over 100 experimental runs. The analysis focuses on comparing the reduction in infection spread across varying vaccination budgets and testing accuracies. Outcomes from each strategy were averaged and then visualized to highlight any differences in effectiveness between targeted and random vaccination approaches.

---

## CHAPTER 3

# Results

---

In this section, we present the findings from our experiments evaluating the effectiveness of targeted and random vaccination strategies within the SIR model framework on the Sociopatterns dataset. The results are illustrated through a series of graphs that depict the dynamics of disease spread across different vaccination strategies and conditions. Each graph is accompanied by a brief description highlighting the key features of the data presented, setting the stage for subsequent analysis and interpretation of the findings.

### 3.1 Gillespie

This section will include the results of all the experiments done on the Gillespie implementation.

#### 3.1.1 Implementation Gillespie

##### 3.1.1.1 endemic vs epidemic

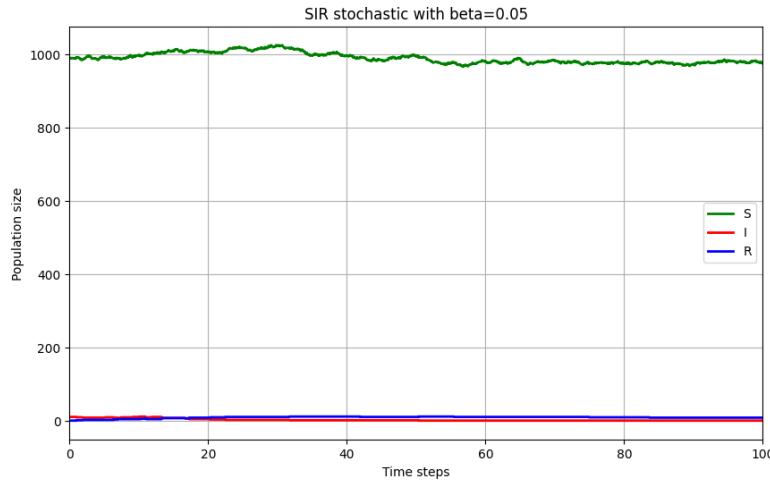


Figure 3.1: Endemic Gillespie phase plot with  $(S, I, R) = (990, 10, 0)$  and  $(\beta, \gamma, \mu) = (0.05, 0.1, 0.01)$

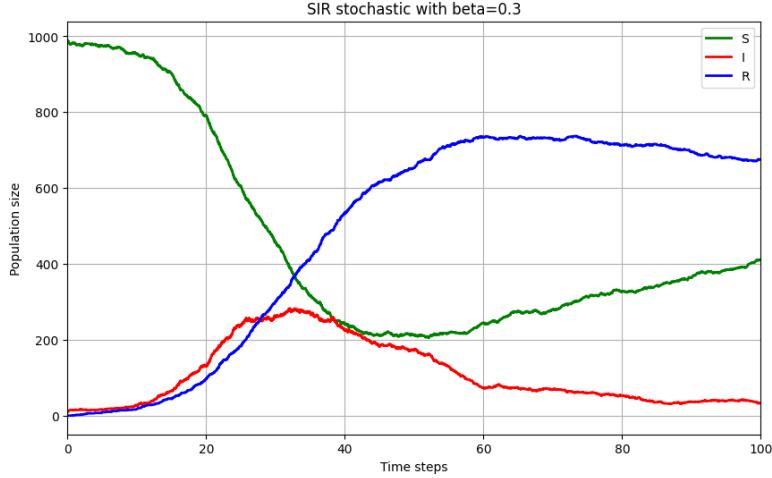


Figure 3.2: Endemic Gillespie phase plot with  $(S, I, R) = (990, 10, 0)$  and  $(\beta, \gamma, \mu) = (0.3, 0.1, 0.01)$

The figures 3.1 and 3.2 illustrate two distinct states of the disease modelled using the SIR framework with Gillespie's algorithm. Notably, the endemic plot shown in figure 3.1 reveals that the population size can exceed the initial population size.

### 3.1.1.2 Comparison deterministic and stochastic

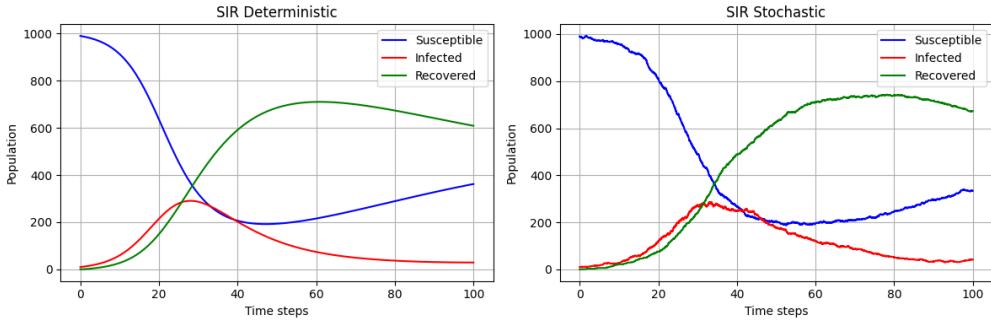


Figure 3.3: Comparison between stochastic (Gillespie) and deterministic (ODE) with ODE  $(S, I, R) = (990/1000, 10/1000, 0)$  and Gillespie  $(S, I, R) = (990, 10, 0)$  and  $(\beta, \gamma, \mu) = (0.3, 0.1, 0.01)$

As shown in the figure 3.3, we can see the noise effect in the stochastic SIR phase plot induced by Gillespie's algorithm.

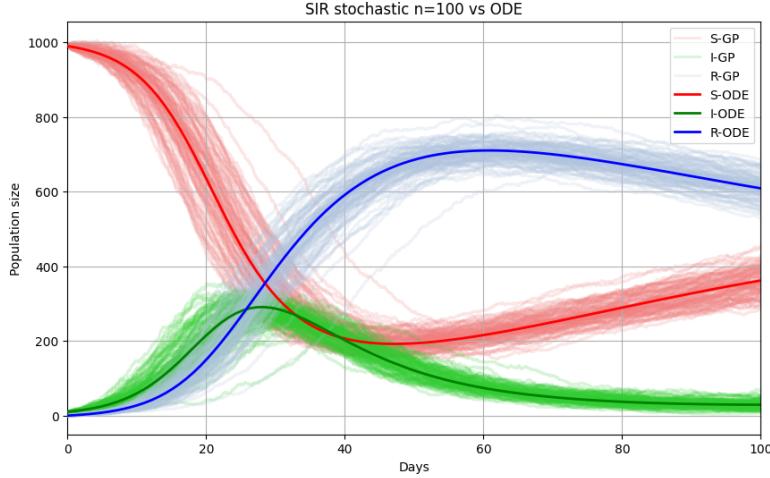


Figure 3.4: Comparison between stochastic (Gillespie) and deterministic (ODE) with ODE ( $S, I, R$ ) = (990/1000, 10/1000, 0) and Gillespie ( $S, I, R$ ) = (990, 10, 0) and  $(\beta, \gamma, \mu) = (0.3, 0.1, 0.01)$  across 100 Gillespie runs

Figure 3.4, shows that the Gillespie implementation varies around the deterministic ODE, with some extreme variations in the 100 runs.

### 3.1.1.3 Noise level

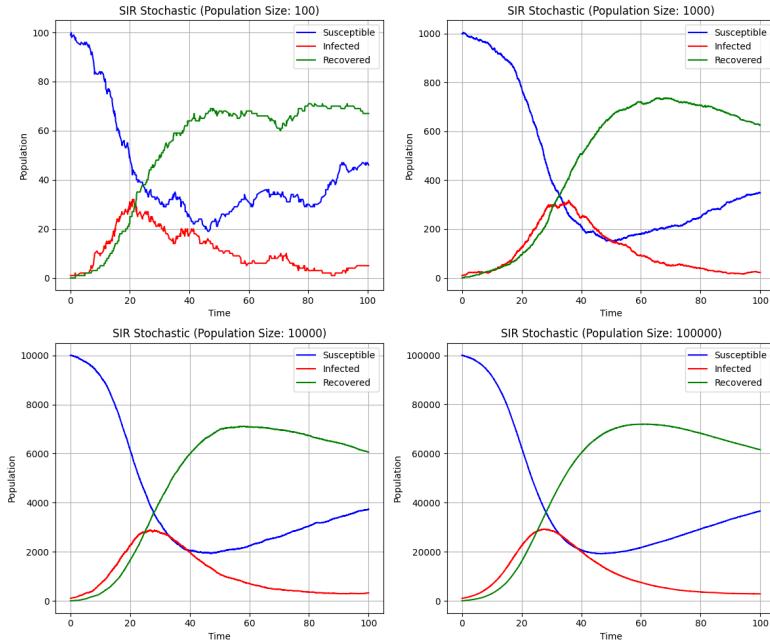


Figure 3.5: Gillespie with differing population size  $S = [10^2 - 1\%, 10^3 - 1\%, 10^4 - 1\%, 10^5 - 1\%]$ ,  $I = S_I/100$ , and  $R = 0$ . The parameters for both models are  $\beta = 0.3$ ,  $\gamma = 0.1$ , and  $\mu = 0.01$

