



REPORT

Assignment 2: Stochastic and Spatial Models

Student:

Anna De Martin, Julia Rotko
Student ID: 16198921, 13603655

Lecturer: Valeria Krzhizhanovskaya

Course:

Introduction to Computational Science

October 26, 2025

Contents

1	Introduction	2
2	Theory	2
2.1	Stochastic Models	2
2.2	Network Models	3
2.2.1	Network Measures	3
2.2.2	Network Types	5
3	Numerical methods	5
3.1	Gillespie's Algorithms (GA)	5
3.2	Power Spectral Density for Resonance Valuation	6
3.3	Network Implementation	7
3.3.1	Constructing Networks	7
3.3.2	SIR Model on Networks	7
4	Results and discussion	8
4.1	Gillespie's Direct Algorithm and Stochastic Hallmarks	8
4.1.1	Implement Gillespie's Algorithm	8
4.1.2	Variability and Negative Co-variance	9
4.1.3	Stochastic Resonance and Increased Transients	10
4.1.4	Extinction events and Critical Community Size	10
4.2	Spatial Models: Epidemic Spread on Networks	11
4.2.1	Implement SIR and Simulate	11
4.2.2	Generate Networks of Equivalent Form	12
4.2.3	Simulate SIR Spread on the Network	12
4.2.4	Sociopatterns Data Network and Vaccination Strategies	14
5	Conclusions	15
A	Additional Figures	16

1 Introduction

Mathematical modeling in epidemiology has significantly advanced understanding of infectious disease dynamics, particularly since the development of the SIR model [7]. The classic SIR model is typically deterministic and assumes homogeneous mixing; however, the real epidemiological phenomena are influenced by random forces, and contacts depend on the underlying spatial structure of a population [6]. With advancements in computational methods and developments in the network science, a classic version of the SIR model can be extended to include stochasticity and modeled spatially using networks, thus closer reflecting the real dynamics of a spreading pathogen. Previous studies have applied a stochastic SIR framework for modeling of the COVID-19 pandemic [8]. Networks have also been used to assess the phase transition between persistence and extinction phases in the SIRS model (a variant of the SIR model) [10]. In another study, researchers developed an SIR model for complex networks to better represent the real-world connectivity structure, and applied it to modeling the COVID-19 pandemic [4].

In this study, we investigated the dynamics of infectious diseases with a stochastic SIR model and networks. First, we extended the classic SIR model with demography by introducing individual-based randomness and studied the key hallmarks of stochasticity [6]. Subsequently, we investigated the epidemiological spread on networks. To begin with, we explored the epidemic spread on synthetically generated networks of three types (Erdős-Rényi, Watts-Strogatz, and Barabási-Albert networks) to investigate how network topology influences the spread. Lastly, we constructed a network from socio-patterns data and assessed the effectiveness of a targeted, degree-based vaccination strategy.

This report is organized as follows: **Section 2, Theory**, discusses incorporation of stochasticity into the SIR model and five key stochastic hallmarks, as well as network measures and types. **Section 3, Numerical Methods**, explains the numerical methods and their implementation in this study. **Section 4, Results**, presents the findings and discusses their interpretation. Finally, **Section 5, Conclusions**, summarizes the work and key findings.

2 Theory

2.1 Stochastic Models

Stochastic models introduce the element of chance into infectious dynamics [6]. This randomness can be incorporated in several ways: as an external noise affecting the variables and parameters of a deterministic model, or as an intrinsic demographic noise coming from natural randomness of events happening to individuals. This report focuses on implementing individual-based randomness in an SIR model with demography, thus generating stochastic fluctuations in the infection process.

This event-driven approach represents populations as integers instead of continuous variables. Moreover, the introduction of stochastic noise requires running several simulations and studying the outcome variations to understand the short- and long-term behavior of the infection dynamics. These two features relate more broadly to the five distinct hallmarks of the stochastic models class, identified and described by Keeling and Rohani [6]:

1. **Variability between simulations:** with stochasticity, different outcomes are predicted from the same initial conditions. In an event-based approach, this is because events are drawn from probabilistic distributions, so that the exact dynamics trajectory cannot be deterministically predicted.
2. **Variances and covariances:** variability leads to larger variations in dynamics which can only be understood by looking at many simulations and at their range of behavior (variance). In addition, the number of infectious and susceptible individuals are subject

to opposing forces (e.g., transmission event $S \rightarrow S - 1, I \rightarrow I + 1$ reduces S while increases I), creating a negative covariance, between them and shifting their averages away from the deterministic equilibrium.

3. **Increased transients:** the stochastic perturbations causing the deviations from the equilibrium are opposed by deterministic restoring forces, pulling the system back to the equilibrium. This produces sustained transient oscillations that persist longer and with greater amplitude than in purely deterministic models.
4. **Resonance:** the transient dynamics lead to increased oscillations, which can occur at frequencies close to the system's natural frequency, leading to amplified fluctuations, and, hence, stochastic resonance.
5. **Extinctions:** the randomness of events for an integer-valued population increases the chance that the disease goes extinct, especially during the early stage of invasion. This can also happen when $R_0 > 1$, i.e. when extinction would not happen in a deterministic scenario.

These features are maintained in an event-based simulation approach [6], where event updates lead to increases or decreases in the number of individuals within each SIR compartment. To generate the sequence of events, we use Gillespie's Algorithm (GA) in its Direct and τ -leap version [2], [3], described in Section 3.1. The GA algorithm generally chooses the next event(s) based on the probabilities of events happening and updates the susceptible, infected, and recovered populations accordingly.

2.2 Network Models

Networks are mathematical models of relationships between individual entities used to represent complex systems [1]. A network is formally represented as a set of nodes, N , and links between them, L . Links in a network can be undirected or directed (have or not have a direction), and weighted or unweighted (have or not have different weights). In this study, we consider undirected, unweighted links.

2.2.1 Network Measures

To quantitatively describe networks, several key properties are commonly measured that provide insights into aspects of network organization, node connectivity and importance, or link structure.

1. Degree

The degree of a node i represents the number of links incident to it, and therefore it quantifies the number of direct connections a node i has with other nodes in the network. Formally, an undirected network can be represented with an adjacency matrix $A = [a_{ij}]$, where $a_{ij} = 1$ when a link exists between nodes i and j (and 0, otherwise). Then, the degree of a node is given by:

$$k_i = \sum_{j=1}^N a_{ij}. \quad (1)$$

The average degree of the network provides the measure of its overall connectivity and is given by:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (2)$$

A related property is the degree distribution, which provides the probability that a randomly selected node in a network has degree k . For a network with N nodes, the degree distribution is given by:

$$P(k) = \frac{N_k}{N}, \quad (3)$$

where N_k is the number of nodes with degree k . The degree distribution is a fundamental property as it provides insights into how connectivity is distributed throughout the network, shaping many of the behaviors of a spreading phenomena in a network.

