UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Coupled epidemic spreading on complex networks and pathogen evolutionary dynamics

Diffusione epidemica in reti complesse con dinamiche evolutive dei patogeni

Relatore

prof. Manlio De Domenico

Correlatori

dott. Sandro Azaele

dott. Valeria D'Andrea

Laureanda

Vittoria Bettio

1227238

Anno Accademico 2022/2023

Abstract

The spreading of an epidemic on the top of a complex network is a phenomenon widely studied with statistical physics's theoretical and computational tools. The evolutionary dynamics of a pathogen is another phenomenon widely studied by means of ecological models. In the real world, the evolution of a pathogen is dramatically related to the evolution of its host: in fact, co-evolution dynamics is a more adequate framework to understand this complex phenomenon. However, the coupling of epidemic spreading (due to social contacts) and pathogen evolution (including mutation, replication and competition for resources) is barely studied in the literature. This thesis work aims to tackle this challenge in the context of complex network dynamics. Specifically, we explore distinct epidemic processes, with more focus on the SIR models, and evolutionary dynamics, with focus on the replicator-mutator equation on the top of a network with varying topology, encoding increasing levels of complexity observed in empirical biosocial systems. In summary, we have built a model coupling epidemic spreading and evolutionary dynamics. We have developed rate equations and verified their agreement with Monte Carlo simulation of the coupled process. In the next future, we will investigate the critical behavior of the coupled process in terms of a few relevant control parameters, such as the pathogen mutation rate and the epidemic basic reproduction number, to better characterize this complex system.

La diffusione di un'epidemia in una rete complessa è un fenomeno ampiamente studiato dal punto di vista teorico e computazionale con strumenti di fisica statistica. La dinamica evolutiva di un patogeno è un altro fenomeno largamente esaminato per mezzo di modelli ecologici. Nel mondo reale, l'evoluzione di un patogeno è drammaticamente legata all'evoluzione nel corrispondente ospite: infatti, la dinamica co-evolutiva è un ambito più adeguato per la comprensione di questo fenomeno complesso. Tuttavia, la congiunzione di diffusione epidemica (dovuta a contatti sociali) e l'evoluzione del patogeno (includendo mutazione, replicazione e competizione per le risorse) è poco studiata nella letteratura. Questo lavoro di Tesi affronta questa sfida nel contesto di dinamiche su reti complesse. Nello specifico, vengono esplorati diversi processi epidemici, con una speciale attenzione al modello SIR, e di dinamica evolutiva, soprattutto utilizzando l'equazione replicator-mutator, su reti complesse con diverse topologie, analizzando livelli crescenti di complessità osservata in sistemi biosociali empirici. In sintesi, viene di seguito costruito un modello che abbina la diffusione epidemica con la dinamica evolutiva. Sono state inoltre sviluppate alcune equazioni differenziali che siano in grado di descrivere la dinamica accoppiata, verificandone l'accordo con la simulazione Monte Carlo. Nel prossimo futuro, si andrà ad indagare il comportamento critico di processi congiunti in termini di pochi ma rilevanti parametri di controllo come il tasso di mutazione del patogeno e il numero di riproduzione di base epidemico, per essere in grado di caratterizzare al meglio tale sistema complesso.

Contents

1 Introduction

Epidemics have long been a major concern for public health and safety. With the increasing interconnectedness of our world, the spread of infectious diseases has become an even more pressing issue. Understanding how epidemics spread and how they can be prevented is crucial for mitigating their impact and protecting vulnerable populations.
One key aspect of epidemics that has received increasing attention in recent years is the role of pathogen mutation as it spreads. Usually, an epidemic spreading is described through the basic reproduction number R_0 whose value determines the contagion state. However, if the pathogen simultaneously mutates the dynamics of the epidemic can significantly change. The epidemic transitions are influenced not only by R_0 but also by the equilibrium between the epidemic and evolutionary timing. If mutations proceed too slowly, the pathogen prevalence, that is the proportion of people in a population infected by the pathogen, might decline before a crucial mutation manifests. If mutations happen too quickly, the pathogen fails to spread again because its evolution becomes volatile, since frequent mutations can result in genetic instability. However, between these two extremes, we identify a range of circumstances in which a pathogen can spread widely.
This thesis aims to create a model that the dynamics of an epidemic described with both a Susceptible-Infected (SI) and a Susceptible-Infected-Recovered (SIR) models on top of a network, as well as the mutation of the pathogen as the epidemic spreads. Our model will be able to take into account the connections between individuals in a social network and the effect of these interactions on the spread of an epidemic. It will also be able to simulate the evolution of the pathogen as it adapts to its new host and the changes in transmission dynamics that this may bring.
Social networks can facilitate the transmission of infections by bringing individuals into close contact with each other, but they can also act as barriers to the spread of diseases by creating geographic segregation or social isolation.
Our model will be developed using a combination of mathematical and computational techniques.
Overall, the goal of this thesis is to contribute to our understanding of complex epidemiological systems on social networks and provide a valuable tool for predicting and mitigating the spread of epidemics. Hopefully, it will be able to provide valuable insights into the dynamics of the co-evolution of complex epidemiological systems on social networks and inform public health policies and strategies.