### 3.1.2 Investigate Simulation Variability and Negative Co-variance

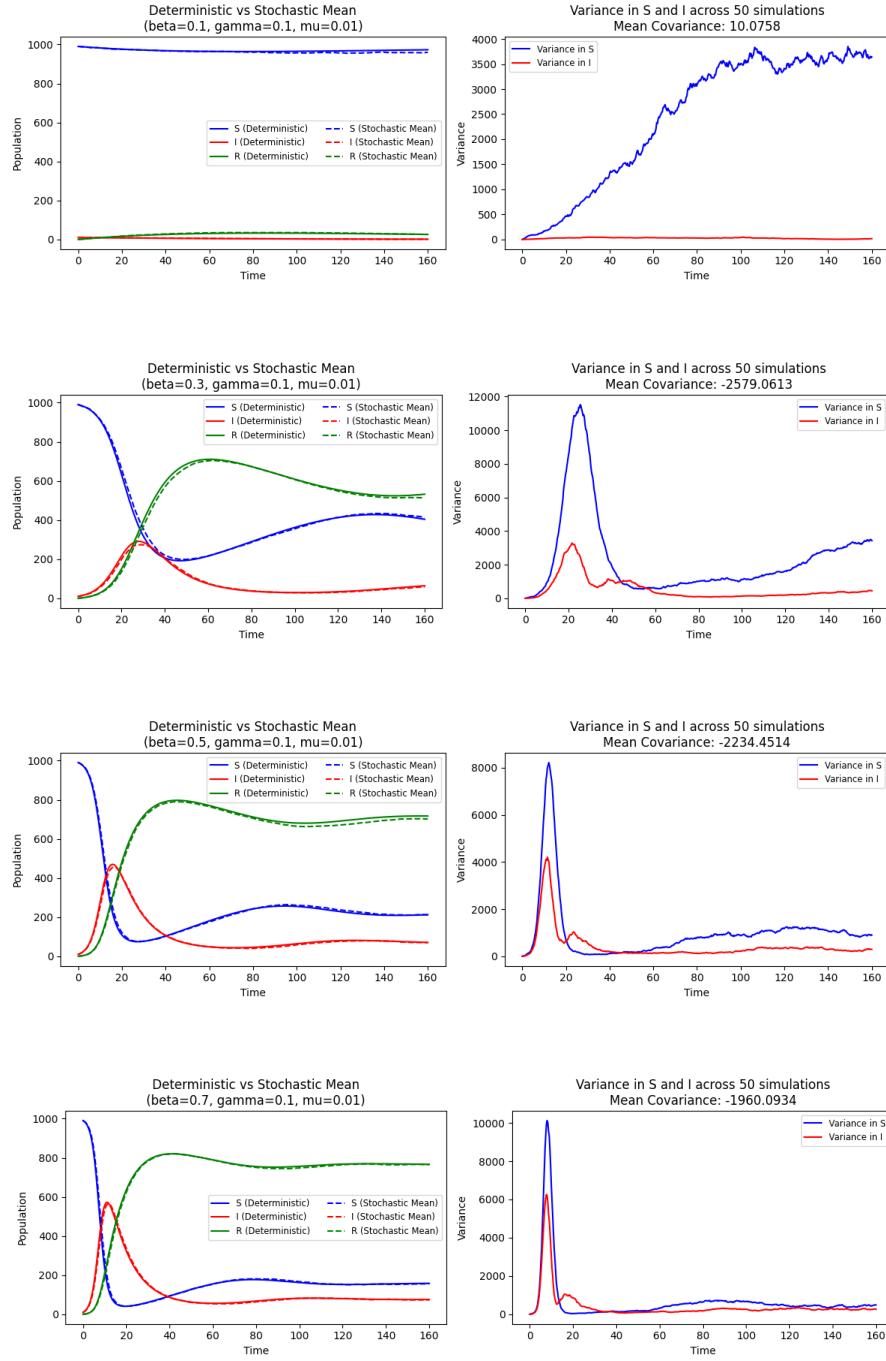


Figure 3.6: Comparison of deterministic and stochastic SIR model dynamics across varying parameter sets. The first column displays the population trajectories of susceptible (S), infected (I), and recovered (R) individuals for each parameter set, illustrating the differences between deterministic (solid lines) and stochastic (dashed lines) outcomes. The second column depicts the variance in susceptible and infected populations across multiple stochastic simulations, highlighting the variability in epidemic dynamics. The parameters used are: (a)  $\beta = 0.1$ ,  $\gamma = 0.1$ ,  $\mu = 0.01$ ; (b)  $\beta = 0.3$ ,  $\gamma = 0.1$ ,  $\mu = 0.01$ ; (c)  $\beta = 0.5$ ,  $\gamma = 0.1$ ,  $\mu = 0.01$ ; and (d)  $\beta = 0.7$ ,  $\gamma = 0.1$ ,  $\mu = 0.01$ . These plots collectively demonstrate how changes in infection and recovery rates affect the epidemic's trajectory and the stochastic variability inherent in disease dynamics.

Figure 3.6 shows the changing dynamics when increasing the value for  $\beta$ . The most notable about the changing  $\beta$  value is the can see a downwards trend in the mean covariance. Additionally, we also notice an increasing peak in the number of infected.

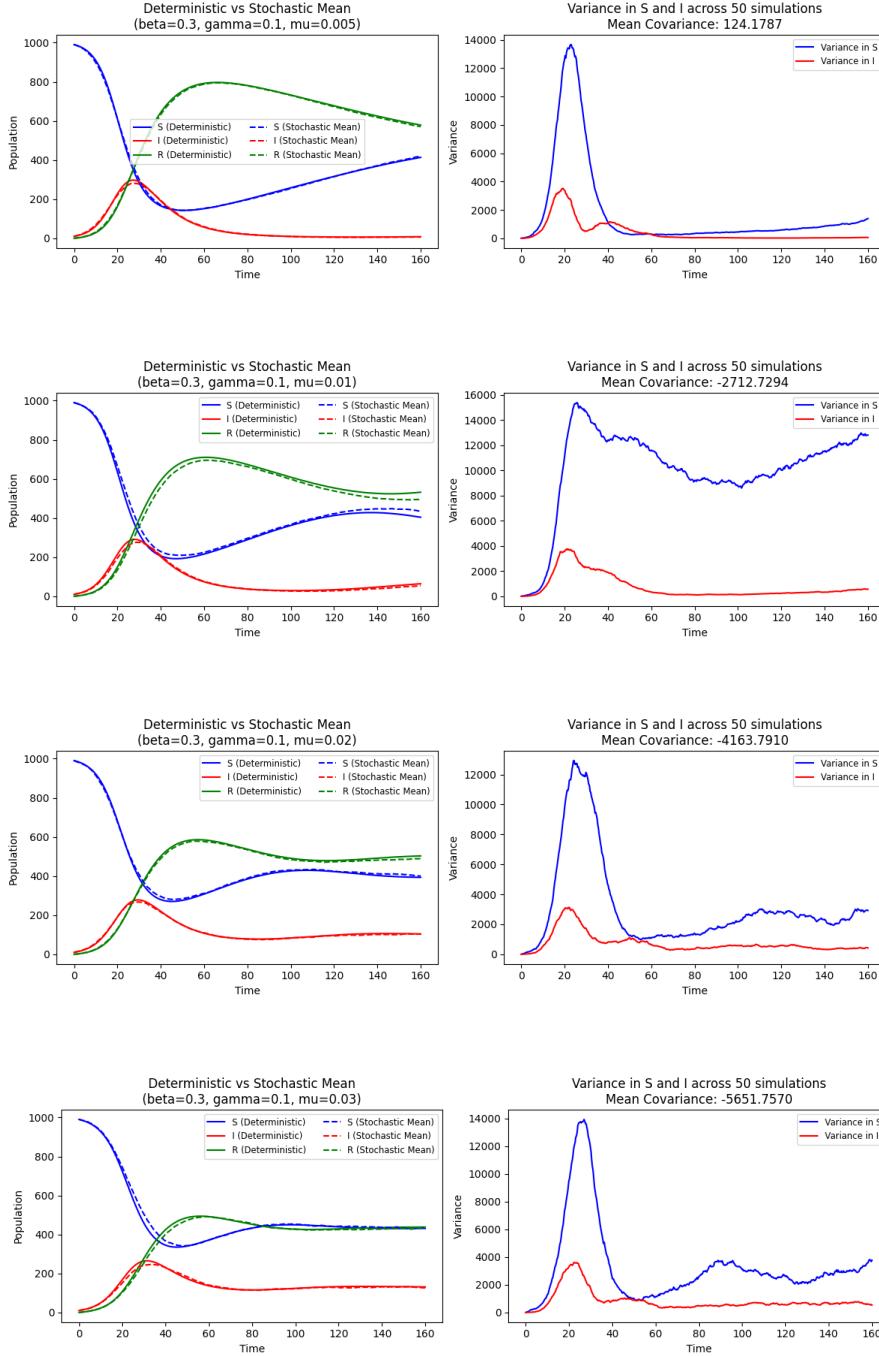


Figure 3.7: Comparison of deterministic and stochastic SIR model dynamics with varying death/birth rates ( $\mu$ ). The first column illustrates the population trajectories of susceptible (S), infected (I), and recovered (R) individuals for each parameter set, comparing the deterministic (solid lines) and stochastic (dashed lines) outcomes. The second column presents the variance in susceptible and infected populations across multiple stochastic simulations, emphasizing the variability in epidemic dynamics due to changes in  $\mu$ . The parameters used are: (a)  $\beta = 0.3, \gamma = 0.1, \mu = 0.005$ ; (b)  $\beta = 0.3, \gamma = 0.1, \mu = 0.01$ ; (c)  $\beta = 0.3, \gamma = 0.1, \mu = 0.02$ ; and (d)  $\beta = 0.3, \gamma = 0.1, \mu = 0.03$ . This set of plots illustrates how varying the birth/death rate impacts the epidemic's trajectory and highlights the stochastic variability in disease dynamics.

Figure 3.7 shows the difference in dynamics when changing the  $\mu$  value (death/birth rate). The notable trend that can be seen from this figure is the (absolute) upward trend in negative

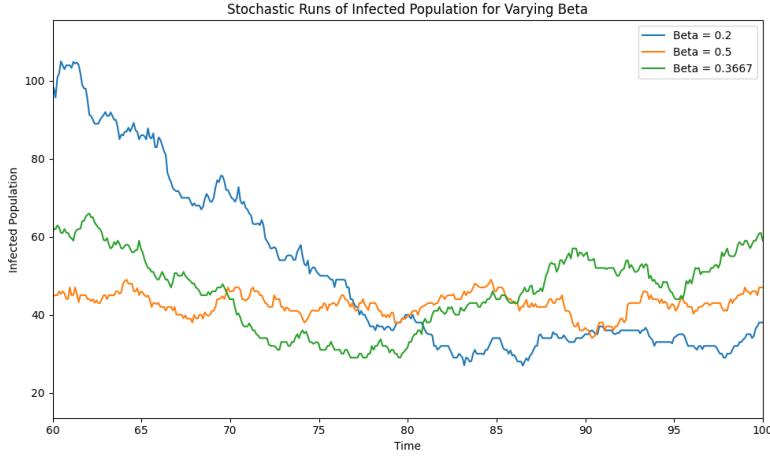


Figure 3.9: Shows 3 (first middle and last) runs of the 60 runs for the different parameter values for  $\beta$  from  $t=60$ ,  $t_{\max}$ ; examples of increased transients

covariance. Where instead of having a decreasing covariance, we have in increasing negative covariance. Additionally, the infected peak seems to settle in at a higher point.

