# Introduction to Computational Science
## *Assignment 2 Project: Modelling the Spread of Infectious Diseases*

Roman Dahm (13598279) and Jialin Dong (13610848)
*Graduate School of Informatics, University of Amsterdam*
*roman.dahm@student.uva.nl*
*jialin.dong@student.uva.nl*

Keywords: SIR model; Gillespie's Direct Algorithm; Spatial, Meta-population, and Network model; Vaccines

Abstract: This paper focuses on stochastic *Susceptible-Infected-Recovered* (SIR) models. By applying Gillespie's Direct Algorithm, it demonstrates the *five hallmarks* of stochastic SIR models. A spatial model using meta-populations is used to analyze the delays in the spread of diseases differences in the dynamics of sub-populations. Finally, we simulate SIR infections on Barabasi Albert, Watts-Strogatz, and Erdos-Reyni networks to discuss their dynamics and propose vaccination strategies.

## 1 INTRODUCTION

Since late 2019, the global coronavirus (COVID-19) pandemic posed a tremendous global public health challenge and caused considerable social and economic damage (Weber et al., 2021). Therefore, it is essential to build and understand predictive computational models for the dynamics of diseases. Particularly stochastic simulations can be a helpful tool to assess, predict, and mitigate real-world epidemics. In the previous report, the dynamics of various variations of the deterministic SIR model, which always produce the same results given initial conditions, have been explored. However, these assumptions do not hold in the real world, as infection traces are not deterministic and depend on many factors such as the number of contacts per person, spatial separation, and chance. This fact is especially pronounced for small population sizes and diseases with low reproductive ratios $R_0$. The stochastic nature of disease spreading can be modelled more realistically using stochastic SIR models, introducing changes and variance around the means.

Fundamentally, there are three ways to build stochastic models: introducing noise to the deterministic model, event-driven approaches, and network-based simulations. In this paper, event-driven and network-based approaches will be considered. For the event-driven approach, Gillespie's Direct Algorithm is used, which computes randomly distributed dynamic events to simulate the spreading of disease. To analyze its characteristics, the *five hallmarks* – variability, negative covariances, increased transients, stochastic resonance, and extinctions are studied.

Network-based approaches are well suited for modelling complex system interactions, such as spatial aspects like geographic and social interactions arising from the fact that an individual can only transmit disease in a restricted area within a population. Different separated sub-populations also come with their unique interactions. By varying the transmission rates between subpopulations, one can obtain insight into the variances of epidemic peaks and delay times. We use three popular types of graph networks: Barabasi Albert, Watts-Strogatz, and Erdos-Reyni. When modelling diseases with the SIR model, every individual can be represented as a node of the population network with edges representing its relationships with other individuals. Network-based simulations reveal that different social and hierarchical structures play a vital role in the dynamics of the real world spreading of diseases.

## 2 METHODOLOGY

This paper absorbs many mathematical methods based on **Infectious Diseases in Humans and Animals** (Keeling and Rohani, 2008) to implement the computer based simulations. We used Python 3.8 as well as the powerful `NDLib`, `DyNetx`, and `NetworkX` libraries to implement network based simulations. Most results such as scatter plots and graphs were plotted using the `Matplotlib` package.

Pearson correlation coefficient analysis was pri-

marily used for the data analysis, which calculates the correlation coefficient and significance level (P-values) for large samples of distributed scatter points.

## 2.1 SIR discrete event model

Although deterministic models are useful for analyzing SIR dynamics, they miss the stochastic nature of disease spreading as it happens in the real world. This is because events such as infections are probabilistic. To give a concrete example, for one infected individual, the number of secondary cases is not deterministic but depends on how many susceptible targets the individual meets that day, how close they get, how likely a target is to get infected and so on. Therefore, deterministic models are more suitable for simulating substantial populations because random fluctuations tend to cancel each other out and reduce randomness. In contrast, the event-driven model offers a more realistic view of dynamics, especially for small populations.

Incorporating demographic stochasticity into a model is a broad topic, as demographic stochasticity is defined as variations in population dynamics. This is due to the randomness of events occurring at the individual level. It is clear that, while the overall tendency of the population process stays stable, individuals may have a different experience. In Gillespie's 1977 paper, he proposed a direct method to offer demographic stochasticity practically, which is drawn from the Monte Carlo method. The algorithm is given in Algorithm 1. Until the termination time $T$ is reached, arrival times for each possible event are computed based on the occurrence rates. Then, the next event with the closest arrival time is chosen, and its dynamics are executed. Finally, the simulation time is incremented by the selected arrival time. The whole process obeys a stochastic simulation.

**Algorithm 1** Gillespie's Direct Algorithm

**Require:** $T$ termination time
**Require:** Labeled events $E_1$,..., $E_n$.
  $t \leftarrow 0$
  **while** $t \leq T$ **do**
    **for** $1 \leq i \leq n$ **do**
      $R_i \leftarrow \text{rate}(i)$ ▷ Rate of event
      $\delta_i = \frac{-1}{R_i} \log(\mathcal{N}(0,1))$ ▷ Arrival time
    **end for**
    $i_{min} \leftarrow \forall_j \delta_{i_{min}} \leq \delta_j$ ▷ Find next event
    perform$(E_{i_{min}})$
    $t \leftarrow t + \delta_{i_{min}}$
  **end while**

In the concrete case for the discrete event SIR model with demography and imports, the events, rates, and dynamics are given in Table 1.

| Event | Rate | Dynamic |
|---|---|---|
| Infection | $\frac{\beta SI}{N_{pop}}$ | S - 1, I + 1 |
| Recovery | $\gamma I$ | I - 1, R + 1 |
| Birth | $\mu N_{pop}$ | S + 1 |
| Death in S | $\mu S$ | S - 1 |
| Death in I | $\mu I$ | I - 1 |
| Death in R | $\mu R$ | R - 1 |
| Imports | $\varepsilon$ | I + 1 |

Table 1: All events, rates, and dynamics for the discrete event model of SIR with demography and imports used for Gillespie's Direct Algorithm. The parameters $\beta$, $\gamma$, $\mu$, and $\varepsilon$ are transmission, recovery, birth and death, importing rates. S, I and R represent the number of susceptible, infected, and recovered individuals. $N_{pop}$ is the total population size ($N_{pop} = S + I + R$).

## 2.2 Five hallmarks

Keeling and Rohani categorized key dynamics of stochastic SIR models in terms of the *five hallmarks*:

1. Variability
2. Negative co-variances
3. Increased transients
4. Stochastic resonance
5. Extinctions

We will furthermore examine the *first passage time* and *critical community size*, which are also unique to stochastic models and arise from the latter. In the following, we will briefly introduce each hallmark. Experimental results are discussed in 3.1.

**Variability**

Since an event's probabilities are drawn from a uniform random distribution based on the deterministic rate, the arrival time is non-deterministic and impossible to predict with high precision. The outcomes will be different for every simulation run. However, the overall dynamics of an epidemic are relatively stable.

**Negative co-variances**

Negative co-variance is the difference between the numbers of infected and susceptible people, as determined by the interactions of stochastic and deterministic dynamics. The population size can influence this

index. The equation for the co-variance between two series of samples $X, Y$ is as follows.

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

where $n$ is the size of the sample sequence, $X_i$ and $Y_i$ are the unit values, and $\bar{X}$ and $\bar{Y}$ are the average values.

**Increased transients**

Compared with the deterministic model, the stochastic model does not have a deterministic equilibrium, but it has similar resilient behaviour around the deterministic equilibrium, where perturbations away from the equilibrium are countered with transient like returns back.

**Stochastic resonance**

A deterministic model can reach an equilibrium state before experiencing decaying oscillations. For the stochastic model, when the frequency of stochastic dynamics is close to the natural frequency after reaching a relative equilibrium, the oscillation amplitude is amplified, which can lead to epidemic oscillations around the endemic state.

**Extinctions**

For a disease with $R_0 > 1$, extinction does not occur in a deterministic model but can occur when using a stochastic model due to random fluctuations that impact the dynamics. Population size and $R_0$ are important factors to consider, as small populations and small $R_0$ increase the chance of extinction. To deal with the chance of extinction, an *import rate* is often added to the model, which will prevent extinction by importing newly infected people from other sources. In this experiment, imports are not taken into account.