The thesis is structured as follows: in Chapters 2, 3 and 4, we review the literature on social networks, complex epidemiological systems and pathogen evolution. Chapter 5 describes the methods used to create and validate our model. Finally, in Chapter 6, we draw conclusions and suggest directions for future research.

2

An overview of networks

At its most basic level, a **network** is a group of points connected by lines in pairs. These points are called vertices or nodes, while the lines are called edges or links. The physical, biological, and social sciences contain numerous systems that can be conceptualized as networks: they can help to better understand phenomena such as the spread of diseases, the structure of the internet, the relationships between individuals, and the flow of information, among others.
Networks can be directed or undirected, the first ones are used to explicit the direction of the link between two nodes while the others do not provide information on the direction of the link, treating the relationship between nodes as a symmetric or bidirectional one. Directed networks are commonly employed in modeling systems such as transportation networks and information networks, where the direction of connections can play a crucial role in understanding the dynamics and behavior of the system. Undirected networks are often used to model systems such as friendship networks and biological networks, where the relationships between nodes can be viewed as reciprocal or non-directional. In our models, we are focusing on undirected networks.
The first step in analyzing the structure of a network is to get in touch with the concepts and tools that describe it. A crucial and valuable network measure is **centrality**, which assesses the importance of vertices or edges within a networked system. Numerous mathematical centrality measures have been developed, each emphasizing different definitions and concepts of centrality in a network. A fundamental yet highly practical example is **degree**, which calculates the number of edges connected to a vertex in a network. We will denote the degree of vertex i by k_i. The **density** ρ of a graph is the fraction between the potential edges present and the maximum number of edges in a graph, it always falls within 0 and 1. Let's assume we have a network with m edges and n nodes, the maximum possible number of edges is $\binom{n}{2}$, so ρ must be:

$$\rho = \frac{m}{\binom{m}{2}} = \frac{2m}{n(n-1)} \tag{2.1}$$

A network is considered **dense** when its density approaches a constant value as n$\to \infty$. In contrast, a network is classified as **sparse** when its density ρ approaches zero as n$\to \infty$.
To measure a network's connectivity, various benchmarks can be employed. The **clustering coefficient** is one of such, it measures the number of closed paths of length three in a network. Along with this, **hubs**, which are vertices that exhibit higher degrees, can be used to determine the significance of specific vertices within a network.
Another crucial concept related to network connectivity is the **giant component**, which refers to a network component that increases proportionally with n. For a vertex i to be excluded from the giant component, it must not be linked to the giant component through any other vertex.
Paths are another critical aspect of network connectivity, indeed a path is a series of vertices connected by edges within a network. The length of a path in a network is the number of edges crossed along the path, and the shortest path between two nodes is called a **geodesic** path.
Finally, the **adjacency matrix** provides a mathematical representation of a network. The adjacency matrix for a simple graph is the matrix with elements A_{ij}, which indicates whether an edge exists between vertices i and j, so that:

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between i and j} \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