## 3.2 Stochastic resonance and increased transients

### 3.2.1 Stochastic resonance general

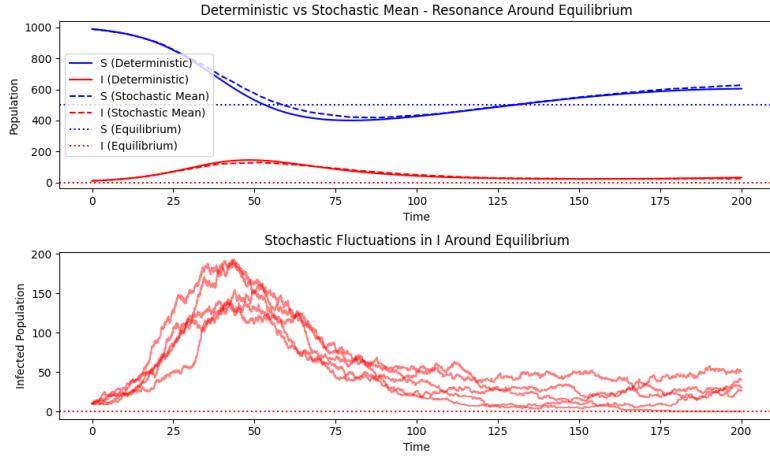


Figure 3.8: Comparison of deterministic and stochastic SIR model outcomes with parameters  $\beta = 0.2$ ,  $\gamma = 0.1$ ,  $\mu = 0.01$ , and initial populations  $S_0 = 0.99$ ,  $I_0 = 0.01$ ,  $R_0 = 0.0$ . The deterministic solution is shown alongside the mean of 50 stochastic simulations (Gillespie algorithm), demonstrating variability in infection dynamics and convergence toward the deterministic equilibrium.

Figure 3.8, shows the stochastic resonance of 50 runs. The resonance seems to be slightly over the equilibrium point. However, the resonance does seem to appear.

### 3.2.2 Model parameters and resonance and examples

Figure 3.9 shows three different examples of infection population for the given beta values. There does not seem to be a clear trend in the varying beta values.

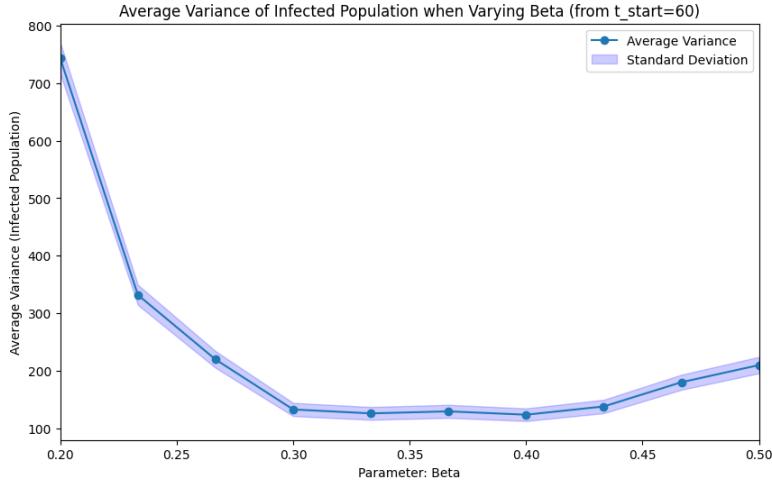


Figure 3.10: Shows the average variance and std of the variance over 60 runs for the different beta values

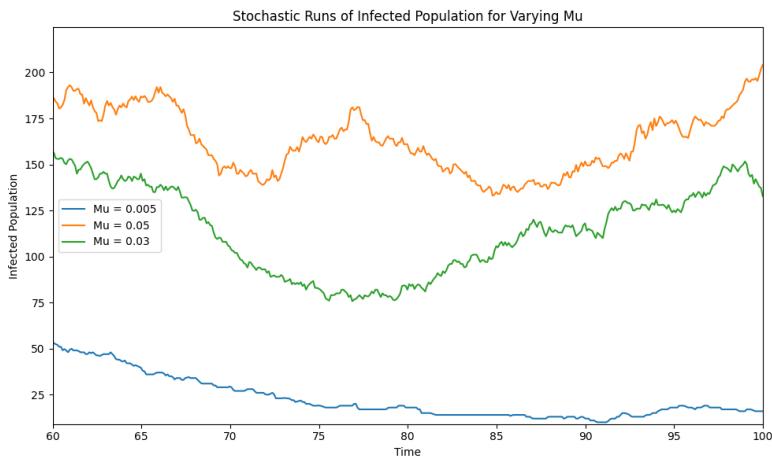


Figure 3.11: Shows 3 (first middle and last) runs of the 60 runs for the different parameter values for  $\mu$  from  $t=60$ ,  $t_{max}$ ; examples of increased transients

Figure 3.10, shows the average variance and standard deviation over 60 runs for the differing beta values. We do clearly notice a very quick downfall in the variance in the first few beta values. Then the averages seem to stagnate before slightly picking up again after  $\beta = 0.5$ .

Figure 3.11, shows three example runs of the varying my values. Here there does seem to be a trend in the infection size in the endemic state. The higher the value for  $\mu$  the higher the infected population seems to get.

Figure 3.12 shows dependency of the infection population size on the value of  $\mu$ . Specifically a trend seems to appear where the higher the value for  $\mu$  gets, the larger the variance of the transients become.

The above figure 3.13, shows examples runs for different population sizes. Again we seem to find a trend where the larger the scale gets, the larger the transients become.

Lastly, the figure above 3.14, again shows an upwards trend in the average variance of the stochastic runs.

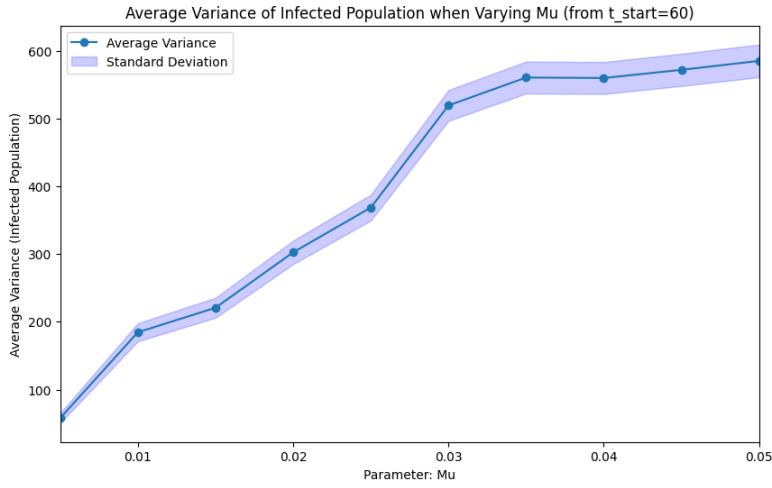


Figure 3.12: Shows the average variance and std of the variance over 60 runs for the different  $\mu$  values

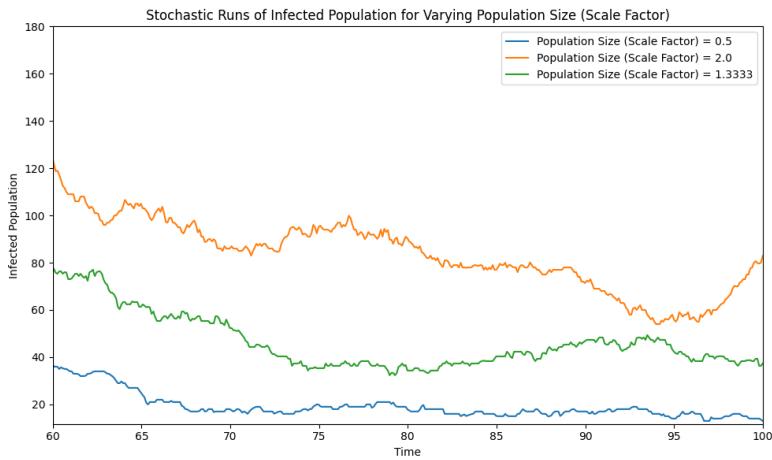


Figure 3.13: Shows 3 (first middle and last) runs of the 60 runs for the different parameter values for the population size from  $t=60$ ,  $t_{max}$ ; examples of increased transients

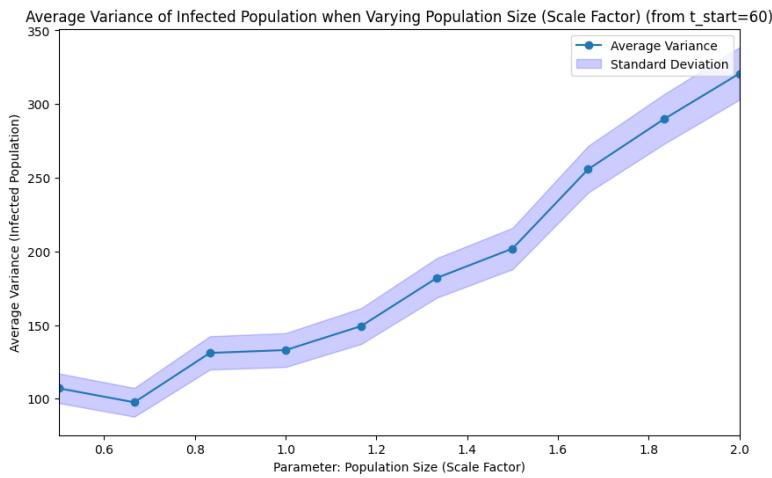


Figure 3.14: Shows the average variance and std of the variance over 60 runs for the different population sizes

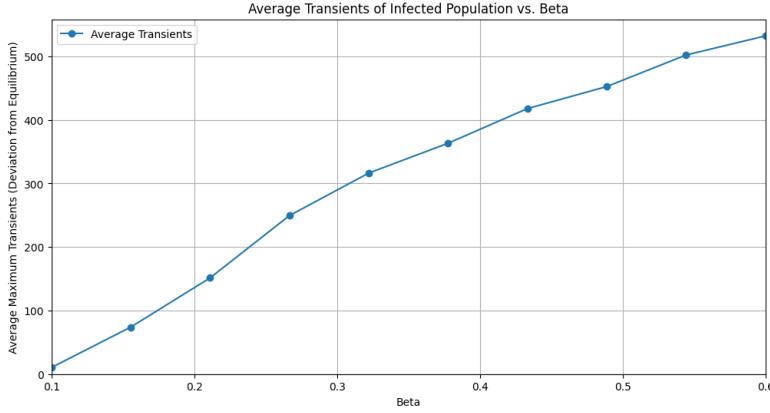


Figure 3.15: Shows the most important parameter to get the largest transients.