**Advanced topic: The first passage time**

The time to extinction can be explored by measuring the average time until the population goes extinct, called "first passage time". To account for the stochastic nature, multiple simulations have to be performed and averaged. As before, imports are not taken into consideration.

**Advanced topic: Critical community size**

Given that small populations and diseases with a small reproductive ratio $R_0$ always suffer from extinction due to random fluctuations, the *critical community size* is a useful persistence metric that determines the minimum population size required to guarantee persistence of disease around the endemic equilibrium.

## 2.3 Meta-population model

As for a real disease spreading, the spatial factor plays a key role in dynamics. For example, the social sphere influences interactions, geographic limitation restricts the disease into a local area, and lockdown cuts off people's contact to prevent the disease from spreading. These can be categorized as heterogeneity, interaction, and isolation. The previous models and experiments assumed that the disease transmission is modeled as a population global rate. In the real world, populations can be modeled as many distinct sub-populations (e.g. metropolitan ares, countries, islands), which can each have different dynamics (e.g. higher transmission rate in large city than in rural areas) and different interactions between any pair of two sub-populations (e.g. tourist favourite island has interaction with many countries, small cities most likely only with neighbour cities). The meta-population model can be used to explore the mechanisms of how sub-populations affect dynamics. In our experiment, we use a naive SIR model without demography.

Similarly, Gillespie's Direct Algorithm is used to simulate the dynamics with events and rates as given in Table 2.

| Event | Rate | Dynamic |
|---|---|---|
| Infection in i | $\lambda_i S_i$ | $S_i$ - 1, $I_i$ + 1 |
| Recovery in i | $\gamma I_i$ | $I_i$ - 1, $R_i$ + 1 |

Table 2: Possible events with their rates and dynamics for the discrete event model for each meta population $i$

Each sub-population $i$ out $n$ sub-populations has its own unique transmission rate $\lambda$ as given by:

$$\lambda_i = \beta_i \sum_{j=1}^{n} \rho_{ij} \frac{I_j}{N_i} \quad (2)$$

where $\rho$ is a n×n matrix where each entry $\rho_{ij}$ represents the force of infection from sub-population $j$ to $i$. $N_i$ is the sum of $S_i$, $I_i$, and $R_i$.

For our experiments, Pearson correlation coefficient analysis was used to explore the delays and epidemic peaks between two sub-populations. This method is mainly used to reflect the degree of linear correlation between two variables. The equation for the Pearson correlation coefficient $P_{X,Y}$ is as follows:

$$P_{X,Y} = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

where $\mu_X$ and $\mu_Y$ are the mean values of X and Y. $\sigma_X$ and $\sigma_Y$ are the standard deviations of X and Y. $\mathbb{E}$ is the expected value.

The value of $P_{X,Y}$ is in the range $[-1, 1]$. The larger the absolute value, the stronger the correlation.

## 2.4 Network model

Networks offer a unified method of thinking about interactions between people or groups, which is especially beneficial when each individual has direct contact to only a tiny fraction of the total population(Newman et al., 2006). Networks are an effective representation to model how illnesses spread in human populations through social interactions. In the network framework, each infected individual can only infect its neighbours. In the following, we will give a short summary on graphs and define the neighbourhood of an individual. Compared to other models with a more macroscopic view, the obvious merit of the network model is that it provides a magnifying glass to observe complicated individual behaviour.

Let a network (graph) be denoted as $G = (N, L)$ with a set of nodes (individuals) $N$ and links $L$. There exists an edge $(v, u) \in L$ from $v \in N$ to $u \in N$ if the individual $v$ has direct contact to individual $u$. For social interactions, directed edges do not make sense and we consider all edges as undirected ($(v, u) \in L \Leftrightarrow (u, v) \in L$).

A very important measure is the degree $k_i$ of a node $i \in N$, which is the number of edges that $i$ shares. The *neighbourhood* of an individual $i \in N$ is thus all the other individuals which share an edge with $i$. As mentioned before, each infected individual can only infect its neighbours which can cause a maximum number of secondary cases bounded by $k_i$ even for large populations. For undirected networks, the *average node degree* $\langle k \rangle$ is defined as the average node degree for all nodes:

$$\langle k \rangle = \frac{1}{|N|} \sum_{i \in N} k_i \tag{4}$$

Another relevant network metric is the average shortest-path length $\langle d \rangle$. Let $d(i, j)$ be the shortest path distance function between two nodes $i, j \in N$, for an undirected graph network, $\langle d \rangle$ can be calculated as follows:

$$\langle d \rangle = \frac{2}{|N|(|N|-1)} \sum_{i,j>i \in N} d(i, j) \tag{5}$$

where $i, j > i \in N$ makes sure that distances between the same nodes in different directions are are only counted once and $\frac{|N|(|N|-1)}{2}$ are the number of node pairs $i, j > i \in N$.

Finally, we introduce the average clustering coefficient given by:

$$\mathrm{C}_G = \sum_{i \in N} \frac{2e_i}{|N| k_i (k_i - 1)} \tag{6}$$

where for a node $i \in N$, $e_i$ is the number of all edges of neighbour nodes, $k_i$ is the degree. This coefficient is used to describe the degree of clumping between vertices in a graph. Specifically, it gives an assessment of the overall degree of agglomeration in a graph.

In the following, we will briefly discuss the different types of graphs that will be used for the experiments.

**Scale-free networks**

In 1999 Barabási and Albert proposed that the number of contacts for individuals is highly heterogeneous. The connections are relatively low for most people, whereas some individuals have plenty of relationships with others. In our context, we also refer to the latter as so-called *hyperspreaders*. As individuals with a high node degree can potentially cause a high number of secondary cases, we investigate the dynamics of Barabási Albert networks, which closely match real world conditions. The mechanism of construction for such a network of $n$ nodes involves linking a newly added node with $m$ other nodes with a preference for those with a high node degree.

**Small World Networks**

A small-world network is inferred from a lattice structure, with a small number of long-range connections added into the model. The Watts-Strogatz network (1998) has relatively high clustering because newly added nodes tend to be linked with the nearest neighbour nodes. In those networks, the transmission happens mostly locally, while some infections can go farther by accidental long-range connections.

**Random Networks**

As its name suggests, random networks neglect the spatial location of different nodes. The Erdos-Reyni network links a new node with equal possibility $p$ with other any other node in the network. In these networks, the clustering phenomenon is not emphasized with a clustering coefficient of $p = \frac{\langle k \rangle}{|N|}$. The average shortest path is small ($\approx \frac{\log |N|}{\log \langle k \rangle}$) because of many long-range connections.

# 3 RESULTS

This section presents the results of all experiments conducted, including phase space, control variable, and Pearson correlation coefficient analysis, as a result of observing stochastic, meta-population, and network models with dynamics, parameters interplay, and oscillations behaviour.

## 3.1 Discrete event model simulation

As shown in Figure 1, the high standard deviation shows that the values differ for each run of the simulation when the population size is small.

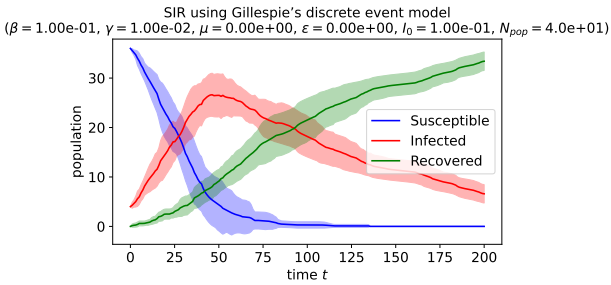
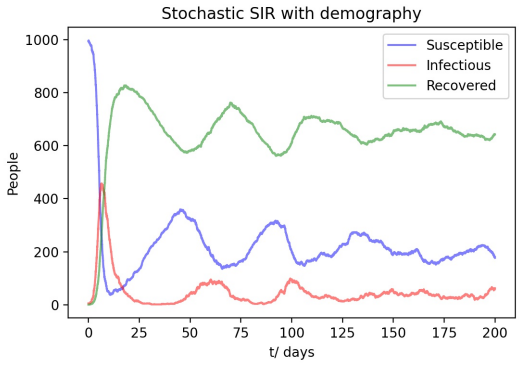


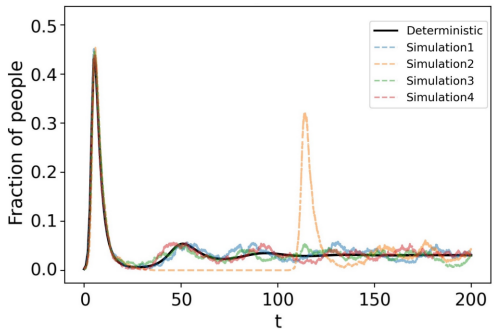

Figure 1: Variability of the stochastic SIR model for a small population $N_{pop} = 40$. The filled confidence bars show the standard deviation over 10 runs of the simulation.