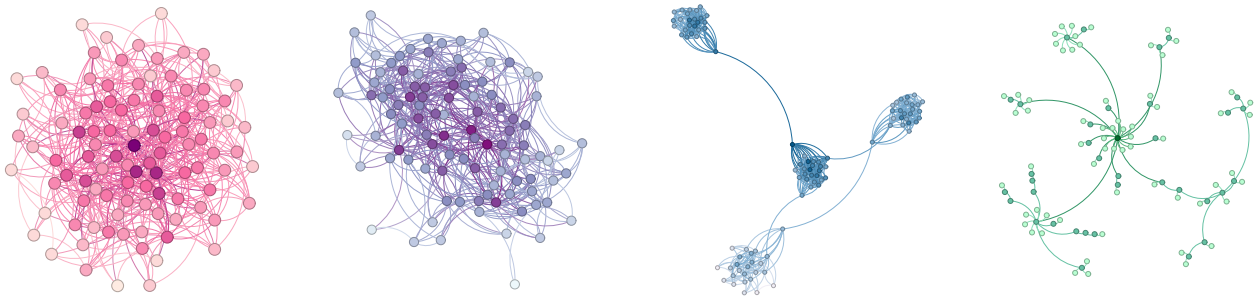
This representation is fundamental for analyzing the structure and properties of a network mathematically.
We now introduce the most basic of network models: **random graphs**. Random graph models involve generating networks that exhibit specific properties of interest, such as predetermined degree distributions, while showing randomness in other aspects.

One of the most famous examples of a random graph is the model referred to as the Erdős-Rényi model Fig. 2.1 a, it fixes the probability of links between vertices instead of the number of edges so that there are still n vertices, but each unique pair is connected by an edge with an independent probability p. This network has the characteristic of being entirely random and can be sparse or dense depending on the probability p. In this type of network, the mean degree is $\langle k \rangle = (n-1)p$.

Another random network is the random geometric graph Fig. 2.1 b in which n nodes are randomly distributed in a metric space, based on a specified probability distribution, and two nodes are linked only if they are within a certain range of distance, such as being smaller than a designated neighborhood radius r. It depict the concept of links between individuals that are geographically close to each other. This type of spatial network is considered the simplest one mathematically, after a lattice.

The Stochastic block model (SBM) Fig. 2.1 c generates networks that exhibit communities, which are groups of vertices linked to each other with specific densities of edges. Assume n is the number of total vertices, k the number of communities, p=(p_1, . . . , p_l) is a probability vector (the prior on the l communities), and w a ll symmetric matrix with entries in [0, 1] representing the connectivity probabilities. The model generates l groups each with a number np_i of nodes and every vertex is connected to the others according to the w matrix (with w_{ii} being the probability of connecting within the same group) [Abbe (2017)].

Another model is based on a "rich-get-richer" mechanism, under the assumption that wealthy individuals make money by investing the money they have. Similarly, in a friendship network, individuals with a higher number of friends are more likely to make more friendships (so more links) with new individuals. This model has been named after Barabási and Albert Fig. 2.1 d. It is a generative undirected network model, in which nodes are added incrementally one by one to a growing network, with each vertex connecting to a set of previously established vertices that have been suitably selected. Every node, as it appears, establishes a fixed number of connections, which is denoted as c (so that no vertex has a degree k<c). Since we are referring to a preferential attachment mechanism, the more connected a node is the more likely it is to receive new connections [Newman (2010)].



(a) Erdos-Renyi (b) Geometric Random (c) Stochastic Block Model (d) Barabasi-Albert

Figure 2.1: Representation of networks, the gradient of colours indicates the degree of nodes, the higher the degree the darker the colour.

3

Epidemic spreading

The connection between social networks and disease transmission is one of the primary reasons why the scientific community has made significant investments in its study. Social connections between people can facilitate the spread of diseases, such as airborne illnesses like influenza or tuberculosis that can be transmitted when individuals share the same air in a room. Physical contact can lead to the transmission of contagious diseases and parasites, while sexually transmitted infections, such as HIV, can spread through sexual contact.
Understanding what happens biologically within an individual, also known as a "host" in epidemiology, when they contract an infection, can be a complex task. The pathogen responsible for the disease tends to replicate in the body, while the immune system works to fight it off, often causing symptoms. Ultimately, either the individual recovers, dies, or remains in a chronic state of permanent infection. Although, in theory, understanding the complete biological picture is necessary to understand how diseases spread within populations, in practice, this is often a daunting task and scarcely attempted. Luckily, there are simpler models for disease transmission that provide valuable insights into disease behavior in many situations. The typical mathematical models used to depict an epidemic simplify the complex dynamics of the disease within an infected individual by reducing it to a few fundamental disease states.