2. Path Length

The distance, or shortest path, between node i and j is given as the minimum number of links connecting the two nodes. The average path length of a network quantifies the overall average distance between all nodes in a network, and for an undirected network is given by:

$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j>i} d_{ij}, \quad \text{with } L_{\max} = \frac{N(N-1)}{2}. \quad (4)$$

where d_{ij} is the distance between nodes i and j , and L_{\max} represents the total number of unique pairs of nodes.

3. Clustering Coefficient

Clustering coefficient is a measure that captures the fraction of the neighbors of a node that are connected to each other. For a node i , the clustering coefficient is given by:

$$C_i = \frac{2e_i}{k_i(k_i-1)}, \quad (5)$$

where k_i is the degree of the node i and e_i is the number of links between the node's neighbors.

4. Centrality Measures

(a) Degree Centrality

Degree centrality is a normalized version of the degree measure and quantifies how connected the node is, relative to the maximum possible degree in a network. It is given by:

$$C_D(i) = \frac{k_i}{N-1}, \quad (6)$$

where N is the total number of nodes and k_i is the degree of node i .

(b) Betweenness Centrality

Betweenness centrality quantifies the relative number of times a node appears on any shortest path between two nodes in the network. It is given by:

$$C_B(i) = \sum_{h \neq i \neq j} \frac{\sigma_{hj}(i)}{\sigma_{hj}}, \quad (7)$$

where $\sigma_{hj}(i)$ is the total number of times a node i lies on any shortest path between h and j , and σ_{hj} is the total number of shortest paths between nodes h and j .

(c) Closeness Centrality

Closeness centrality quantifies how close a node is to all other nodes in the network, measured as the inverse of the sum of shortest path distances to all other nodes. It is given by:

$$C_c(i) = \frac{1}{\sum_{j=1}^N d_{ij}}, \quad (8)$$

where d_{ij} is the distance from each node j to node i .

2.2.2 Network Types

To model contact structures in human populations, we employ three common network models, representing different types of connectivity patterns.

1. Random Networks (Erdős-Rényi)

A random network, also known as an Erdős-Rényi (ER) network, denoted as $G(N, p)$ consists of N nodes where each possible pair of nodes is independently connected with probability p . This model creates homogeneous networks where connections are distributed uniformly at random. The degree distribution of an ER network follows a binomial distribution:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \text{ or for larger } Ns \quad P(k) \approx \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}. \quad (9)$$

For large N , the binomial degree distribution approximates a Poisson distribution, where $\langle k \rangle$ is the average degree in the network. This distribution indicates that most nodes have a degree close to the average degree $\langle k \rangle$, resulting in little variability in degree distribution for ER networks.

2. Small-World Networks (Watts-Strogatz)

A Watts-Strogatz (WS) small-world network, denoted $G(N, k, p_{\text{rewiring}})$ is constructed by starting with a ring network of N nodes where each node is connected to its k nearest neighbors. Each edge is then randomly rewired with probability p_{rewiring} : one endpoint remains fixed while the other is reconnected to a randomly chosen node, avoiding self-loops and duplicate edges. The degree distribution is centered around the initial degree k , with increasing variance as p_{rewiring} increases. This rewiring process creates the characteristic "small-world" property of high clustering and short average path lengths, so local communities with rapid global communication. As $p_{\text{rewiring}} \rightarrow 1$, WS networks becomes similar to ERs.

3. Scale-Free Networks (Barabási-Albert Network)

A the scale-free or Barabási-Albert (BA) network is constructed with rules of growth and preferential attachment. At each timestep, the network is grown by adding a new node with m ($\leq m_0$, initial nodes) number of links, connecting it to m existing nodes in the network. The nodes are chosen according to preferential attachment, i.e., the probability that a new node connects to node i depends on the degree of node i and is given by:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (10)$$

Thus, new nodes are more likely to connect to nodes that already have more links. The degree distribution of a Barabási-Albert network follows a power-law distribution:

$$P(k) \sim k^{-\gamma}, \quad (11)$$

which results in a few nodes, called hubs, being very highly connected, while most other nodes have only a few connections.

3 Numerical methods

3.1 Gillespie's Algorithms (GA)

We implement the event-driven stochastic approach using Gillespie's method, first in its direct formulation, presented in Gillespie's first paper [2], and then in the more efficient τ -leap version [3].

Consider an SIR model with demography defined by six events: births at rate μN , infections at rate $\beta X \frac{Y}{N}$, recoveries at rate γY , and deaths in each compartment at rates μX , μY , and μZ for susceptible, infected, and recovered populations, respectively. The GA generates a sequence of events, by choosing the next event based on the single event rate (R_i). The following pseudocode outlines the implementation of the Direct Method, which uses the cumulative (sum) rate of all possible events (R_{tot}) to calculate the time until the next event (δt), and a second random number to select which event occurs.

Algorithm 1 Gillespie's Direct Algorithm for the Stochastic SIR Model

Require: Initial populations (X_0, Y_0, Z_0), rates β, γ, μ , termination time T

```

1:  $t \leftarrow 0$ 
2: while  $t < T$  do
3:   Compute individual rates  $R_1, \dots, R_6$ 
4:    $R_{tot} \leftarrow \sum_{i=1}^6 R_i$ 
5:   Draw two random numbers  $u_1, u_2 \sim U(0, 1)$ 
6:    $\Delta t \leftarrow -\frac{\ln(u_1)}{R_{tot}}$                                  $\triangleright$  Time until next event (exponential distribution)
7:   if  $\text{then } \sum_m^{p-1} < P \leq \sum_m^p$ 
8:     Update population compartments based on event p
9:    $t \leftarrow t + \delta t$ 
10:  end if
11: end while

```

The simulation time for the direct method scales with population size (N), as the cumulative rate R_{tot} increases with N , leading to shorter time steps (δt) and a higher number of required iterations to reach time T .

The τ -leap version is a more efficient implementation, because it sets a fixed time step (τ) during which multiple events are allowed to occur. This method improves computational feasibility for larger populations.

Algorithm 2 τ -Leap GA for SIR with demography

Require: Initial populations (X_0, Y_0, Z_0), rates β, γ, μ , termination time T , time step τ

```

1:  $t \leftarrow 0$ 
2: while  $t < T$  do
3:   for  $1 \leq i \leq 6$  do
4:     Compute current rate  $R_i$ 
5:     Sample event counts:  $M_i \sim \text{Poisson}(\max(R_i \tau, 0))$ 
6:   end for
7:   Update populations  $X, Y, Z$ , based on  $M$ 
8:    $t \leftarrow t + \tau$ 
9: end while

```

The fixed time step τ can be used as a control parameter for stochastic noise. A very small τ shows highly stochastic dynamics close to the Direct Method, while a large enough τ averages out random fluctuations, leading to smoother, closer-to-deterministic dynamics. But if τ is too large, too many events per leap occur. Then, τ must be carefully chosen to balance accuracy, computational time and the necessary stochastic element. Given this discretionary nature of τ , we prefer the direct method throughout the project, as long as computationally feasible.

The stochastic results were compared to deterministic trajectories obtained by solving the ordinary differential equations (ODEs) for the classic SIR with demography approach through the `solve_ivp` solver from the `scipy.integrate` library, which implements Runge-Kutta numerical integration methods (specifically, by default RK45).