After adding demography and imports, one can observe oscillatory behaviour for all three compartments as depicted in Figure 2a. This phenomenon is consistent with the deterministic model. Similarly, Figure 21 (see Appendix) compared the graphs for the number of infections with the deterministic model for multiple runs of the stochastic simulations Figure 2b shows the combined simulations. It can be seen that during early stage, four stochastic simulations nearly have the same trend, except for the second simulation. This is possibly because the infected people recovered or accidentally died before transmission. However, there is also a high peak for the second simulation about at about $t$ =120 days. The most likely cause for this is that imports of infected individuals from other sources caused a sporadic epidemic outbreak.

According to deterministic modeling, the changing direction of S and I is not obvious in different stages. Therefore, to verify if the negative co-variance stay the same trend, it selects two important time points: one is the early stage, which is around the epidemic period, and another is the 365 days, which is used to observe long time span relation. From Figure 3, it seems no obvious difference between two stages. The susceptible and infected have a clear negative relation with each other. This results accord with the reality in case of total population is stable.




(a) SIR model with demography and imports. ($S_0 = 995$, $I_0 = 5$, $R_0 = 0$, $\beta = 1.4$, $\gamma = 0.3$, $\mu = \frac{1}{60}$, $\varepsilon = 0.001$, time = 200)




(b) Four simulations of SIR model. ($S_0 = 2000$, $I_0 = 5$, $R_0 = 0$, $\beta = 1.8$, $\gamma = 0.3$, $\mu = \frac{1}{60}$, $\varepsilon = 0.001$, time = 200)

Figure 2: Stochastic SIR model simulations

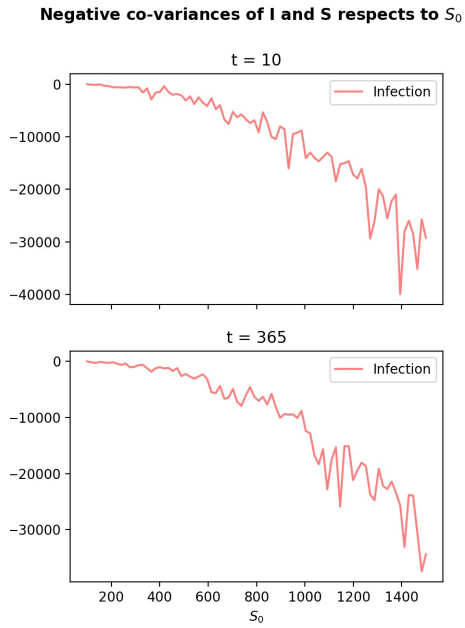



Figure 3: Negative co-variances analysis at different stages. ($S_0$ = [from 100 to 1500], $I_0 = 5$, $R_0 = 0$, $\beta = 1.4$, $\gamma = 0.3$, $\mu = \frac{1}{60}$, $\varepsilon = 0.001$, $time_1 = 10$, $time_2 = 365$ )

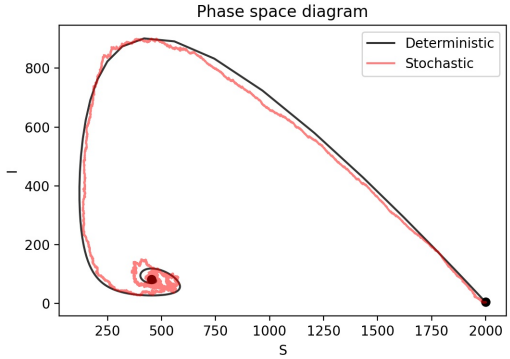



Figure 4: Phase space analysis of stochastic and deterministic process.
($S_0 = 2000$, $I_0 = 5$, $R_0 = 0$, $\beta = 1.4$, $\gamma = 0.3$, $\mu = \frac{1}{60}$, $\varepsilon = 0.001$, time = 200)

Figure 4 shows the phase space diagram comparing the stochastic and deterministic model, where observed states – points $(I_t, S_t)$ of the number $I_t$ of infections at time $t$ and $S_t$ the number of susceptible individuals at time $t$, are plotted in the two dimensional state space. The state space trajectory of the stochastic model closely follows the dynamics of the deterministic model. Even if it deviates from the trajectory, it is pulled back quickly by transient-like returns. This shows that the stochastic process has a decent degree of resilience in its dynamics. More importantly, th stochastic model is showing considerable fluctuation around the fixed point. This shows that the additio of stochastic noise during the process causes othe wise very weak oscillations to amplify and excee the deterministic threshold amplitudes and temporaril leave the fix point, which is a phenomenon known a *stochastic resonance*.

Figure 5 illustrates the relationship between th number of extinctions per year for ten combined simu lations with respect to the population size $N_{pop}$ and th reproductive ratio $R_0$. The result shows that both popu lation size and $R_0$ have a negative relationship with th chance of extinction. When the population size or $R$ is small, extinctions occur in nearly every simulatio For larger population sizes and $R_0$, extinctions are fa less likely to occur.

Figure 6 makes 30 times stochastic simulations and calculates the average first passage time for different population sizes, combined with number of extinction that occurred. It illustrates that when the size of the population is below around 800, the first passage time can be prolonged. However, after that, the first passage time has a decreasing trend. During the whole process, the extinction time has a strong negative relationship with population size.

Diseases with a more variable infectious period

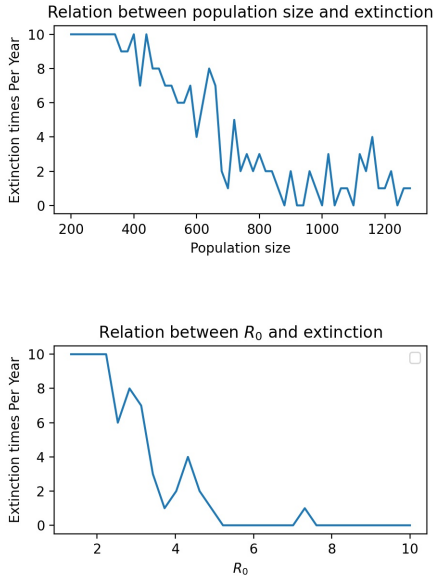



Figure 5: Analyze the impacts of population size and $R_0$ on extinction times per year.
($S_0$ = [from 200 to 1200 increases by 20] , $I_0 = 5$, $R_0 = 0$, $\beta$ = [0.4 to 3], $\gamma = 0.3$, $\mu = \frac{1}{60}$, $\varepsilon = 0.001$, time = 365)

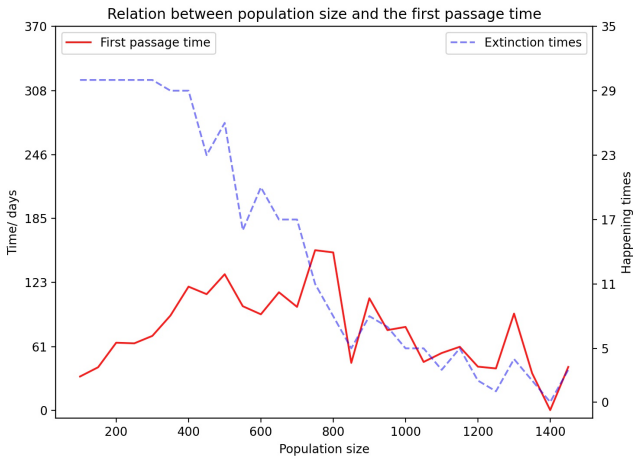



Figure 6: The first passage time respects to population size.
($S_0$ = [from 100 to 1500 increases by 50] , $I_0 = 5$, $R_0 = 4.7$, $\beta = 1.4$, $\gamma = 0.3$, $\mu = \frac{1}{60}$, $\varepsilon = 0.001$, time = 365, simulation times = 30)

have more significant variability in the number of secondary cases, and therefore a greater risk of extinction when infectious numbers are low. This leads to the critical community size as a persistence measure for

the minimum population size $N_c$ required to avoid extinction.