3.1 SI model

In its simplest form, this model only includes two states: Susceptible and Infected. A susceptible person is someone who has not yet contracted the disease but is at risk of infection through contact with an infected person. An infected individual, on the other hand, is someone who already has the disease and can potentially transmit it to a susceptible individual through contact.
The conventional method assumes that every individual has an equal chance of coming into contact with any other person per unit of time. This assumption implies that people meet randomly, which is far from how the world operates. In reality, people have contacts with only a small percentage of the population, and that fraction is not chosen randomly. This is precisely why networks play a critical role in the spread of disease. Despite this, a basic understanding of the traditional approaches is beneficial in our exploration of network epidemiology. Therefore, we will examine its fundamental principles briefly.

Assume S(t) is the number of individuals susceptible to the infection at time t and I(t) the number of infected. If the chance to be infected is uniform across the population, we can assume S and I are the average number of susceptible and infected individuals if we ran the process several times. Imagine that disease transmission occurs solely through chance encounters, with each individual having an average rate of β (a parameter that we will properly define later) contacts with random individuals per unit of time, resulting in the spread of the disease. Subsequently we present a type of flow chart that will be used to illustrate the possible state transitions:

$$\mathrm{S} \xrightarrow{\beta} \mathrm{I}$$

Transmission of the disease occurs solely when an infected individual interacts with someone susceptible. Assuming that there are N individuals in the population, the likelihood of encountering a susceptible individual at random is $\frac{S}{N}$. Hence, an infected person comes into contact with an average

of $\frac{\beta S}{N}$ susceptible individuals per unit of time. Given that there is an average of I infected individuals at any given time and that the number of susceptible individuals goes down at the same rate as I grows, we can formulate differential equations for the rate of change of I and S as follows:

$$\dot{S} = -\frac{\beta}{N} SI \qquad \dot{I} = \frac{\beta}{N} SI \ , \tag{3.1}$$

with the conservation of population condition, since no individual can appear or disappear over time:

$$S + I = N \qquad \forall t \tag{3.2}$$

Adding this condition to the previous equation, we can write:

$$\dot{I} = \beta \frac{I}{N}(1 - \frac{I}{N}), \tag{3.3}$$

that gives essentially the logistic growth equation:

$$I = \frac{I_0 e^{\beta t}}{N - I_0 + I_0 e^{\beta t}}, \tag{3.4}$$

where I_0 is the value of I at t = 0.
As illustrated in Fig. 3.1, in the initial stage of the disease, when a significant portion of the population is susceptible, the amount of infected individuals shows an exponential growth that gradually reaches a point of saturation as the number of susceptible ones decreases.

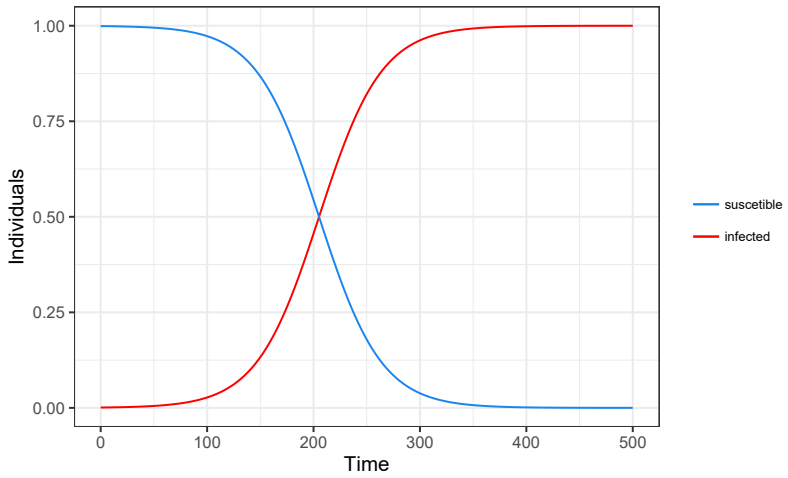


Figure 3.1: Logistic curve of the SI epidemic model (β: 0.034)

However, the definition of β given is not verisimilar. It is unreasonable to assume that any two people can have a chance encounter in the real world. The likelihood of two randomly selected individuals from the global population meeting each other is insignificant and can be considered negligible. Network models are useful since we can see links as potential contacts instead of assuming that contact is possible with the entire population. In a new model, considering individuals linked by edges in a network, we can redefine the **transmission rate** β as the probability per unit time that an infection spreads between an infected and susceptible individual linked by an edge in the network. As before, β is an inherent characteristic of a disease, with some diseases being more easily spread than others, resulting in higher transmission rates However now, it is also influenced by the social and behavioral characteristics of a real population.