3.2 Power Spectral Density for Resonance Valuation

The addition of stochastic noise amplifies the system's natural oscillations, often leading to larger amplitude fluctuations than those observed in the deterministic case. To study this

resonant behavior, we use a Power Spectral Density (PSD) analysis. The PSD describes how the power (or intensity) of a signal is distributed over different frequency components, making it useful for identifying dominant oscillation frequencies in stochastic systems.

In this report, the PSD is used to quantify the strength of stochastic oscillations and compare them to the system's theoretical natural frequency. To calculate it, we use the `welch` function from the `scipy.signal` library. The Welch method works by splitting the signal into overlapping segments. Each segment is smoothed using a window function, which gradually reduces the signal near the edges to prevent artificial high-frequency spikes. In our case, the default Hann window is used. Then, the Fast Fourier Transform (FFT) is applied to each segment, which decomposes the signal into its frequency components. The PSD is obtained by averaging the results across all segments, giving a smoother and more reliable estimate of the signal's power across frequencies. This Welch approach is especially helpful for noisy or non-stationary signals.

3.3 Network Implementation

3.3.1 Constructing Networks

The NetworkX library is a Python library that provides tools for creation and analysis of complex networks [5]. Using NetworkX, we generated three types of synthetic networks to analyze structural differences between different network types., by fixing either the number of nodes N and or the average degree k_{avg} .

Erdős-Rényi network was generated using the `erdos_renyi_graph` function with N nodes and controlling the connection probability $p = k_{avg}/(N - 1)$.

Watts-Strogatz network was constructed using the `watts_strogatz_graph` function, which requires N nodes, the number of k nearest neighbors to which each node is connected, and a rewiring probability $prewiring$.

Barabási-Albert network was generated with the `barabasi_albert_graph` function with N nodes and $m = k_{avg}/2$ links added per node, based on preferential attachment.

To describe the structural properties of the constructed networks, we computed several key network measures. Nodes degree distributions and their variance were extracted using `G.degree()` method and degree centrality was calculated with `nx.degree_centrality()`. We also measured other metrics with the following functions: average clustering coefficient (`nx.average_clustering()`), general connectivity through the shortest path length (`nx.average_shortest_path_length()`), and betweenness centrality (`nx.betweenness_centrality()`).

The NetworkX library was also used to construct the empirical network from socio-patterns data. For this network, a link was added for every pair $[i, j]$ of an adjacency data matrix for which $A[i, j] > 0$ (otherwise, no link). The final network consisted of 374 nodes and 1265 links.

3.3.2 SIR Model on Networks

We simulated the epidemic spread with the SIR model using the NDlib (Network Diffusion Library) Python library, which provides a method for simulation and analysis of the diffusion processes on complex networks [9]. For both synthetic and empirical networks the SIR model was configured using the NDlib `ep.SIRModel()` function with parameters of transmission rate β , recovery rate γ , and the initial fraction of infected individuals.

To study the spread on synthetically generated networks, we used the `model.iteration_bunch()` function to execute all time steps at once and obtain a complete epidemic trajectory. On the other hand, for vaccination strategy experiments, we used the `model.iteration()` function to iterate the model one step at a time. The stepwise approach enabled dynamic intervention by assessing the statuses of the nodes for testing with `model.status.values()` before every iteration step.

4 Results and discussion

4.1 Gillespie's Direct Algorithm and Stochastic Hallmarks

4.1.1 Implement Gillespie's Algorithm

As mentioned above in Section 3.1, different formulations of the GA exist. In this project, we initially implement the Direct Method, as it offers a good balance between simplicity and computational efficiency, especially for larger population sizes N . Using fixed infection period $\frac{1}{\gamma}$ and birth rate μ , but varying β and N , we obtain the first stochastic realizations of the SIR model.

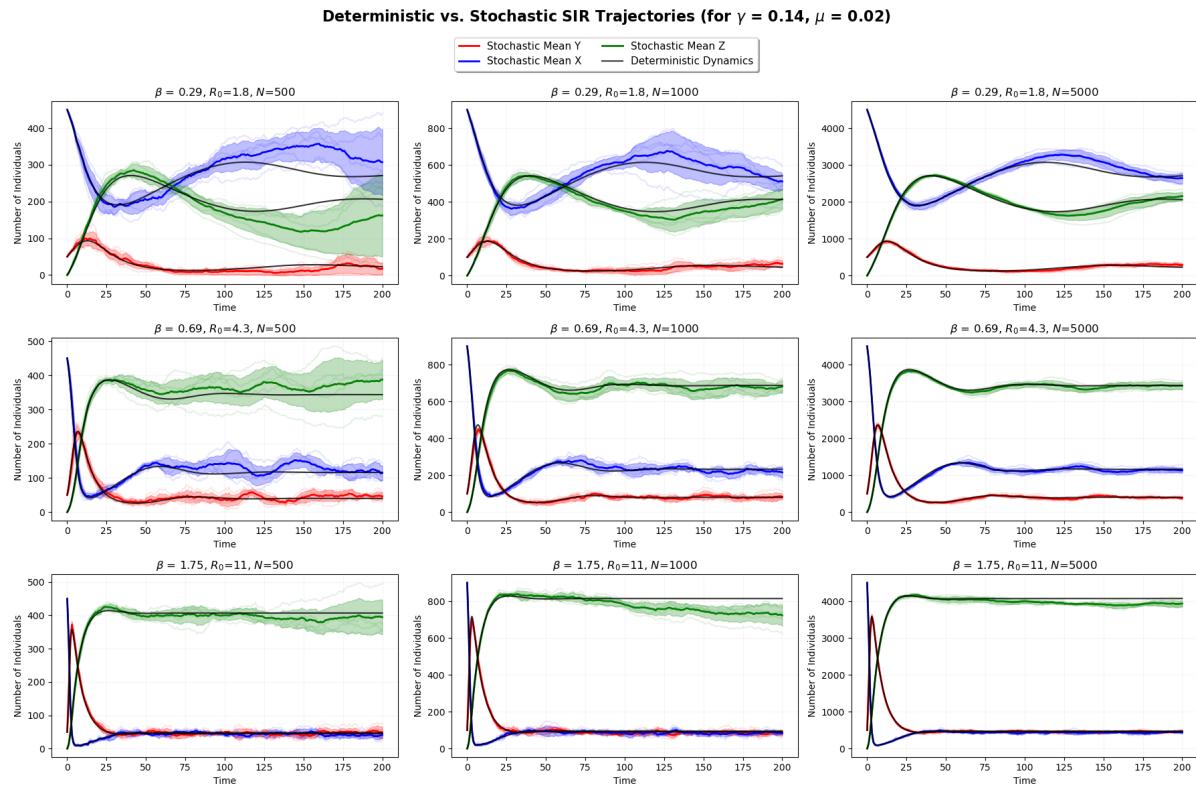


Figure 1: Comparison between deterministic trajectories (black) and stochastic ones (colored) for different values of population size $N \in [500, 1000, 5000]$ (rows) and transmission rate $\beta \in [0.29, 0.69, 1.75]$, and thus $R_0 \in [1.8, 4.3, 11]$ (columns). Other parameters are fixed at $\gamma = 1/7$, $\mu = 1/60$, with initial infected fraction $i_0 = 0.1$ for each population size. Faded lines represent five individual stochastic simulations, the bold line shows the mean across realizations, and the shaded area indicates the mean \pm standard deviation

As N increases, stochastic noise becomes smaller, and the dynamics progressively resemble those of the deterministic ODEs model. This is expected because both population size N and model parameters (e.g., β) determine the relative weight of random individual events. When N is small, each random event (e.g., one infection or one recovery) represents a large relative change in the system's state, leading to a highly noisy trajectory. Conversely, for large N or high total reaction rates (R_{tot}), events occur very frequently and δt becomes small, resulting in trajectories that appear nearly continuous.