We define $N_c$ as the population size for which the number of extinctions per year is below 2. To show the dependencies of $N_c$ with respect to $R_0$ and $\gamma$, we iteratively find the critical community size $N_c$ for combinations of $R_0$ in the range $[1,21]$ and $\gamma$ in the range $[50,550]$.

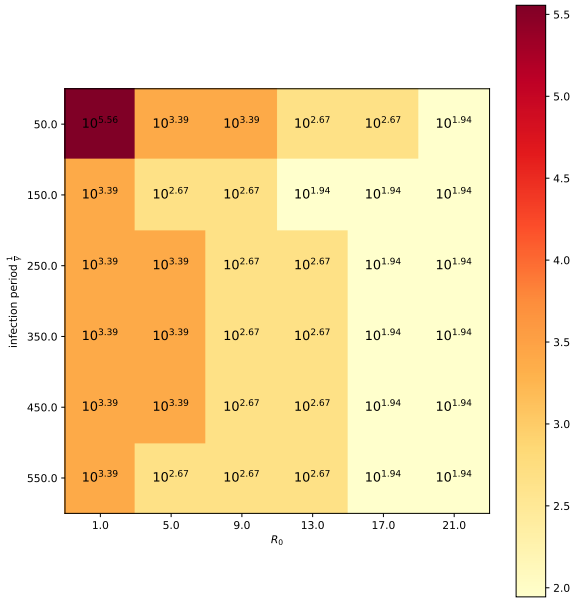


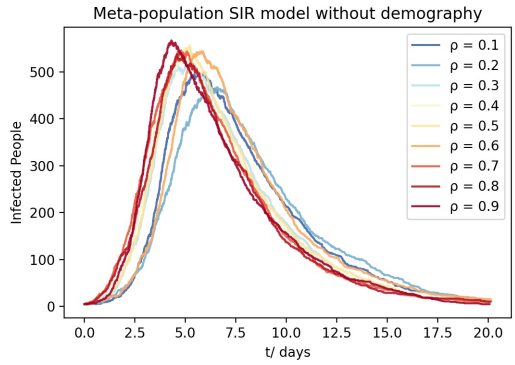

Figure 7: The critical community size $N_c$ with respect to different values of $R_0$ and $\gamma$. Colors denote the exponents to the base 10.

Figure 7 shows the results for the critical community size $N_c$ that were found empirically through simulation. For all values of $\gamma$, the critical community size for $R_0 = 1$ needs to be considerably higher than for $R_0 = 21$ ($10^3 < N_c < 10^6$ compared to $10^2 < N_c < 10^3$). Also, shorter infection periods require higher critical community sizes, as individuals do not remain infectious for long enough to spread the disease.

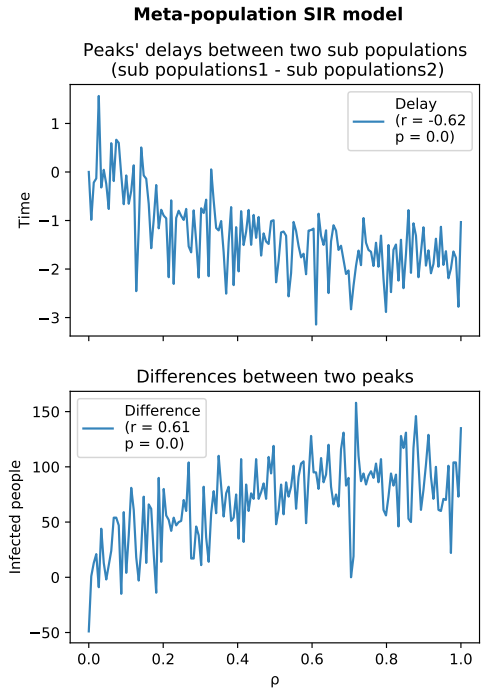
## 3.2 The influence of sub-populations

In our experiments, we first used $n = 2$ sub-populations and tested the impact of different $\rho$ on their dynamics.

Figure 8a gives a clear exhibition of how $\rho$ influences the epidemic of population 1, where $\rho$ is a measure of the strength of interaction between populations. In this part, $\rho_1 2$ measures the relative strength of transmission to subpopulation 1 from subpopulation 2. As $\rho$ increases, the epidemic happens earlier with a higher peak. It shows that the transmission from another city can bring a source of infection, but the starting and ending time for the whole epidemic is stable.




(a) Effects of $\rho$ on population 1 within two populations



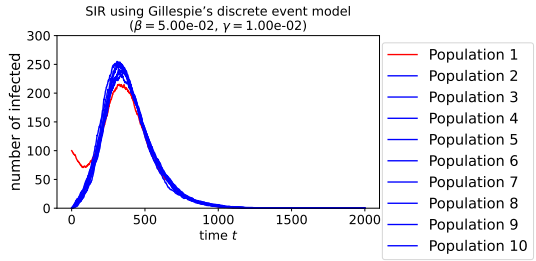

(b) Further experiments

Figure 8: Two subpopulations stochastic model

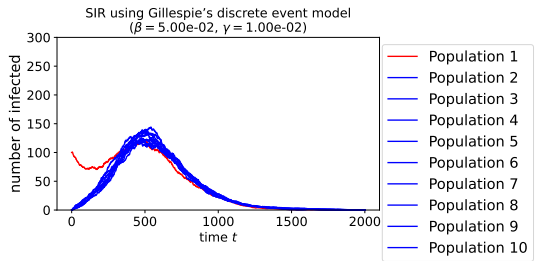
To verify the guess from Figure 8a, it applies Pearson correlation coefficient analysis. It calculates the correlation coefficient and significance level (p-value) and finds the p-value equals 0, so the correlation coefficient is highly statistically significant. Although the r-value is medium, it still has a relatively negative relation with delay time and positive relation with infected people peak difference.

In addition to the $n = 2$ sub-populations considered so far, we also consider a meta-population model with $n = 10$ sub-population and varying interaction matri-

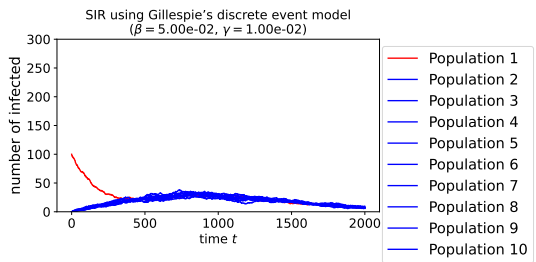
(a) High bidirectional interaction



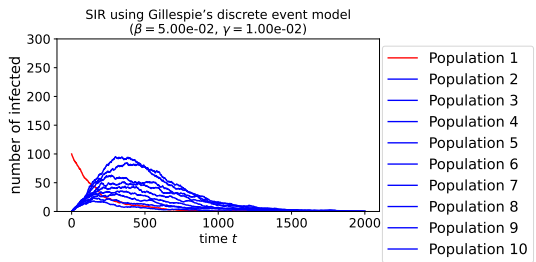

(b) Medium bidirectional interaction




(c) Low bidirectional interaction




(d) Unidirectional 1




(e) Unidirectional 2

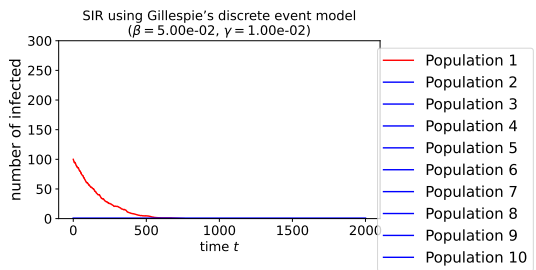



Figure 9: Effect of different interaction matrices $\rho$ on a meta-population SIR without demographics with $n = 10$ sub-populations. ($\beta = 0.05$, $\gamma = 0.01$)

ces $\rho$. For our experiments, we test bidirectional interactions with $\rho_{ij} \in \mathcal{N}(0.75, 0.1), \rho_{ij} \in \mathcal{N}(0.5, 0.1)$, as well as $\rho_{ij} \in \mathcal{N}(0.25, 0.1)$ with $\rho_{ii} = 1$ for all of them, denoted as $\rho^{high}_{bidirectional}$, $\rho^{medium}_{bidirectional}$, and $\rho^{low}_{bidirectional}$ respectively. Furthermore, unidirectional interaction $\forall i < j : \rho_{ij} = 0 \wedge \rho_{ij} \in \mathcal{N}(0.5, 0.5)$ and $\forall j < i : \rho_{ij} = 0 \wedge \rho_{ij} \in \mathcal{N}(0.5, 0.5)$ with $\rho_{ii} = 1$ for all of them denoted as $\rho^{1}_{unidirectional}$ and $\rho^{2}_{unidirectional}$ are tested.