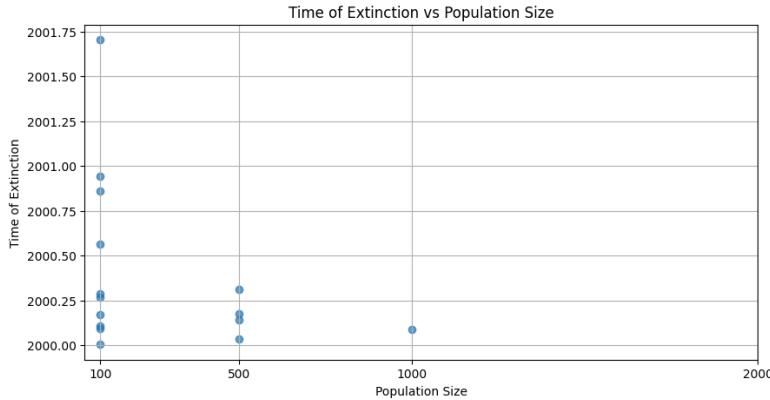


Figure 3.16: Extinction points for 10 runs per population size.  $(\beta, \gamma, \mu) = (0.3, 0.1, 0.01)$

### 3.2.3 Increased transients

The figure above, 3.15, shows the trend of increasing transients as the beta value increases.

### 3.2.4 Extinction events and Critical Community Size

Figure 3.16 shows that there appears to be an exponential trend between the population size and the number of extinctions per population size.

## 3.3 SIR Model on Networks

### 3.3.1 SIR Model on BA/WS/ER Networks

For Figures 3.17, 3.18, and 3.28, we observe how the infected population  $I$  changes with varying parameters across different network types.

In Figure 3.17 (SIR on BA), as we increase the infection rate  $\beta$ , we see a notable increase in the peak infected population. Higher values of  $\beta$  lead to a quicker rise in the infected population. In contrast, varying  $\gamma$  shows that higher recovery rates result in a quicker decline in the infected population. Finally, changing the initial infected population demonstrates that a larger initial count leads to a higher peak in infection.

In Figure 3.18 (SIR on WS), similar trends are observed. Increasing  $\beta$  again correlates with a rapid increase in the infected population. As we vary  $\gamma$ , we see that lower recovery rates sustain the infection for longer periods, resulting in a higher cumulative infected population. Changes

in the initial infected population also shows that starting with more infected individuals leads to a more pronounced and earlier peak.

For Figure 3.28 (SIR on ER), we see that varying  $\beta$  results in similar patterns: higher infection rates increase the total infected population over time. When examining  $\gamma$ , a consistent trend emerges where lower recovery rates maintain a higher number of infected individuals. Furthermore, variations in the initial infected population have the effect that with higher initial values, it results in a larger peak infected population and a more extended period of elevated infection levels.

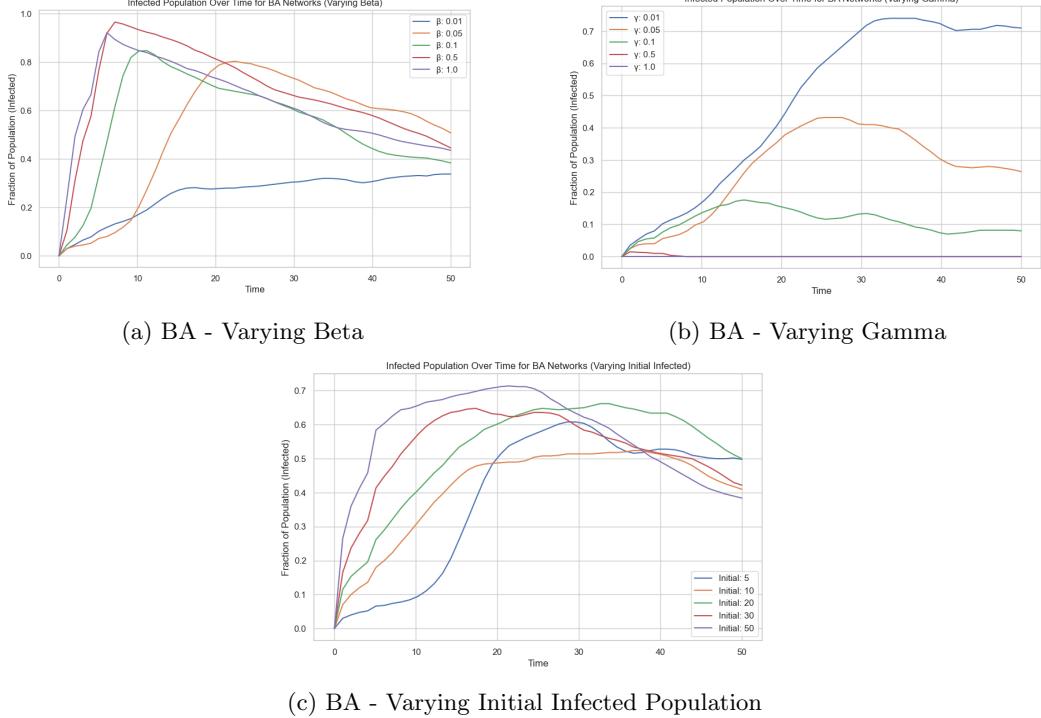


Figure 3.17: SIR Model on Watts and Strogatz (WS) network with varying parameters: Beta, Gamma, and Initial Infected Population. Standard parameters are  $\beta = 0.1$ ,  $\gamma = 0.01$ , and Initial Infected Population = 5.

### 3.3.2 SIR Model on Real-Life Network

#### 3.3.2.1 Epidemic Scenario

As seen in both Figure 3.20 and Table 3.1, the disease reaches its peak at Timestep 15, where the number of infected individuals is at its highest. Following this peak, the infection rate begins to decline, entering the Dying Down phase at Timestep 16 and further transitioning into the Middle Dying Down stage at Timestep 141. Eventually, the infection rate falls below the level of susceptible individuals, leading to complete disease extinction by Timestep 266.

#### 3.3.2.2 Non-Epidemic Scenario

As seen in both Figure 3.21 and Table 3.2, the disease struggles to establish itself within the population, as evidenced by the initial infection count remaining low. By Timestep 5, the infection quickly dies out.

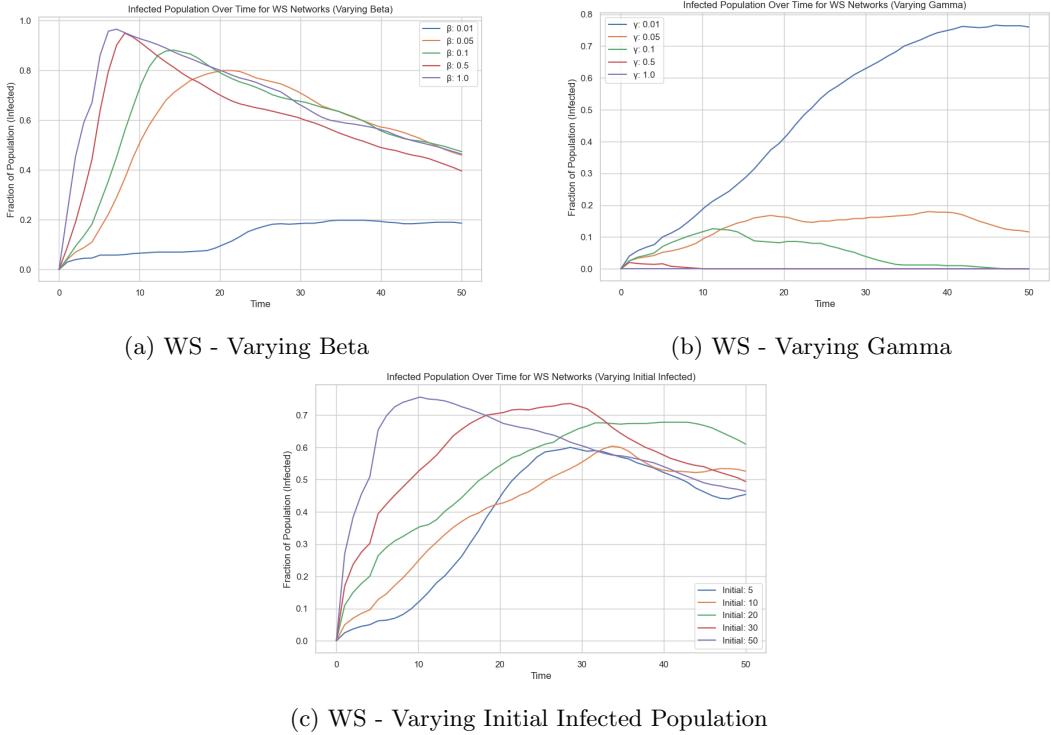


Figure 3.18: SIR Model on Watts-Strogatz (WS) network with varying parameters: Beta, Gamma, and Initial Infected Population. Standard parameters are  $\beta = 0.1$ ,  $\gamma = 0.01$ , and Initial Infected Population = 5.