We also explore the τ -leap version of the GA, as a way to control noise levels. By increasing the fixed time step τ , we coarsen the simulation, smoothing out random fluctuations by averaging events that occur within each interval. However, excessive coarsening (large τ) reduces accuracy, as too many events are assumed to occur instantaneously (Appendix, Figure A.1).

4.1.2 Variability and Negative Co-variance

We investigate the variance in the number of infected individuals for the same population size N and basic reproductive ratio R_0 , as in Figure 1. The variance was normalized to allow comparability, and therefore reflects the variance relative to the population size. The results are shown in the Figure 2.

The findings indicate that the variance decreases as N increases, for any value of R_0 . This decrease in variance is a one of the five hallmarks of stochasticity [6]: as the population size N increases random fluctuations become proportionally smaller, leading to lower relative magnitude of variability. Therefore, in larger populations the model tends to approach a deterministic dynamic, whereas in smaller populations stochastic forces dominate.

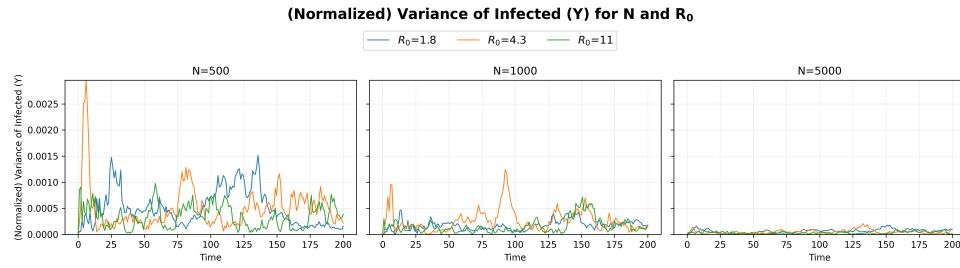


Figure 2: Variance of infected individuals Y for different population sizes N and values of the reproductive ratio R_0 . Variance is normalized by population size to facilitate comparison across different N . Largest fluctuations are shown at small population sizes and low transmission rates.

We also investigate the covariance between the number of susceptibles X and the number of infected Y . The results are displayed in the Figure 3. Again, the covariance was normalized with respect to the population size N to allow comparability, and thus it reflects a relative covariance.

The findings show that negative covariance between X and Y decreases as the population size N increases. This is explained with the fact that in smaller populations stochastic forces tend to dominate and thus the relatively higher noise in the system pushes X and Y in opposite directions. In contrast, the magnitude of covariance decreases as the population size N increases. In addition, we observe that for lower values of R_0 , the system is more sensitive to noise, which leads to the negative covariance being more pronounced. This has an important implication for the average dynamics: as the covariance becomes more negative, it effectively reduces the rate at which new cases arise, leading to a marginally higher mean of susceptibles over time.

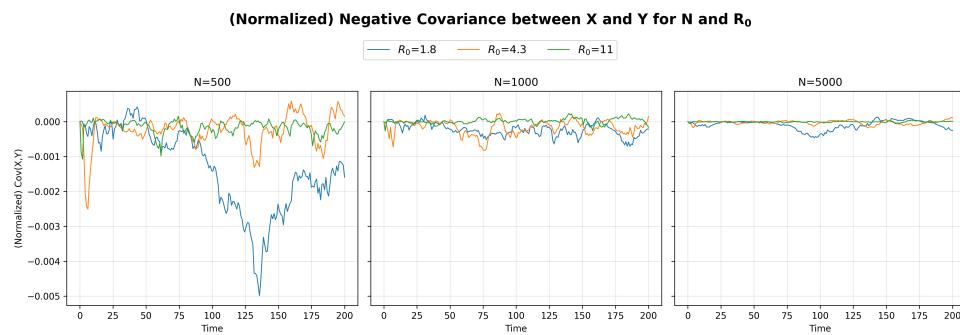


Figure 3: Covariance between infected Y and susceptible X individuals for different population sizes N and values of the reproductive ratio R_0 . Covariance is normalized by population size and remains negative throughout: as infections increase, susceptibles decrease.

4.1.3 Stochastic Resonance and Increased Transients

Focusing on the oscillatory dynamics of the infected fraction near the endemic equilibrium (I^*), we investigate the transients and resonance phenomena. In the deterministic model, the fixed point acts as a stable spiral for $R_0 > 1$, and the infected trajectory oscillates toward equilibrium (black lines in Appendix, Figure A.2). When stochastic noise is introduced, it can amplify the intrinsic oscillations of the system, exhibiting stochastic resonance, while also opposing the deterministic forces causing transients-like returns to equilibrium

To observe how this amplification interacts with the population size and transmission intensity, we analyse the power spectral density, which represents how the strength of stochastic oscillations is distributed across different frequencies and allows comparison with the system's natural frequency. Each panel in Figure 4, corresponds to a different R_0 , with the dashed line marking the theoretical natural frequency f_n . For lower transmission rate $\beta = 0.29$ and reproductive number $R_0 = 1.8$, we observe the strongest evidence of stochastic resonance, where a sharp spectral peak aligns closely with the natural frequency: random fluctuations reinforce the system's intrinsic oscillations. At intermediate transmission levels ($\beta = 0.69$ and $R_0 = 4.3$), all population sizes show weaker peaks near the natural frequency, and this effect diminishes as N increases. For high transmission rates ($\beta = 1.75$ and $R_0 = 11$), stochastic fluctuations primarily add noise rather than amplifying specific frequencies. The high infection rate dominates stochastic effects, producing deterministic-like cycles with minor random perturbations. Overall, resonance occurs under conditions of moderate transmission intensity and smaller population sizes, where stochastic fluctuations are strong enough to amplify the system's natural oscillations.