Sampling from a Gaussian distribution more closely matches real world conditions. To ensure correctness, all entries are clipped to bounds of $[0, 0.99]$. Figures 9a, 9b, and 9c show the results when using bidirectional interaction and Figures 9d and 9e show the effect of unidirectional interaction. The red graph shows the only sub-population $i$ with initial infections.

For very low or non existent interaction from the initial sub-population $i$ to the others, the disease starts to eradicate as individuals recover and not enough new infections are caused by other sub-populations as can be seen in Figures 9c and 9e for low bidirectional interaction and unidirectional interaction that does not take interactions from $i$ to other sub-populations into account.

Comparing high and medium bidirectional interaction (Figures 9a and 9b) we observe that there are slight variations in the number of infected in each sub-population and a greater delay for $\rho^{medium}_{bidirectional}$ due to the different interaction strengths and an epidemic occurs.

Finally, comparing unidirectional interaction from $i$ to other sub-population without interaction back, the disease eradicates in the initial sub-population $i$ and causes small epidemics in sub-populations before they also eradicate.

## 3.3 Dynamics in network model

This section will present and discuss the experimental results obtained from simulations of a stochastic SIR network model implemented with `NDLib` on three different types of graph networks (Barabasi Albert, Watts-Strogatz, Erdos-Reyni).

Also, to visualize the whole process of how disease spread in the real network, three real simulations are finished as Figure 22, Figure 23, and Figure 24 in **APPENDIX**. Where two infected nodes are selected at the start, they show phase dynamics during 30 days. It exhibits different properties of the different network models. For the Barabasi Albert network, most nodes are converging into the central point, owing to a high clustering phenomenon. In addition, the average shortest path is low, which means the infected can spread via fewer people to anyone. The Watts-Strogatz net-

work also performs a similar clustering phenomenon, but the distribution of nodes tends to scatter towards some small groups. So the average shortest path is not so good, representing the infected is less likely to spread far nodes. Erdos-Reyni has a bad clustering since many nodes are very far from the centre and only rely on one edge connecting to others. The random distribution indicates heterogeneity in the number of contacts and a lack of clustering, which means that it is hard to predict the spreading of infection.

The most effective measures to stop or slow down an epidemic are early quarantine and early vaccination. Hence special care is given about measuring the early behaviour of the epidemic.

### 3.3.1 Effect of network initialization

First, we discuss the impact of network initialization, namely the choice and number of nodes initially infected. Trivially, a more significant fraction of infected nodes will lead to a faster spread of infections. Therefore, we focus on the Effect of different choices of initially infected nodes $I_0$ for different types of networks while keeping $|I_0| = 0.1 \cdot N_{pop}$ fixed.

We will discuss the following strategies for the choice of nodes in $I_0$:

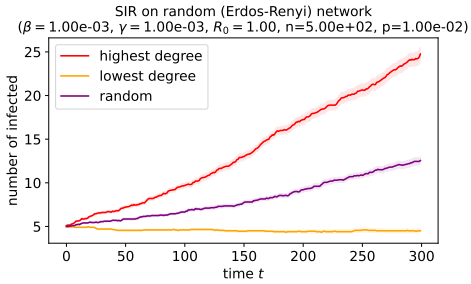
1. Choose nodes with the *highest degree*, we expect the disease to spread faster due to the high number of neighbours
2. Choose nodes with the *smallest degree*, we expect the disease to spread slower or go extinct due to the low number of neighbours
3. Choose nodes *randomly*, which will be a compromise between 1 and 2 as for large sample sizes, the expectation will approximate the mean node degree in the graph. However, for smaller samples, there can be significant variation.

Figure shows the effect of the different initialization on all three types of graphs. All simulations were performed for the same disease with $\beta = 0.001, \gamma = 0.001, R_0 = 1$ on a population size of 500 ignoring demography or imports. Note that the absolute number of infections is much more significant for the 10b network, as $\langle k^2 \rangle = 403.76$ is much more extensive for a scale-free graph. Focusing on the relative differences between the initialization strategies, we observe that on the Erdos-Renyi network (10a), the highest degree initialization leads to faster spread of the disease as expected. The lowest degree initialization spreads much slower, and random initialization is in between both extremes.

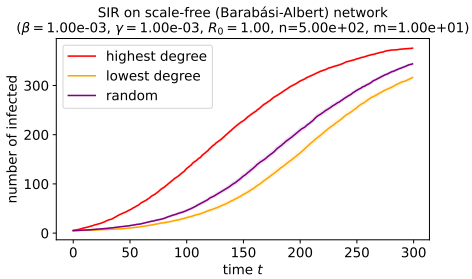
For the scale-free Barabasi-Albert network (10b), highest degree initialization also leads to faster spread, with the lowest degree and random initialization spreading approximately at the same speed.

Finally, random initialization spread the fastest for the small world Watts-Strogatz network (10c) with the other strategies spreading equally fast.

(a) Erdos-Renyi ($\langle k \rangle = 5.21$, $\langle k^2 \rangle = 29.84$)




(b) Barabasi-Albert ($\langle k \rangle = 19.60$, $\langle k^2 \rangle = 403.76$)




(c) Watts-Strogatz ($\langle k \rangle = 10.00$, $\langle k^2 \rangle = 110.00$)

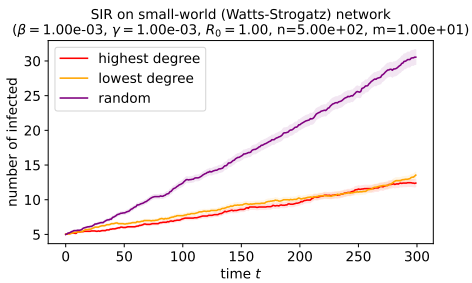



Figure 10: Effect of choosing random, lowest degree, and highest degree initialization of infected nodes $I_0$ of the same disease for three types of networks. Results were averaged over 20 simulations.

### 3.3.2 Vaccination strategies

For the following experiments, we chose random initialization of nodes with a fixed seed to evaluate the effectiveness of the proposed vaccination strategies on the different networks.

We evaluate three vaccination strategies:

1. Random vaccination

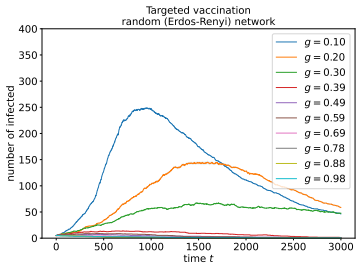



Figure 11: Targeted Vaccination on Erdos-Renyi

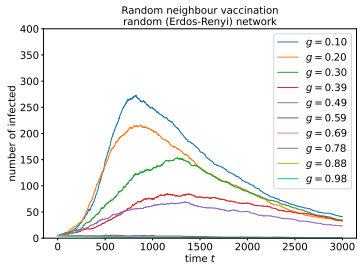



Figure 12: Random neighbour vaccination on Erdos-Renyi

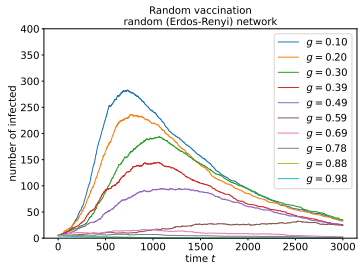



Figure 13: Random vaccination on Erdos-Renyi

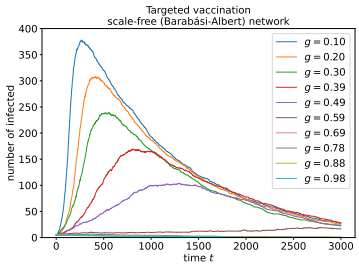



Figure 14: Targeted Vaccination on Barabasi-Albert

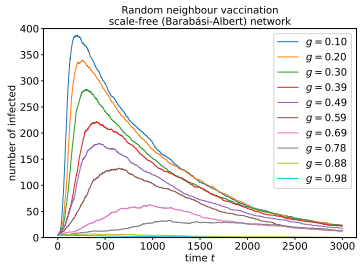



Figure 15: Random neighbour vaccination on Barabasi-Albert

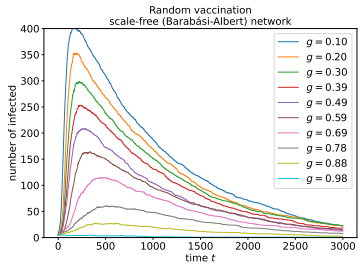



Figure 16: Random vaccination on Barabasi-Albert

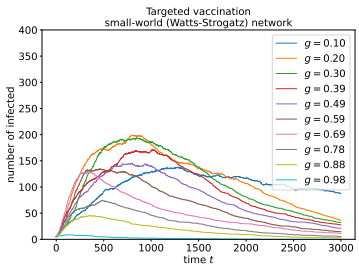



Figure 17: Targeted Vaccination on Watts-Strogatz

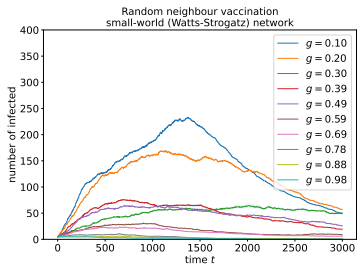



Figure 18: Random neighbour vaccination on Watts-Strogatz

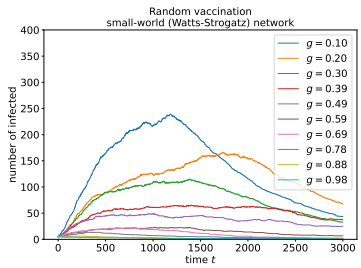



Figure 19: Random vaccination on Watts-Strogatz

Figure 20: Results for vaccination of random, highest degree, and random neighbour nodes of the same disease for Erdos-Renyi ($\langle k \rangle = 4.87$, $\langle k^2 \rangle = 29.84$, $g_c = 0.95$), Barabasi-Albert ($\langle k \rangle = 19.60$, $\langle k^2 \rangle = 403.76$, $g_c = 0.98$), and Watts Strogatz ($\langle k \rangle = 10.00$, $\langle k^2 \rangle = 110.00$, $g_c = 0.97$). Parameters for the SIR models are $N_{pop} = 500$, $\beta = 0.003$, $\gamma = 0.001$, with $R_0 = 3.00$ and $I_0 = 0.01$ initially infected. Each vaccination strategy chooses precisely a fraction $g$ unique individuals. All results were averaged from 3 repeated simulations.

2. Targeted vaccination
3. Random neighbour vaccination

A prerequisite for strategy 2 is that the node degrees in the network are known, such that the *hyper-spreaders* – nodes with many neighbours can be targeted for vaccination. Strategies 1 and 3 are of high practical interest, as they do not require global network information such as the node degrees.

Considering the experimental results from observing the early spreading dynamics for different initialization of infected nodes, we hypothesize that *random* and *random neighbour* vaccination likely works best for Watts-Strogatz networks. In contrast, *targeted vaccination* likely works well for Barabasi-Albert and Erdos-Renyi networks.

For all experiments, the fraction of initially infected is $I_0 = 0.01$ and a fraction $g$ of unique susceptible individuals is vaccinated. Note that *uniqueness* is an important assumption, as a naive implementation of random neighbour vaccination is highly likely to

choose the same high degree neighbours many times, leading to a much lower number of vaccinated individuals and incomparable results. A correct implementation of random neighbour vaccination must therefore repeat the random selection process until a unique fraction $g$ is found. While computationally more expensive, the assumption is well justified as individuals are not vaccinated many times under real world conditions.

We use the same parameters for SIR, $N_{pop} = 500$, $\beta = 0.003$, $\gamma = 0.001$, with $R_0 = 3.00$, yet the absolute numbers of infected individuals differs significantly between the different types of networks due to their properties. Since we will not compare absolute numbers of different types of networks, but rather focus on the differences between vaccination strategies per network, we will ignore this fact.

First, it shall be noted that the theoretical critical vaccination percentage $g_c$ is provided by

$$g_c = 1.0 - \frac{\langle k \rangle}{\langle k^2 \rangle R_0} \tag{7}$$

Figures 11, 12, and 13 show the effects of targeted vaccination, random neighbour vaccination, and random vaccination for the Erdos-Renyi network. The theoretical critical vaccination percentage is $g_c = 0.95$. For the Erdos-Renyi network, targeted vaccination performs very well. While for low fractions of targeted vaccinations $g \leq 0.1$ there is no significant difference between vaccination strategies, for $g = 0.39$ an epidemic is mitigated. In comparison, the peak number of infections for $g = 0.39$ using random neighbour and random vaccination are around 100 and 150 out of 500, respectively. Random neighbour vaccination similarly decreases the number of infected over random vaccination, albeit less pronounced.

In general, a clear trend of slowing down the spread of the disease can be seen for targeted and random neighbour vaccination, with peak infections decreasing and reaching their maximum at later points in time.

Similarly, Figures 14, 15, and 16 show the effects on the Barabasi Albert network with a theoretical critical vaccination percentage $g_c = 0.98$. Like for the Erdos-Renyi network, targeted vaccination performs best and random neighbour vaccination performs better than random vaccination but worse than targeted vaccination. We observe that peak infection numbers are reduced and that the curve is flattened. Also, there is no significant difference for very low fractions $g \leq 0.3$.

In contrast, Figures 17, 18, and 19 show the effects on the Watts-Strogatz network with a theoretical critical vaccination percentage $g_c = 0.97$. Unlike the previous networks, random and random neighbour vaccination perform roughly equally well, while targeted vaccination performs equal or worse across all values of $g$. This result is also in-line with our hypothesis. One possible explanation for this could be that nodes with highest degrees tend to have more long range connections which introduce local infections in spatially distant neighbourhoods – breaking containment.

Finally, it must be noted that due to the stochastic process of choosing the nodes to vaccinate as well as the stochastic nature of the SIR simulation, more repetitions would have been better to increase the reliability of the results.

# 4 DISCUSSION AND CONCLUSION

We looked at the implications of three significant changes to the standard deterministic SIR model on model dynamics. To begin, stochasticity was modeled into the dynamics using an event-driven technique, which was accomplished using Gillespie's Direct Algorithm. The development of variability, negative covariances, enhanced transients, stochastic resonance, and extinctions were described as five characteristic properties of stochastic SIR models. Furthermore, it has explored the first passage time and critical community size influence combined with extinction.

Secondly, by adopting a meta-population model and exploring the influence of the number of populations and the level of contact on the disease spreading dynamics, the simple SIR model's assumption of homogeneous population mixing is loosened. The interaction factor $\rho$ determines how a disease spreads among the designated sub-populations. Lower values of $\rho$ makes the epidemic delay across sub-populations and can reduce the amplitude of the infection peak, whereas by Pearson correlation analysis, they only have medium linear relationship.

Thirdly, using different types of networks with different characteristics, we showed that spatial and social forces dramatically impact the dynamics of disease transmissions with otherwise identical parameters and thereby demonstrated the added possibilities for more realistic and highly detailed modeling on a per individual basis. We explored the early time behaviour with different initializations of infected individuals and used the observations to design three different vaccination strategies and evaluate their effectiveness on the tree types of graph networks under consideration.

Finally, in the experiments presented, we highlighted the special dynamics that underlie different kinds of stochastic SIR models. We conclude that sufficient problem sizes and sensible parameterization allow more relevant insights and an outstanding level of

detail over the simpler deterministic models. However, this gained flexibility comes at the price of computational complexity. More precisely, simulating a large graph network required a lot of working memory and many CPU cycles. Moreover, to obtain reliable results, simulations should be performed multiple times. Compared to numeric integration of deterministic models, this is a significant overhead.

# APPENDIX

See them in the next page.

# REFERENCES

Keeling, M. and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.

Newman, M., Barabási, A.-L., and Watts, D. (2006). *The structure and dynamics of networks*, volume 9781400841356 of *Princeton studies in complexity*. Princeton University Press, Princeton.