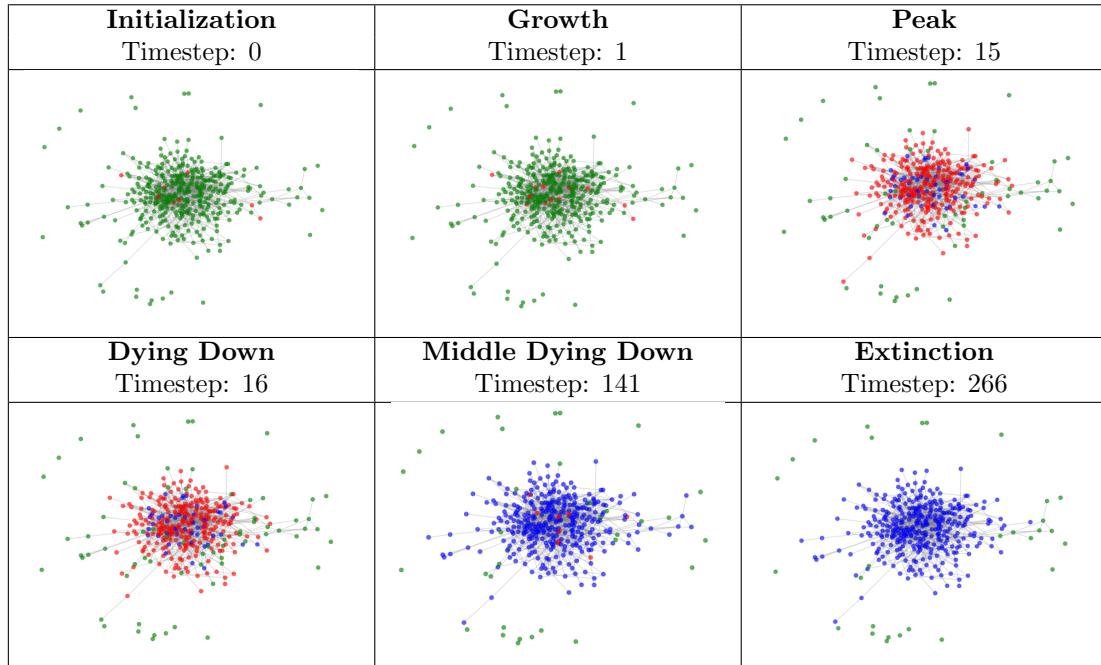


Table 3.1: Snapshots of the disease spread at six key stages: Initialization, Growth, Peak, Dying Down, Middle Dying Down, and Extinction. Each image illustrates the network state, with nodes color-coded by state (green for susceptible, red for infected, and blue for recovered).

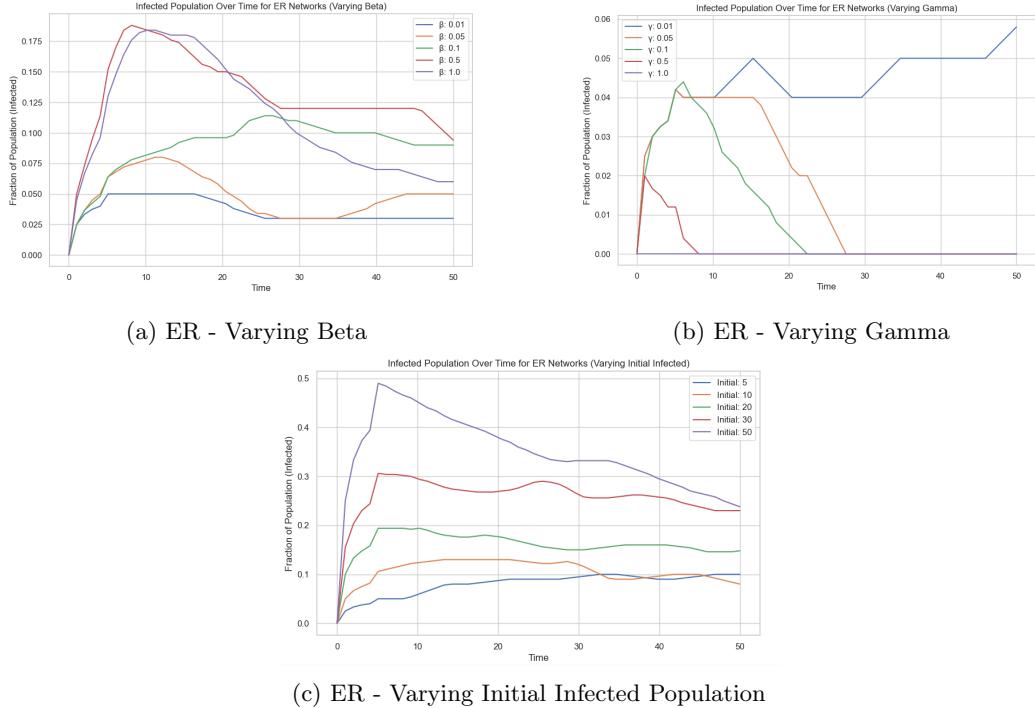


Figure 3.19: SIR Model on Erdos-Renyi (ER) network with varying parameters: Beta, Gamma, and Initial Infected Population. Standard parameters are  $\beta = 0.1$ ,  $\gamma = 0.01$ , and Initial Infected Population = 5.

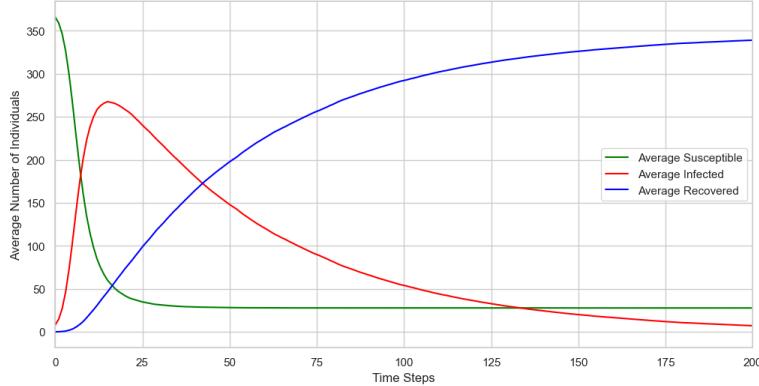


Figure 3.20: SIR Graph of Network Spread under the epidemic scenario with  $\beta = 0.1$  and  $\gamma = 0.02$

### 3.4 Network Statistics

Figure 3.22 to 3.26 showcase different network statistics, clearly delineating the various metrics for each network type. A comprehensive interpretation of these statistics will be provided in the subsequent chapters.

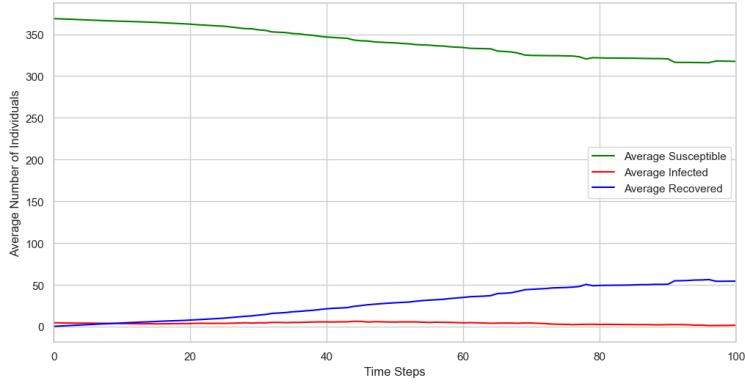


Figure 3.21: SIR Graph of Network Spread under the epidemic scenario with  $\beta = 0.01$  and  $\gamma = 0.1$

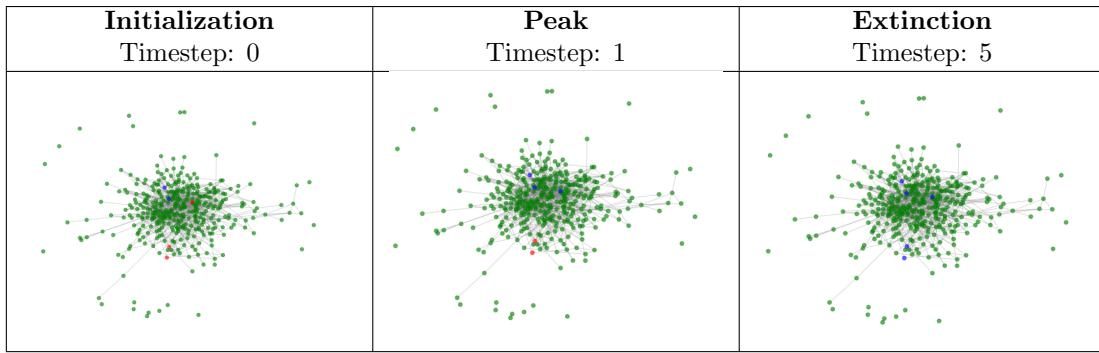


Table 3.2: Snapshots of the disease spread at three key stages: Initialization, Peak, and Extinction. Each image illustrates the network state, with nodes color-coded by state (green for susceptible, red for infected, and blue for recovered).

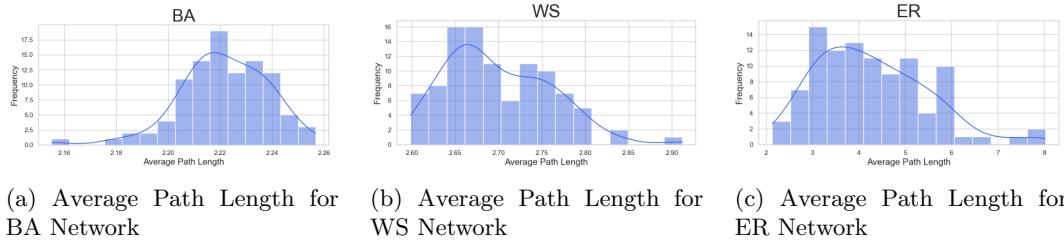


Figure 3.22: Average path length across different network types: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER). Each has been plotted 100 times and the average has been taken.

Figures 3.27a to 3.27e illustrate various network statistics as a function of population size across different network types: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER).

### 3.5 Vaccination Strategies

Figure 3.28 illustrates the effects of varying vaccination budgets on infection rates, comparing the targeted vaccination strategy against the random vaccination strategy. The targeted approach consistently results in a lower number of infections compared to the random strategy across all budgets. As the vaccination budget increases, the number of infected individuals decreases for both strategies; however, the reduction is more pronounced with the targeted strategy. More-

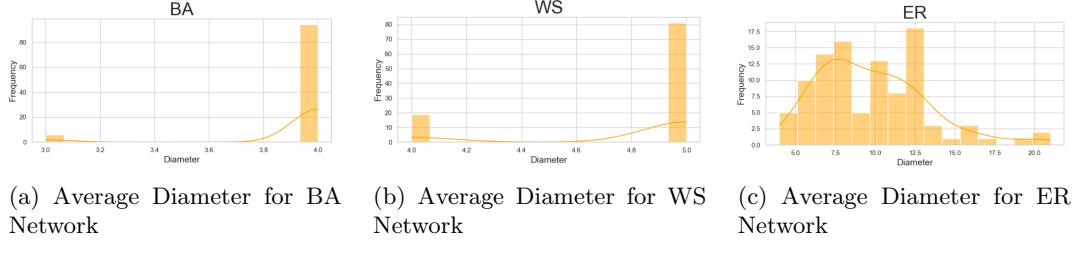


Figure 3.23: Average diameter across different network types: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER). Each has been plotted 100 times and the average has been taken.