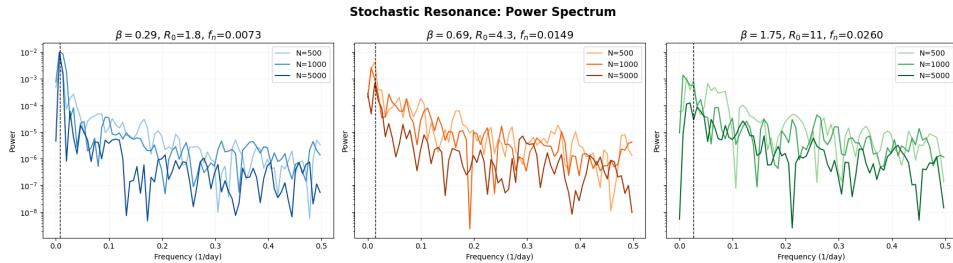


Figure 4: Power spectral density for the three different R_0 values across population sizes. Vertical dashed lines indicate the theoretical natural frequency (f_n). Stochastic resonance effects are shown for lower values of R_0 and smaller population sizes

The phenomenon of increased transients reflects the interplay between the stochastic and deterministic forces. Figure 5 shows the mean maximum deviation from the endemic equilibrium across stochastic realizations, for different values of N and R_0 . At low transmission ($R_0 = 1.8$), the amplitude remains high and relatively constant across all population sizes. At moderate transmission ($R_0 = 4.3$), the amplitude decreases monotonically as population size increases. For high transmission ($R_0 = 11$), amplitudes are smallest overall and also decrease with population size, even if the effect is less pronounced. These results show that near the epidemic threshold (lower R_0 , hence β), the system exhibits persistently large stochastic fluctuations and it is less sensitive to population size, while higher transmission rates produce smaller relative fluctuations that diminish predictably as populations grow larger.

4.1.4 Extinction events and Critical Community Size

To analyze the fifth and final hallmark of stochasticity, we compute the probability of extinction as the proportion of simulations where the infected population dropped below one individual within the observation period. We run multiple simulations across different population sizes N and three levels of R_0 , by using the τ -leap method to reduce computational time, setting

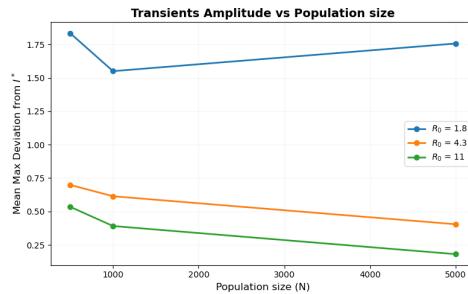
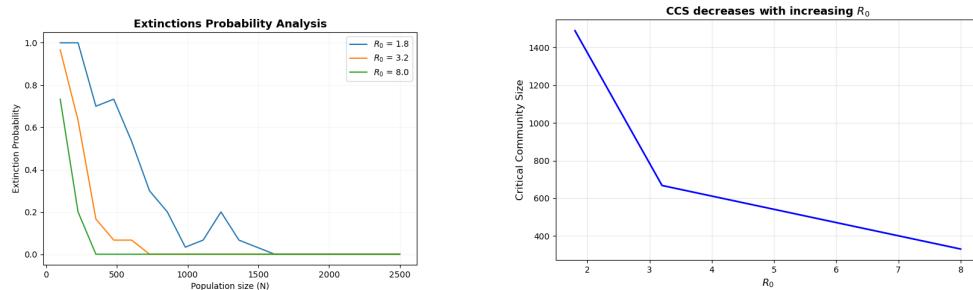


Figure 5: Mean maximum deviation from equilibrium for the three transmission rates as a function of N . Large fluctuations are present at lower R_0 values and, overall, decrease with network size

$\tau = 0.1$, as suggested by [6].

Figure 6a shows the probability of early extinction across the tested N range. For all R_0 values, the probability of early extinction decreases rapidly as N increases, as discussed by [6]. The lowest $R_0 = 1.8$ exhibits the highest extinction probability (starting at 1.0) and requires the largest population to reach a near-zero extinction rate. Conversely, for higher R_0 values, extinction probability reaches zero more quickly (already for $N \geq 500$).

To explore the relationship between R_0 and population size further, we estimate the Critical Community Size (CCS), defined as the minimum population size that does not suffer disease transmission. We obtain it by interpolating the extinction frequency curve to find the population size N where we expect one extinction event per year. Figure 6b shows that CCS decreases with increasing R_0 : a stronger disease requires a smaller population to survive the initial stochastic noise.



(a) Probability of disease extinction for each transmission rate ($R_0 \in [1.8, 3.2, 8]$) and N ranging from 100 to 2500, calculated as the fraction of simulations (out of 30) where the infected population fell below 1 within 365 days. Smaller populations exhibit higher extinction, with the effect most pronounced at lower R_0 values.

(b) Critical community size (CCS) as a function of basic reproductive number, defined as the population size at which one extinction event occurs per year, and estimated by linear interpolation. CCS decreases with increasing R_0 , indicating that diseases with higher transmission rates can persist in smaller populations.

4.2 Spatial Models: Epidemic Spread on Networks

4.2.1 Implement SIR and Simulate

To investigate the spread of diseases on various network models, we initialize a simulation framework that will work for two initial conditions: a standard approach where a random fraction of nodes is initially infected, and a targeted approach where specific individuals, according to centrality measures, are initially infected. The analysis focuses on the three networks classes described in Section 2.2.2: the Barabási-Albert (BA) scale-free network, the Watts-Strogatz (WS) small-world network, and the (Erdős-Rényi) (ER) random network.

4.2.2 Generate Networks of Equivalent Form

Before simulating the SIR dynamics, we first analyze the networks' structures, since their properties can provide initial insights into infection spreading dynamics. We focus on statistics that can help us understand how network structure influences epidemic spreading, such as centrality measures, degree variance, clustering coefficient and average path length. For instance, knowing whether nodes are highly connected or at a short distance will allow the disease to spread faster.

The comparative plot (Figure 7) shows that as the probability p increases, in ER graphs average degree and clustering grow linearly while path lengths and betweenness centrality drops, creating a homogeneous population where all nodes have similar roles; WS networks, created by rewiring an initial ring structure with probability p_{rewiring} , maintain high local clustering at low rewiring, while reducing paths length as rewiring increases; BA networks, grown via preferential attachment, exhibit the most heterogeneous structure with increasing degree variance and low clustering, creating highly connected hub nodes while the majority of nodes remain poorly connected.