Weber, T., Hurley, J., and Adăscăliței, D. (2021). *COVID-19: implications for employment and working life*. Employment and labour markets. Publications Office, Luxembourg.

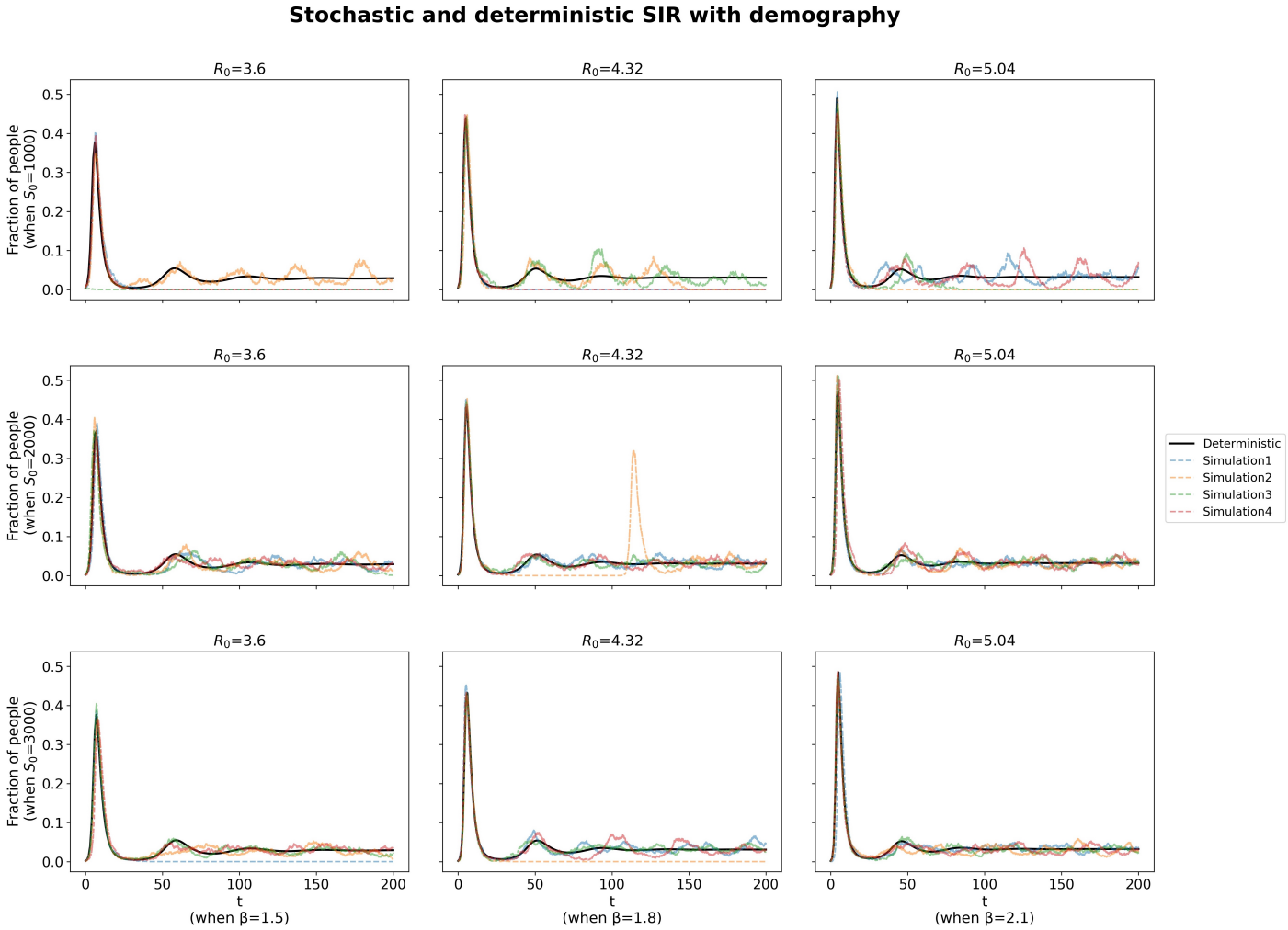



Figure 21: Four simulations of stochastic SIR model with demography
($S_0$ = [1000, 2000, 3000], $I_0$ = 5, $R_0$ = 0, β = [1.5, 1.8, 2.1], γ = 0.3, $\mu = \frac{1}{60}$, ε = 0.001, time = 200)

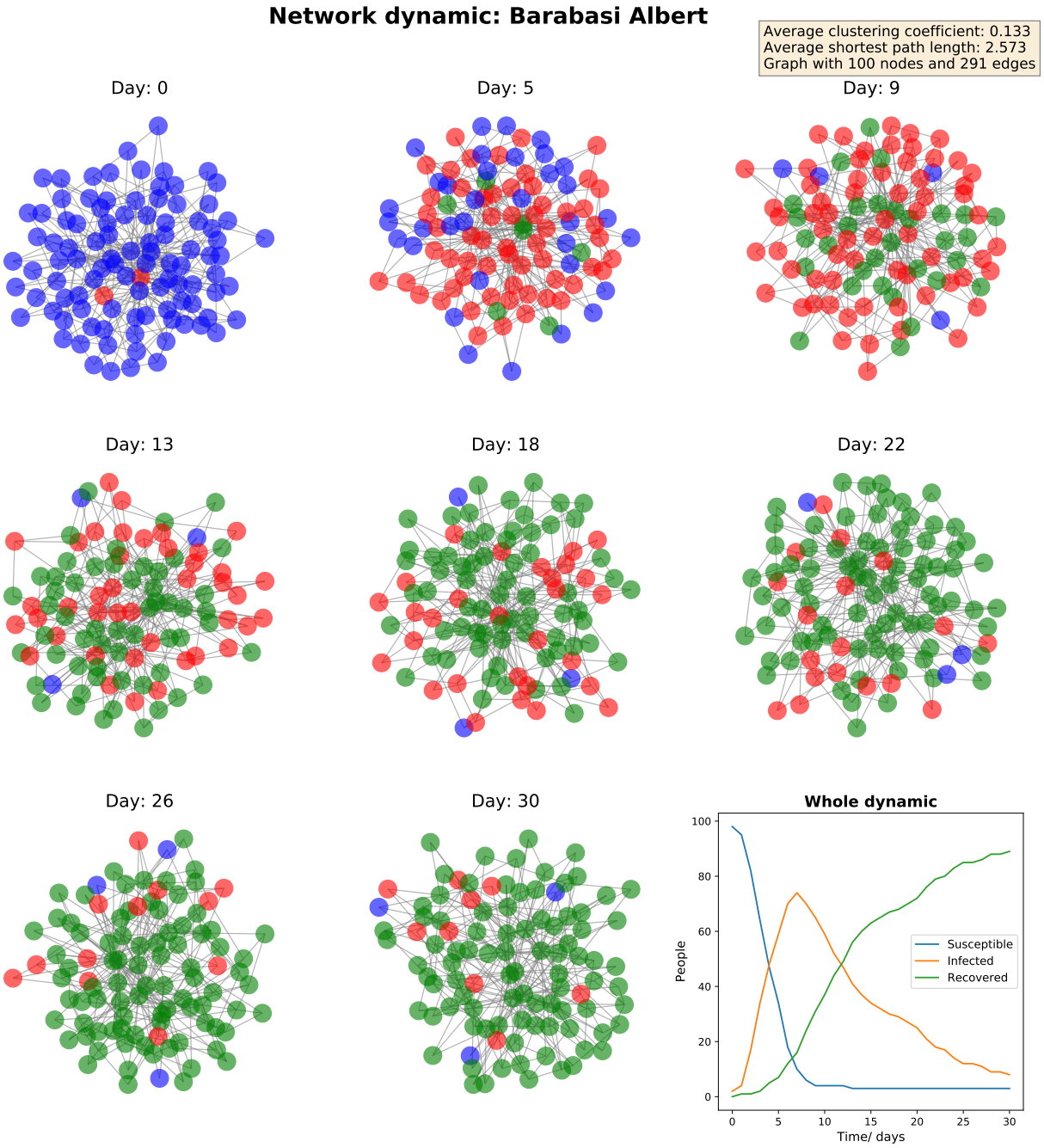



Figure 22: Real Barabasi Albert network simulation of SIR model
($S_0 = 100$, $I_0 = 2$, $R_0 = 0$, $\beta = 0.2$, $\gamma = 0.1$, time = 30)