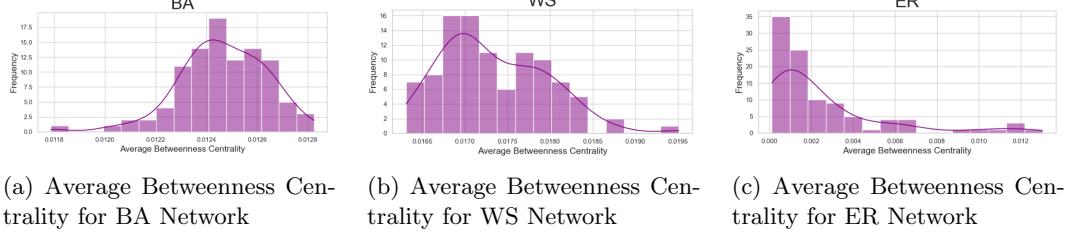


Figure 3.24: Average betweenness centrality across different network types: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER). Each has been plotted 100 times and the average has been taken.

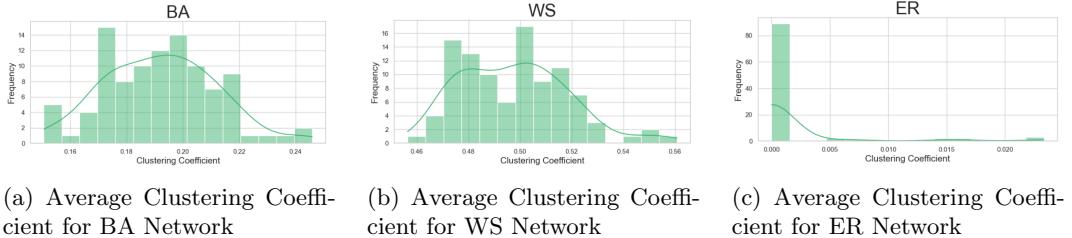


Figure 3.25: Average clustering coefficient across different network types: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER). Each has been plotted 100 times and the average has been taken.

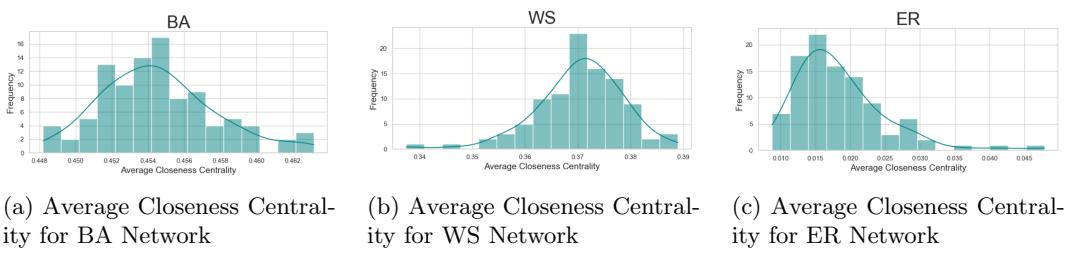


Figure 3.26: Average closeness centrality across different network types: Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER). Each has been plotted 100 times and the average has been taken.

over, when analyzing the probabilities of testing accuracy, the targeted strategy continues to outperform the random vaccination method.

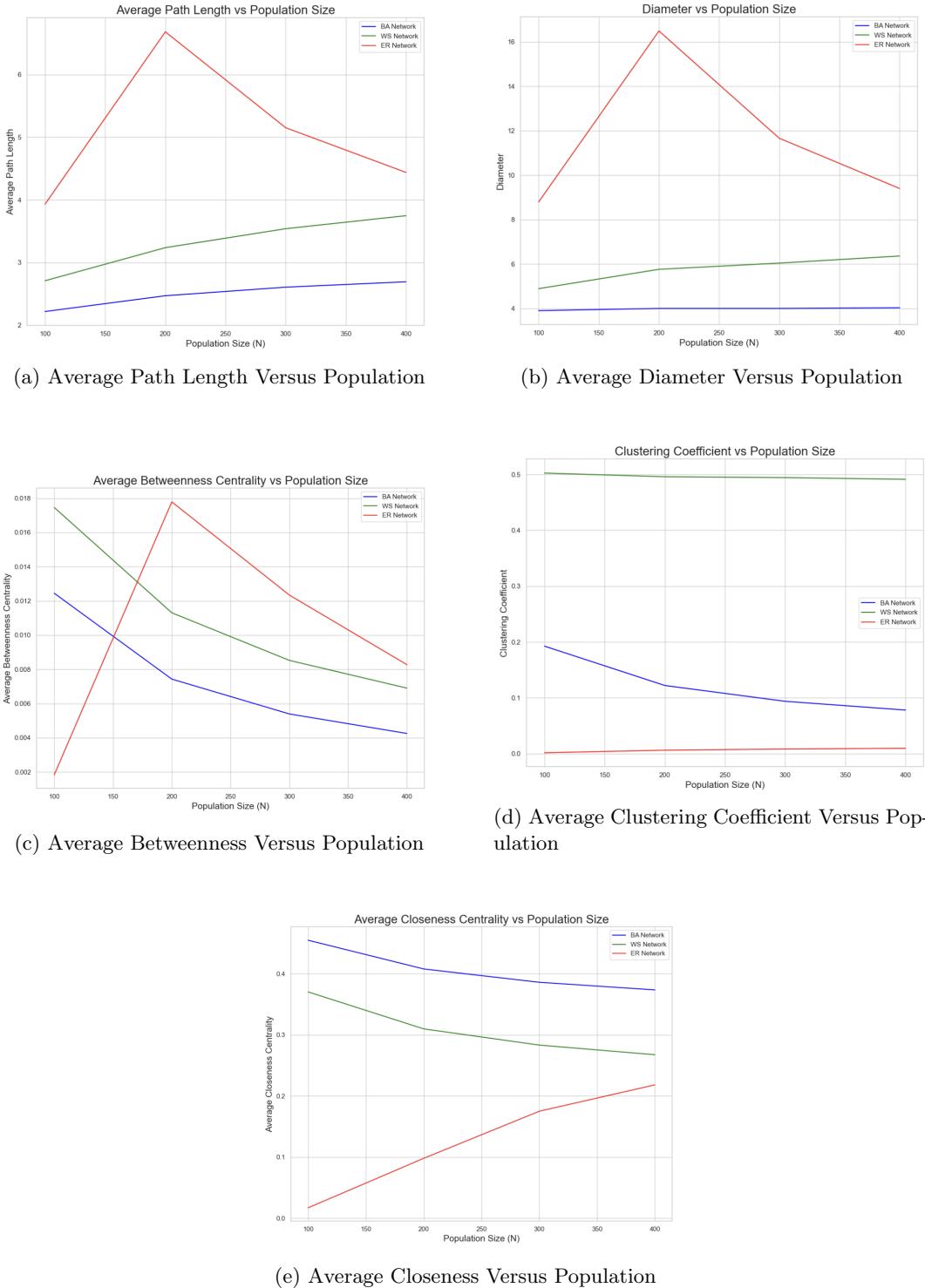


Figure 3.27: Comparison of Network Statistics Across Varying Population Sizes. This figure illustrates the relationships between average path length, average diameter, average betweenness centrality, average clustering coefficient, and average closeness centrality as the population size varies (0, 100, 200, 300, and 400).

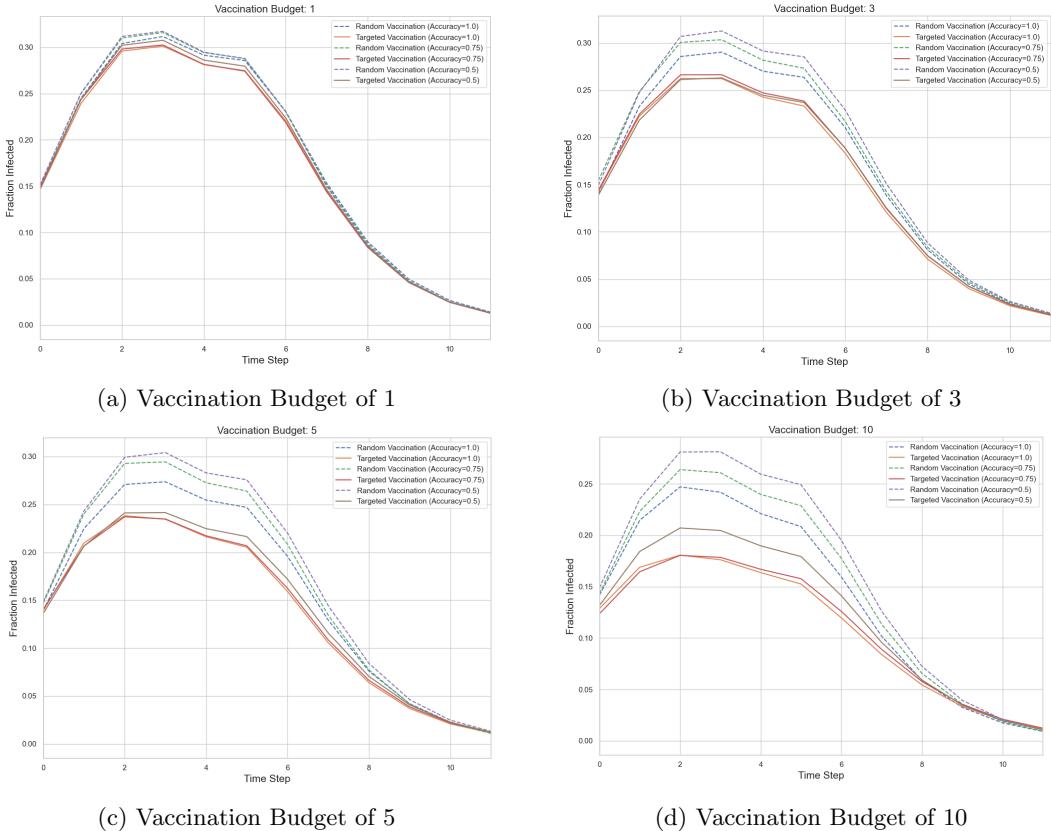


Figure 3.28: Impact of Vaccination Budget on Infection Rates. The graphs illustrate the relationship between different vaccination budgets (1, 3, 5, and 10) and probability that testing is accurate. Each plot shows the fraction of the population that remains infected over time.