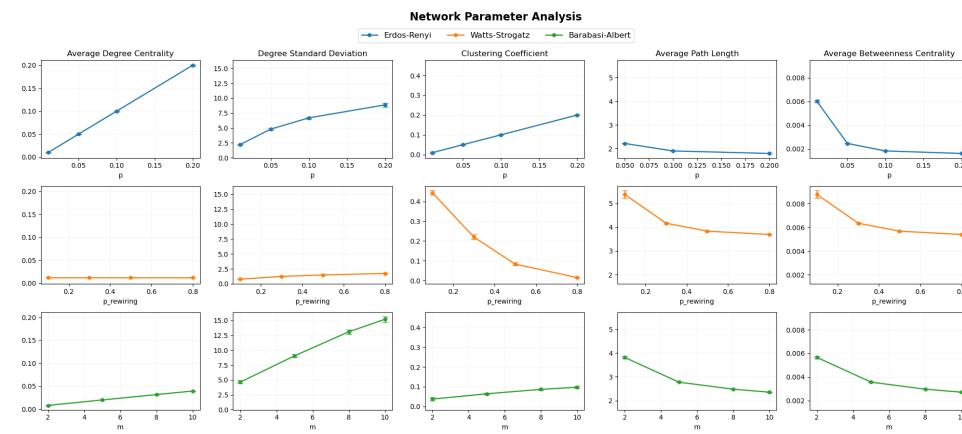


Figure 7: Structural properties across the three graph models. Ten instances of each network model with 1000 nodes were generated by varying parameters: connection probability $p \in [0.01, 0.05, 0.1, 0.2]$ for ER, rewiring probability $p_{\text{rewiring}} \in [0.1, 0.3, 0.5, 0.8]$ (joining each node to its $k = 6$ nearest neighbors) for WS, and preferential attachment parameter $m \in [2, 5, 8, 10]$ for BA. The five network measures provide insights on how topological features can influence epidemic spreading

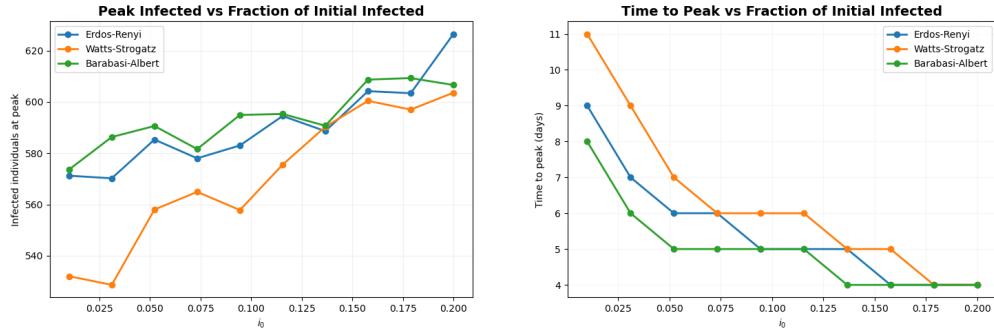
Then, infections in ER networks should spread uniformly across nodes, through localized outbreaks or a rapid global spread based on the rewiring probability in WS networks; while for BA networks outbreaks dynamics are related to the presence of hub nodes, that if targeted might make interventions more effective but also create vulnerability to rapid initial spread.

4.2.3 Simulate SIR Spread on the Network

Given this background analysis, we perform SIR simulations on comparable networks, with 1000 nodes and average degree $k_{\text{avg}} = 6$ (implying $p = 0.006$, $p_{\text{rewiring}} = 0.3$, $m = 3$).

We initially infect varying fractions of random individuals (i_0). Figures 8a and 8b show the number of infected individuals at peak and how long does it take for the infection to reach such peak, respectively. Overall, the WS network showed the lowest peak infection fraction and the slowest time-to-peak across most of the i_0 values. As suggested by the structural analysis, the higher clustering level in this network could act as a local barrier, making it harder for the infection to break out. The ER and BA networks showed a higher number of infected individuals at the peak and a faster pace of spread, as i_0 increased.

Next, we investigate the effect of targeted infection, simulating a scenario where the top 3 central nodes (based on degree, betweenness, and closeness centrality) are targeted and initially



(a) Peak number of infected individuals as a function of initial infection fraction. SIR simulations were run on ER, WS, and BA networks ($N = 1000$ nodes, $\beta = 0.2$, $\gamma = 1/7$) with initial infection fraction i_0 varying from 0.01 to 0.2. Peak infections increase with i_0 across all network types.

(b) Time to peak infections as a function of initial infection fraction. For the same simulation parameters, time to reach peak infections decreases with larger i_0 , as a higher number of initially infected individuals accelerate epidemic progression.

infected. We test which centrality measure provides the fastest spread on each network model, by comparing the infected curves for targeted strategy with a random selection spread. The optimal initial seeding strategy is dependent on the network type. In the ER network, targeting nodes with the highest degree centrality produces the most infections (compared to other centrality measures), probably because, despite ER homogeneity, nodes with a higher-than-average degree could act as accelerators in transmission. For WS network, targeting the highest betweenness centrality nodes results in the fastest epidemic (still, compared to the other centrality measures), as these bridging nodes connect otherwise separated communities, enabling rapid transition from local clusters to global spread. Finally, in BA networks, infecting nodes with the highest closeness centrality, i.e. the well-connected hubs, maximizes both the spread rate and final epidemic size. Notably, BA was the only network type where targeted seeding consistently outperformed random seeding, producing higher infection peaks every time. This highlights how critical hubs are in scale-free networks: infecting them creates explosive outbreaks that are hard to stop, but, identifying and vaccinating these super-spreaders through vaccination would be particularly effective.

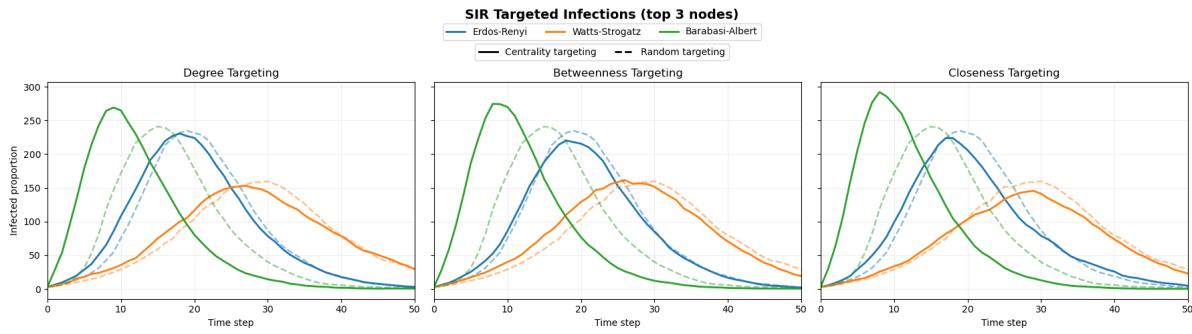
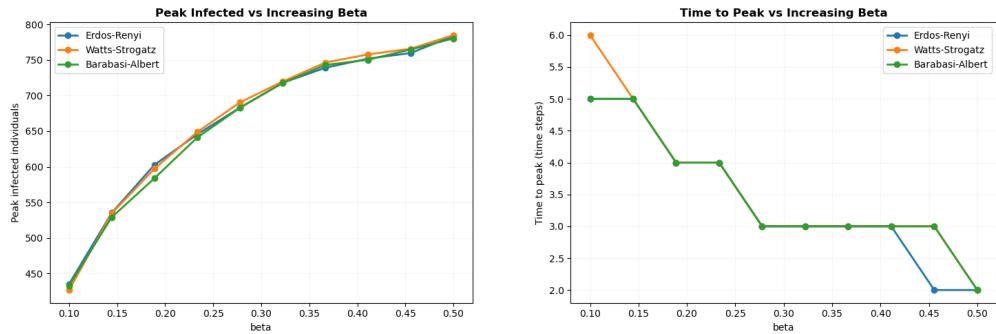


Figure 9: Truncated infection curves comparing targeted and random initial seeding strategies, with fixed parameters $N = 1000$, $\beta = 0.2$, $\gamma = 1/7$, and i_0 determined by infecting the 3 most central nodes. Targeted seeding produces higher or faster peaks, with the effect most pronounced in BA networks, because of hubs.