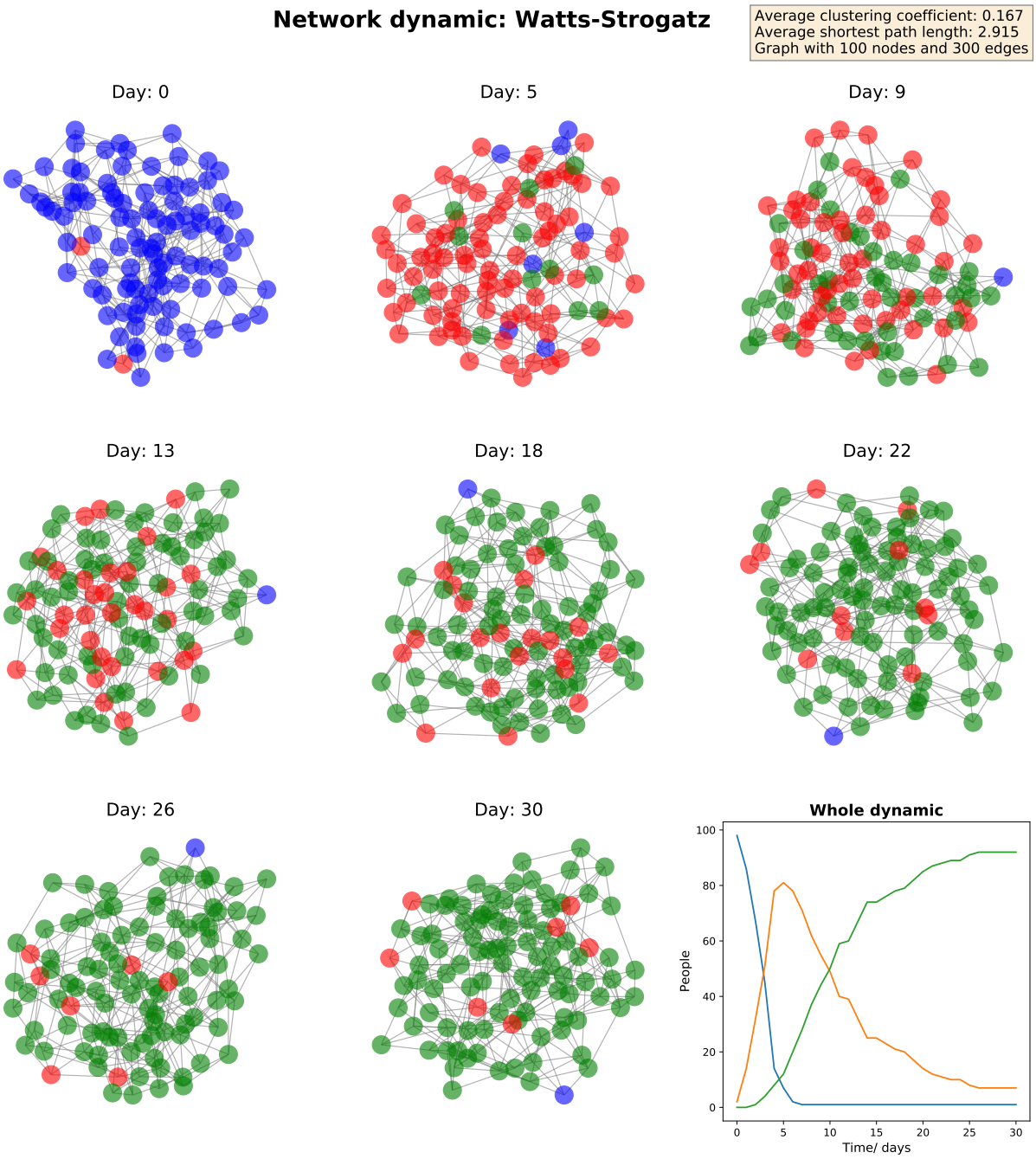



Figure 23: Real Watts-Strogatz network simulation of SIR model
($S_0 = 100$, $I_0 = 2$, $R_0 = 0$, $\beta = 0.2$, $\gamma = 0.1$, time = 30)

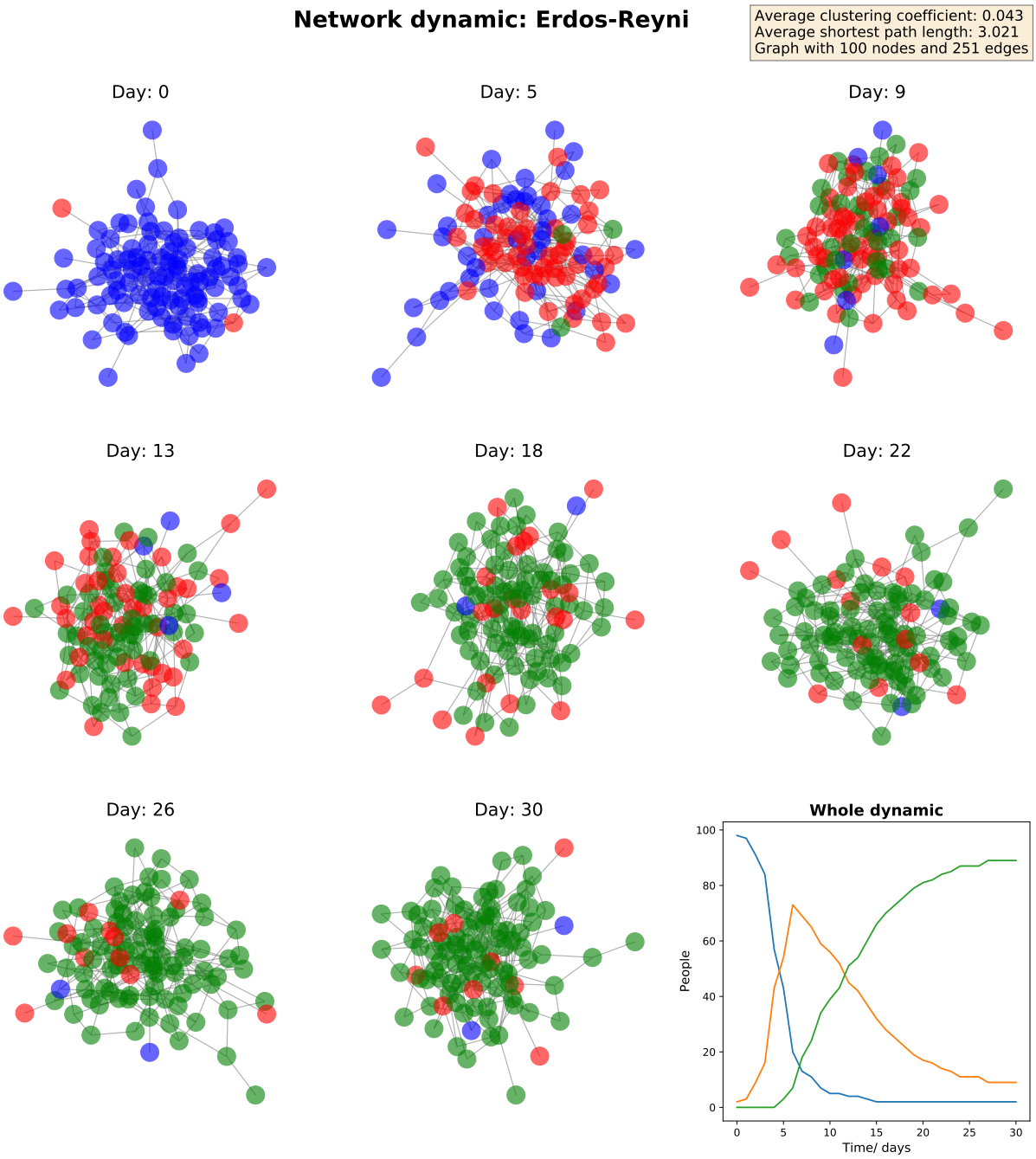



Figure 24: Real Barabasi Albert network simulation of SIR model
($S_0 = 100$, $I_0 = 2$, $R_0 = 0$, $\beta = 0.2$, $\gamma = 0.1$, time = 30)