---

## CHAPTER 4

---

# Discussion

---

This chapter presents a comprehensive analysis of the results obtained in the previous chapter. We will interpret the findings in the context of the models and strategies employed, providing insights into their implications. Furthermore, we will address any discrepancies observed in the results, considering factors such as network topology and the stochastic nature of disease transmission.

### 4.1 Implementation of Gillespie

In comparing the Gillespie algorithm to the standard ODE model, the key difference is in how the Gillespie algorithm captures the inherent randomness of the system. While the ODE approach provides a smooth, deterministic trajectory for the spread of infection, the Gillespie algorithm introduces stochasticity, allowing for fluctuations that reflect the probabilistic nature of individual interactions and infection events. This stochastic behaviour enables the Gillespie model to simulate variations in infection peaks and troughs that would not appear in a deterministic model, providing a more nuanced view of possible epidemic outcomes.

Another notable aspect of the Gillespie algorithm is its sensitivity to population size, which directly influences the level of noise in the system. In smaller populations, random events such as individual recoveries or infections lead to higher variability, or "noise," in the infection dynamics. This noise diminishes with larger populations, where random events tend to average out, producing dynamics that more closely resemble those predicted by the ODE model. This dependence on population size highlights the strengths of the Gillespie algorithm for modelling small or isolated populations where stochastic effects play a significant role.

### 4.2 Stochastic resonance and increased transients

The concept of stochastic resonance plays a critical role in understanding the dynamics of the Gillespie algorithm, particularly in how fluctuations can amplify responses in a system. In the context of epidemic modelling, an increase in the transmission rate, represented by the parameter  $\beta$ , leads to a lower negative average covariance between the susceptible ( $S$ ) and infected ( $I$ ) populations. This reduction in covariance indicates that as the infection spreads more aggressively, the susceptible population diminishes in a more regular and predictable manner. Consequently, the system experiences smaller fluctuations around its equilibrium, resulting in damped transients and more stable long-term behaviour.

Conversely, increasing the death rate ( $\mu$ ) introduces greater variability into the dynamics of the model. A higher  $\mu$  leads to larger negative covariance between  $S$  and  $I$ , as the loss of individuals from both populations due to mortality can create pronounced oscillations in infection

dynamics. This increased covariance signifies that changes in the susceptible population are more tightly coupled with changes in the infected population, leading to pronounced transients. As a result, the system experiences heightened fluctuations and variability, amplifying the effects of stochasticity in smaller populations.

These interactions between the parameters highlight the complex behaviour of the Gillespie algorithm in simulating real-world dynamics. Understanding how varying  $\beta$  and  $\mu$  affects the covariance between susceptible and infected populations can provide valuable insights into epidemic control strategies and the factors influencing outbreak dynamics.

### 4.3 Extinction events and Critical Community Size

The first graph, depicting the relationship between average transients of the infected population and the transmission rate ( $\beta$ ), reveals an upward trend. This indicates that higher transmission rates result in more substantial fluctuations within the infected population before reaching extinction. This behavior supports the hypothesis that higher  $R_0$  values lead to more prolonged and dynamic infection periods, potentially delaying the extinction due to the continuous generation of new cases.

In contrast, the second graph illustrates the time of extinction across varying population sizes, showing a trend where larger populations tend to reach extinction more rapidly than smaller ones. This might initially seem counterintuitive, but it suggests that in larger populations, while the disease spreads more quickly (due to higher contact rates), it also burns out faster once susceptibles are exhausted or herd immunity thresholds are reached. This rapid exhaustion of susceptibles in larger populations could expedite the extinction process, aligning with our discussions on how population size and  $R_0$  interact to influence disease dynamics in closed settings.

These observations provide clear visual evidence that both the transmission rate ( $\beta$ ) and population size ( $N$ ) significantly impact the time and dynamics of virus extinction in closed populations, supporting the core ideas discussed in the lectures about stochastic effects leading to different extinction outcomes despite  $R_0 > 1$ .

### 4.4 SIR Model on Networks

In this section, we will further explore the results obtained from implementing the SIR model on networks, as presented in Chapter 3.

#### 4.4.1 SIR Model on BA/WS/ER Networks

Throughout this section, we discuss the findings derived from the SIR model simulations performed on Barabási-Albert (BA), Watts-Strogatz (WS), and Erdős-Rényi (ER) networks. The results showcase how varying parameters such as infection rate ( $\beta$ ), recovery rate ( $\gamma$ ), and initial infected population impact the dynamics of disease spread across these networks.

In Figure 3.19a (SIR on BA), as we increase the infection rate  $\beta$ , we see a notable increase in the peak infected population. Higher values of  $\beta$  lead to a quicker rise in the infected population. In contrast, varying recovery rates  $\gamma$  shows that higher recovery rates result in a quicker decline in the infected population. Finally, changing the initial infected population demonstrates that a larger initial count leads to a higher peak in infection. It's important to note that the Y-axis scale for the BA network is particularly high, indicating a significant proportion of the population can become infected, which highlights the efficient spread characteristics of BA networks.

In Figure 3.19b (SIR on WS), similar trends are observed. Increasing  $\beta$  again correlates with a rapid increase in the infected population. As we vary  $\gamma$ , we see that lower recovery rates sustain the infection for longer periods, resulting in a higher cumulative infected population. Changes in the initial infected population also show that starting with more infected individuals leads to a more pronounced and earlier peak. The scale on the Y-axis for WS is more moderate compared

to BA, reflecting the unique structure of the network, which supports sustained infection but with a lower maximum infected fraction than BA.

For Figure 3.19c (SIR on ER), we see that varying  $\beta$  results in similar patterns: higher infection rates increase the total infected population over time. When examining  $\gamma$ , a consistent trend emerges where lower recovery rates maintain a higher number of infected individuals. Furthermore, variations in the initial infected population have the effect that higher initial values result in a larger peak infected population and a more extended period of elevated infection levels. The Y-axis scale for ER networks is lower than both BA and WS, suggesting that while infection can still spread, the overall efficiency is reduced due to the random nature of connections, leading to fewer highly connected nodes available for the infection to propagate.

In summary, while the trends of higher infection rates with higher  $\beta$  and slower declines with lower  $\gamma$  remain consistent across networks, the absolute values of infected populations and the duration of infections can vary significantly. Therefore, when analyzing or presenting the results, it's essential to highlight these scale differences to avoid misinterpretations of the graphs. The differing scales across the network types reinforce the importance of considering these factors when discussing the implications of the findings, which is crucial for accurately conveying the dynamics of disease spread in diverse network structures.

#### 4.4.2 SIR Model on Real-Life Network

The simulation results of the SIR model on real-life networks provide insights into how disease spread dynamics can vary depending on the parameters set within the model. Figures 3.20 and 3.21 illustrate the differences in the behavior of the infected population  $I$  under both epidemic and non-epidemic scenarios.

In 3.20, which represents the epidemic scenario with  $\beta = 0.1$  and  $\gamma = 0.02$ , we observe a pronounced peak in the number of infected individuals. The infected population rises sharply, indicating a rapid transmission of the disease throughout the network. This is followed by a gradual decline as individuals either recover or become removed from the susceptible population. The average number of susceptible individuals decreases significantly as the infection spreads, highlighting the aggressive nature of the outbreak. The rapid rise and subsequent decline suggest that without effective intervention strategies, the disease could lead to a substantial burden on the population.

Conversely, 3.21 depicts a non-epidemic scenario where the parameters prevent a significant outbreak. The infected population remains low and fluctuates minimally, indicating that the conditions are not conducive for widespread infection. This scenario underscores the effectiveness of lower infection and higher recovery rates in controlling the disease spread.

These findings illustrate the critical role of network dynamics in shaping infection trajectories. The contrast between the epidemic and non-epidemic scenarios emphasizes how varying initial conditions and transmission characteristics can lead to vastly different outcomes. Understanding these dynamics is essential for developing effective strategies for disease management and intervention in real-life settings.

## 4.5 Network Statistics

In this section, we analyze our findings from Figures 3.22 to 3.26 to discuss how the statistics differ across network types.

Starting with the **average path length** (Figure 3.22), we observe that the Barabási-Albert (BA) network exhibits the shortest average path lengths, indicating that it facilitates quicker connectivity between nodes compared to the other networks. This is due to its scale-free nature, which promotes the emergence of hubs that enhance connectivity. In contrast, the Watts-Strogatz (WS) network shows a wider distribution of path lengths, reflecting its rewired connections that can lead to both short and long paths. The Erdős-Rényi (ER) network displays the longest

average path lengths, suggesting that it is less efficient in connecting nodes, consistent with its random connection structure.

Looking at the **average diameter** (Figure 3.23), we find that the BA network again has the smallest diameter, reinforcing its efficiency in connecting nodes. The WS network shows a slightly larger diameter than BA but still maintains a relatively efficient connection due to its rewiring process. The ER network demonstrates the largest diameter, aligning with its characteristic of more isolated connections.

Examining **betweenness centrality** (Figure 3.24), we note that the BA network presents a higher average betweenness centrality, indicating the presence of nodes that play significant roles as bridges in the network. This characteristic is critical for controlling the flow of information or diseases. The WS network showcases a more moderate average, while the ER network has the lowest betweenness centrality, reflecting its less organized structure and fewer influential nodes.

When considering the **clustering coefficient** (Figure 3.25), the BA network again excels with the highest average, showing its propensity for tightly-knit communities. This clustering is indicative of the network's structure where nodes are more likely to form connections with other connected nodes. The WS network maintains a moderate clustering coefficient, influenced by its rewiring, which can create clusters without eliminating long-range connections. The ER network exhibits the lowest clustering coefficient, indicative of its random and less connected nature.