Finally, starting from a fixed initial infection fraction ($i_0 = 0.1$), we compare the dynamics across different infection rates (β s). As expected, in Figures 10a and 10b, increasing β accelerates infection and increases the infected peak infected in all types of networks. However, while all networks experience epidemics of similar sizes as transmission rate varies, the speed of the outbreaks initially differs, also based on the networks' properties, with the WS network

consistently exhibiting the slowest spread.



(a) Peak number of infected individuals as a function of transmission rate β ranging from 0.1 to 0.5 (starting from $i_0 = 0.1$). Peaks increase with transmission rate across all topologies.

(b) Time to reach peak infections decreases with increasing β . WS network exhibits slowest peak, for lower values of β .

4.2.4 Sociopatterns Data Network and Vaccination Strategies

We conducted vaccination experiments on the network derived from socio-patterns data. First, we visualized the network and investigated some of its main measures to assess the network type. The socio-patterns network is visualized in Figure A.3 (see Appendix). Table 1 summarizes the network's fundamental metrics. We further investigated the distributions of the degree and betweenness centrality for this network. The findings indicate that both degree and betweenness centrality distributions closely follow a power-law distribution, which are characteristic for a Barabási-Albert network.

Network Characteristics for Sociopatterns Data

Network Measure	Value
Average Degree	6.765
Degree Variance	41.784
Average Betweenness Centrality	0.005692
Average Clustering Coefficient	0.296

Table 1: Key network metrics.

Next, we performed vaccination experiments to investigate the effectiveness of a targeted vaccination strategy that prioritizes the most influential individuals (nodes) in the network. In these experiments, we compared degree-based targeted vaccination strategy with a null strategy. In the null strategy, a number of individuals (both susceptible and infected) was randomly selected for vaccination ($Vacc \in [1, 3, 5, 10]$). In the targeted strategy, individuals were first ranked according to their degree, and a number of individuals was subsequently picked from the top for testing ($Tests/Iter \in [3, 5, 7, 12]$). The testing aimed to determine each individual's infection status (infected vs. susceptible). Individuals identified as susceptible were then vaccinated, up to the limit specified by the current value of $Vacc$. In addition, the accuracy of testing was varied ($Test Accuracy = [0.5, 0.75, 1.0]$), which introduced the possibility of both false positives and false negatives in identifying susceptible and infected individuals.

The results of the experiments for different values of $Vacc$, $Tests/Iter$ and $Test Accuracy$ are presented in Figure 11. The degree-based strategy is visualized with red-shaded lines, and the null strategy is visualized with blue-shaded lines. The degree-based strategy outperforms the null strategy whenever its infected curve falls below that of the null strategy. The findings

indicate that the degree-based strategy outperformed the null strategy for most of the experimental setups, and particularly for $Vacc = 10$ and $Tests/Iter = 12$ across all levels of test accuracy. These results suggest that a degree-based targeted vaccination strategy can be more effective than a random, null strategy.

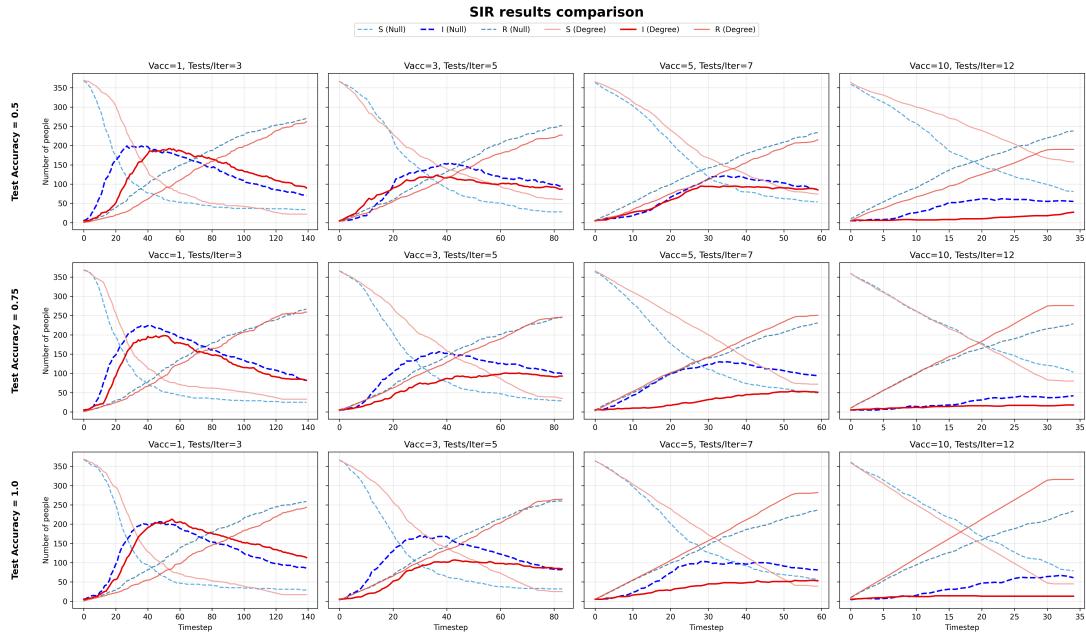


Figure 11: Results of vaccination experiments. Comparison across varying values of $Vacc \in [1, 3, 5, 10]$, $Tests/Iter \in [3, 5, 7, 12]$ and $Test Accuracy \in [0.5, 0.75, 1.0]$. The degree-based strategy outperforms the null when its infected curve (dark red) lies below that of the null (dark blue)

5 Conclusions

In this study, we investigated the dynamics of infectious diseases modeled with the SIR framework by incorporating both demographic stochasticity and spatial structure through network models. By exploring the five key hallmarks of stochasticity, we found that stochastic effects are most pronounced in small populations (low N) and near the epidemic threshold (low R_0), where demographic fluctuations dominate deterministic forces. These conditions produce high variance in epidemic trajectories, negative correlations between compartments, oscillatory behavior through stochastic resonance, large-amplitude transients, and potentially disease extinctions. As population size increases and transmission intensity grows (high β), stochastic dynamics increasingly resemble deterministic ones, with reduced relative variability and stable endemic equilibrium.

Our network analysis revealed that epidemic spreading is influenced by network structure and characteristics. We found that strategic targeting of specific individuals, whether for initial infection or vaccination, is significantly more effective than random selection in both accelerating and controlling disease spread. Targeted initial seeding strategies accelerated epidemic dynamics compared to random initialization, with most pronounced effects in scale-free networks where infecting central hubs increased both outbreak speed and size. Moreover, we found that targeting highly connected individuals effectively outperforms the null strategy in most experimental setups, highlighting the importance of prioritizing vaccination of most connected people for better control of the infection spread.