Finally, the analysis of **closeness centrality** (Figure 3.26) shows that the BA network provides the highest average closeness centrality. This suggests that, on average, nodes in this network can reach other nodes more quickly, which can be advantageous for efficient communication and resource distribution. The WS network also presents a decent average, while the ER network once again shows the lowest, reflecting its randomness which can hinder rapid access to all nodes.

Furthermore, Figures 3.27a to 3.27e present how these network statistics vary with population size across different network types. The BA network continues to show superior performance in terms of connectivity metrics, maintaining shorter average path lengths and diameters as the population increases. The average betweenness centrality for the BA network decreases with larger populations, suggesting that the network structure becomes more decentralized as it grows. In contrast, the WS and ER networks exhibit increases in average path lengths and diameters with population size, indicating reduced efficiency in connectivity.

The average clustering coefficient for the BA network remains consistently high, while both the WS and ER networks show a decline, reflecting their differing connectivity patterns. The closeness centrality metrics reveal that as population size increases, the BA network's nodes initially benefit from better connectivity but gradually align more closely with the performance seen in the WS and ER networks, which tend to have lower closeness centrality values.

This structured comparison highlights how different network topologies impact key metrics in the SIR model simulation, influencing both disease spread and control strategies. Each metric illustrates a unique aspect of the network's structure, revealing the underlying dynamics that can inform interventions in real-world epidemiological scenarios.

## 4.6 Vaccination Strategies

Figure 3.28 illustrates the effects of varying vaccination budgets on infection rates, comparing the targeted vaccination strategy against the random vaccination strategy. The targeted approach consistently results in a lower number of infections compared to the random strategy across all budgets.

As we analyze the graphs in Figure 3.28, we observe that increasing the vaccination budget reduces the fraction of infected individuals for both strategies. However, the targeted vaccination strategy shows a more pronounced reduction in infection rates. For instance, with a vaccination budget of 1 (subfigure a), the infection rate is noticeably lower for targeted vaccination when compared to random vaccination. This trend continues with increasing budgets of 3 (subfigure

b) and 5 (subfigure c), further highlighting the efficacy of targeted vaccination. By the time we reach a budget of 10 (subfigure d), the gap between the two strategies widens, demonstrating that targeted vaccination remains significantly more effective in controlling the spread of the infection.

Moreover, when analyzing the probabilities of testing accuracy, it is evident that the targeted strategy consistently outperforms the random vaccination method. Higher testing accuracy correlates with a greater reduction in the infected population, particularly for the targeted approach. This suggests that not only is the strategy of targeting individuals based on their connections beneficial, but the effectiveness is further amplified when combined with accurate testing.

The findings from the simulation underscore the importance of strategic vaccination approaches in managing infectious diseases within networked populations. The targeted vaccination strategy emerges as a superior method, providing greater control over infection spread, especially as resources increase.

---

## CHAPTER 5

# Conclusion

---

In this chapter, we will leverage the findings from our experiments and discussions to address our two primary research questions.

### 5.1 Conclusion Gillespie-Related Research Question

*How does the Gillespie method compare to standard ODE models, both including demography, and what phenomena emerge when applying the Gillespie method?*

The experiments in this study address the research question by examining how variations in key epidemiological parameters—specifically infection rate ( $\beta$ ), recovery rate ( $\gamma$ ), birth/death rate ( $\mu$ ), and population scaling—affect infection dynamics under the Gillespie method, a stochastic approach, in comparison to deterministic ODE models. Our findings show that infection rate  $\beta$  exerts the strongest influence on transient infection peaks, leading to larger fluctuations that eventually stabilize at equilibrium, while recovery and birth/death rates ( $\gamma$  and  $\mu$ ) promote quicker stabilization by dampening infection peaks.

Stochastic simulations run via the Gillespie method reveal phenomena not captured by ODE models, such as high variability in infection dynamics, especially in smaller populations where random fluctuations are more pronounced. This variability highlights the impact of demographic factors and stochastic effects in real-world epidemic modelling, showing that the Gillespie method captures a range of potential outcomes more reflective of natural variability than deterministic approaches. These insights provide a clearer understanding of how both modelling approaches can be leveraged to improve the accuracy and applicability of epidemic predictions and intervention planning.

### 5.2 Conclusion Network-Related Research Question

This study has shown how different network topologies influence the dynamics of infectious disease spread modelled using SIR simulations, alongside the impact of a targeted vaccination strategy on a real-world network.

The overarching research question guiding this segment was:

*How do different network topologies influence the spread of infectious diseases modelled using SIR simulations, and what impact does a targeted vaccination strategy have on a real-world network?*

Our findings affirm the hypothesis that network structure plays a critical role in shaping the spread of infectious diseases. Specifically, networks characterized by heterogeneous connectivity, such as the Barabási-Albert (BA) model, demonstrated a significantly more rapid and extensive

spread of infections compared to the more homogeneous structures exemplified by the Watts-Strogatz (WS) and Erdős-Rényi (ER) models. The BA network's scale-free property fosters the emergence of hubs that facilitate efficient disease transmission, leading to pronounced peaks in the infected population. In contrast, the WS and ER networks exhibited a more gradual increase in infections, highlighting their relative resistance to rapid disease propagation.

Furthermore, the implementation of a targeted vaccination strategy within the real-world Sociopatterns network yielded significant reductions in infection rates. Our analysis revealed that targeted vaccinations consistently outperformed random vaccination strategies, particularly as the vaccination budget increased. This underscores the importance of strategically selecting individuals for vaccination based on their connectivity within the network, thus optimizing resource allocation and enhancing the effectiveness of public health interventions.

In conclusion, the dual exploration of network topology and targeted vaccination strategies provides valuable insights into the complex dynamics of disease spread. The findings from this study not only contribute to the theoretical understanding of network influences on infectious diseases but also offer practical implications for public health strategies, emphasizing the necessity of tailored interventions in managing disease outbreaks effectively.

---

## Bibliography

---

- [1] Réka Albert and Albert-László Barabási. “Statistical mechanics of complex networks”. In: *Reviews of Modern Physics* 74.1 (2002), pp. 47–97.
- [2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. “Error and attack tolerance of complex networks”. In: *Nature* 406.6794 (2000), pp. 378–382.
- [3] Roy M Anderson and Robert M May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992.
- [4] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers. “When individual behaviour matters: homogeneous and network models in epidemiology”. In: *Journal of the Royal Society Interface* 4.16 (2007), pp. 879–891.
- [5] Albert-László Barabási and Réka Albert. “Emergence of scaling in random networks”. In: *Science* 286.5439 (1999), pp. 509–512.
- [6] Alain Barrat, Marc Barthélémy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge, UK: Cambridge University Press, 2008.
- [7] Béla Bollobás. *Random Graphs*. London: Academic Press, 1985.
- [8] Reuven Cohen, Shlomo Havlin, and Daniel ben-Avraham. “Efficient immunization strategies for computer networks and populations”. In: *Physical Review Letters* 91.24 (2003), p. 247901.
- [9] Radek Erban, Jonathan Chapman, and Ioannis G. Kevrekidis. “A practical guide to stochastic simulations of reaction-diffusion processes”. In: *SIAM Journal on Applied Mathematics* 70.4 (2007), pp. 1–40. DOI: [10.1137/070699652](https://doi.org/10.1137/070699652).
- [10] Paul Erdős and Alfréd Rényi. “On random graphs”. In: *Publicationes Mathematicae* 6 (1959), pp. 290–297.
- [11] Paul Erdős and Alfréd Rényi. “On the evolution of random graphs”. In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960), pp. 17–60.
- [12] Stephen Eubank, Hasan Guclu, VS Anil Kumar, et al. “Modelling disease outbreaks in realistic urban social networks”. In: *Nature* 429.6988 (2004), pp. 180–184.
- [13] Paul Fine, Ken Eames, and David L Heymann. “Herd immunity: history, theory, practice”. In: *Epidemiologic Reviews* 24.1 (2011), pp. 1–20.
- [14] Linton C Freeman. “A set of measures of centrality based on betweenness”. In: *Sociometry* (1977), pp. 35–41.
- [15] 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. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3). URL: <https://www.sciencedirect.com/science/article/pii/0021999176900413>.
- [16] et al. Hollingsworth T. D. “Generalised Gillespie Algorithms for Simulations in a Rule-Based Epidemiological Model Framework”. In: *PLOS Computational Biology* 16.3 (2020), pp. 1–23. DOI: [10.1371/journal.pcbi.1007643](https://doi.org/10.1371/journal.pcbi.1007643). URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007643>.

- [17] Matt J. Keeling and Ken TD Eames. “Networks and epidemic models”. In: *Journal of the Royal Society Interface* 2.4 (2005), pp. 295–307.
- [18] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press, 2008.
- [19] Stanley Milgram. “The small world problem”. In: *Psychology Today* 2.1 (1967), pp. 60–67.
- [20] Mark EJ Newman. “Spread of epidemic disease on networks”. In: *Physical Review E* 66.1 (2002), p. 016128.
- [21] Mark EJ Newman. “The structure and function of complex networks”. In: *SIAM Review* 45.2 (2003), pp. 167–256.
- [22] Mark EJ Newman. “The structure of scientific collaboration networks”. In: *Proceedings of the National Academy of Sciences* 98.2 (2001), pp. 404–409.
- [23] Romualdo Pastor-Satorras and Alessandro Vespignani. “Epidemic spreading in scale-free networks”. In: *Physical Review Letters* 86.14 (2001), pp. 3200–3203.
- [24] Romualdo Pastor-Satorras and Alessandro Vespignani. “Immunization of complex networks”. In: *Physical Review E* 65.3 (2002), p. 036104.
- [25] Gert Sabidussi. “The centrality index of a graph”. In: *Psychometrika* 31.4 (1966), pp. 581–603.
- [26] Marcel Salathé and Sebastian Bonhoeffer. “The effect of opinion clustering on disease outbreaks”. In: *Journal of The Royal Society Interface* 5.29 (2008), pp. 1505–1508.
- [27] Christian L. Vestergaard and Mathieu Génoin. “Temporal Gillespie Algorithm: Fast Simulation of Contagion Processes on Time-Varying Networks”. In: *PLOS Computational Biology* 11.10 (2015), pp. 1–28. DOI: 10.1371/journal.pcbi.1004579. URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004579>.
- [28] Duncan J. Watts and Steven H. Strogatz. “Collective dynamics of ‘small-world’ networks”. In: *Nature* 393.6684 (1998), pp. 440–442.