Together, these findings demonstrate that stochastic effects and network heterogeneity are essential considerations for a deeper understanding of disease dynamics. A more precise modeling, that accounts for both demographic randomness and contact structure, is fundamental if we aim to design effective intervention strategies.

References

- [1] Barabási, A.-L. (2016). *Network Science*. Cambridge University Press. <https://www.networksciencebook.com/>
- [2] Gillespie, D.T. (1977) Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, 81, 2340–2361. <https://doi.org/10.1021/j100540a008>
- [3] Gillespie, D.T. (2001) Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems. *Journal of Chemical Physics*, 115, 1716–1733. <https://doi.org/10.1063/1.1378322>
- [4] Goel, R., Bonnetain, L., Sharma, R., & Furno, A. (2021). Mobility-based SIR model for complex networks: with case study of COVID-19. *Social Network Analysis and Mining*, 11(1). <https://doi.org/10.1007/s13278-021-00814-3>
- [5] Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference (SciPy 2008)* (pp. 11–16).
- [6] Keeling, M. J., & Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press. <https://doi.org/10.2307/j.ctvcm4gk0>
- [7] Mata, A. S., & Dourado, S. M. P. (2021). Mathematical modeling applied to epidemics: an overview. *São Paulo Journal of Mathematical Sciences*, 15(2), 1025–1044. <https://doi.org/10.1007/s40863-021-00268-7>
- [8] Pájaro, M., Fajar, N. M., Alonso, A. A., & Otero-Muras, I. (2022). Stochastic SIR model predicts the evolution of COVID-19 epidemics from public health and wastewater data in small and medium-sized municipalities: A one year study. *Chaos, Solitons & Fractals*, 164, 112671. <https://doi.org/10.1016/j.chaos.2022.112671>
- [9] Rossetti, G., Milli, L., Rinzivillo, S., Sîrbu, A., Pedreschi, D., & Giannotti, F. (2017). NDlib: a python library to model and analyze diffusion processes over complex networks. *International Journal of Data Science and Analytics*, 5(1), 61–79. <https://doi.org/10.1007/s41060-017-0086-6>
- [10] Saif, M. A. (2019). Epidemic threshold for the SIRS model on the networks. *Physica A: Statistical Mechanics and Its Applications*, 535, 122251. <https://doi.org/10.1016/j.physa.2019.122251>

A Additional Figures

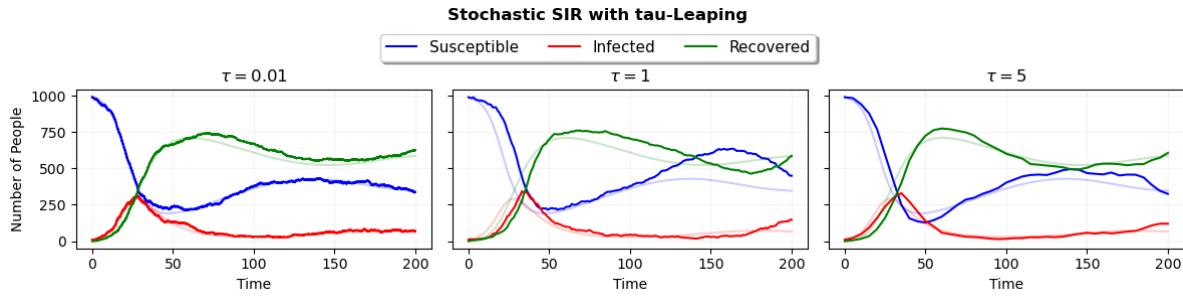


Figure A.1: Stochastic SIR dynamics with τ -leaping, for $\tau \in [0.01, 1, 5]$, over 200 time units, compared to deterministic trajectories (faded lines). As τ increases stocastics dynamics resemble the deterministic ones.

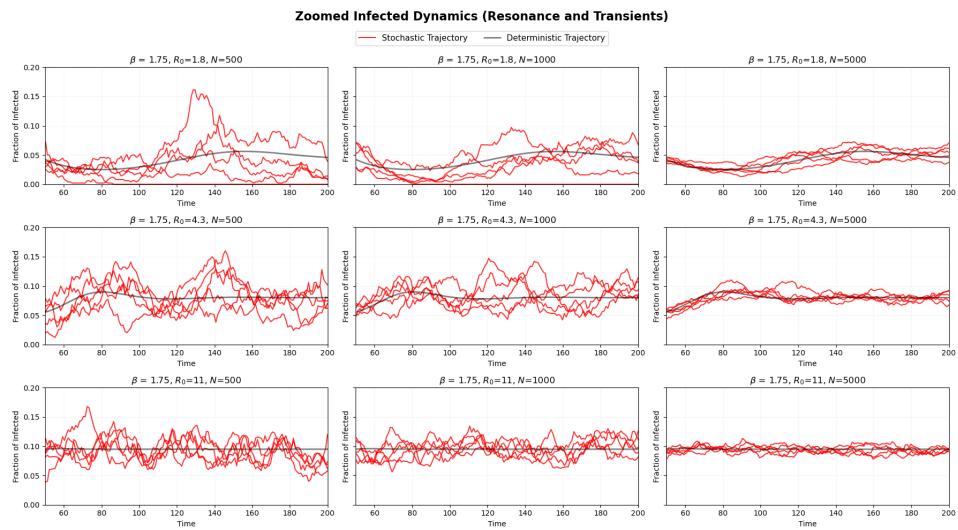


Figure A.2: Zoomed dynamics for fraction of infected across parameters combinations ($N \in [500, 1000, 5000]$ and $\beta \in [0.29, 0.69, 1.75]$). Stochastic trajectories (in red) show increasing variability and amplitude of fluctuations at smaller population sizes (top row) and lower transmission rates (left columns).

Network of Sociopatterns Data

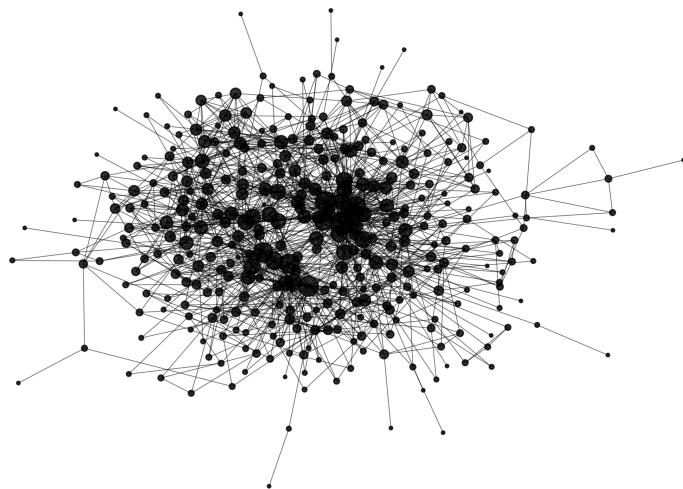


Figure A.3: Network visualization of socio-patterns data.