First of all, let's take into account a fixed network, not changing over time, and focus on its giant component since any other outbreak will merely impact a small fraction of the population and eventually extinguish. If we consider a vertex i that does not belong to the giant component, then according

to the hypothesis assuming that the epidemic only occurs within the giant component, its S_i, the probability of a node being susceptible, value is zero. On the other hand, if i is part of the giant component, we can formulate a differential equation for S_i by calculating the likelihood of i being infected between t and t + dt. For i to become infected, it has to contract the disease from an already-infected neighboring individual j, which occurs with a probability of $I_j = 1$ - S_j, and then j has to transmit the disease during the given time interval, which occurs with a probability of βdt. By multiplying these probabilities and then summing them over all neighbors of i, we obtain the total likelihood of i being infected, which is $\beta S_i \sum_j A_{ij} I_j$, where A_{ij} represents an entry in the adjacency matrix. So we obtain the non-linear differential equations:

$$\dot{S}_i = -\beta S_i \sum_j A_{ij}(1 - S_j) \qquad \dot{I}_i = \beta(1 - I_i) \sum_j A_{ij} I_j \tag{3.5}$$

Correlations between the disease states of different vertices can be handled with an approximate method that gives reasonably accurate results in practice, although not ideal, and generates equations that can be solved through analytical means. Additionally, this approach can be extended to other epidemic models such as the SIR model, as we will show. The approximation introduced was to assume that all vertices of the same degree have the same probability of infection at any given time.
Let's examine a susceptible vertex A. For A to become infected, it must contract the infection from one of its network neighbors. The likelihood of a specific neighbor B being infected is determined by the neighbor's degree. By hypothesis, vertex A is not infected; hence B cannot have contracted the disease from A. If B is infected, it must have caught the disease from one of its remaining neighbors. As a result, this decreases the degree of B by one. Consequently, B will have the same probability of being infected as the average vertex with a degree one less. In other words, B's probability of infection is determined by its excess degree, which refers to the number of edges it has apart from the edge we followed from A to reach it. Hence, B's probability of infection is represented by x_k, where k denotes the excess degree (we indicate q_k as the probability distribution of the excess degree), not the total degree.

That means that the average probability that the neighbor is infected is:

$$v(t) = \sum_{k=0}^{+\infty} q_k \frac{I_k(t)}{N} \tag{3.6}$$

Supposing the neighbor is infected, the probability of the disease being transmitted to vertex A during the given time interval is βdt. Thus, the total likelihood of transmission from a single neighbor during the time interval is $\beta v(t)$dt, and the likelihood of transmission from any neighbor is k$\beta v(t)$dt, where k represents the number of A's neighbors. Consequently, the rate at which S_k changes is expressed as:

$$\dot{S}_k = -\beta k v S_k \tag{3.7}$$

This equation can be solved exactly. Since $S_k(t)$ measures the probability that a vertex of a given degree is susceptible, thus, as we might expect, the vertices with higher degrees are the ones that become infected first.

3.2 SIR MODEL

The SI model represents the most basic form of an infection model, which can be expanded to better fit specific diseases or to enhance its realism. One common extension involves incorporating recovery from the disease.
In the SI model, individuals remain infected and infectious indefinitely once they contract the disease. However, for many actual diseases, people recover over time as their immune system combats the disease agent. Additionally, people typically retain their immunity following recovery, making them unable to contract the disease again. To reflect this pattern, a new third state is necessary, often labeled as R(t) for recovered. The resulting model is known as the SIR model.
To describe the model we need to define an additional parameter γ (called **recovery rate**) as the rate

at which individuals recover while contacts between individuals are assumed to happen, as before, at an average rate β per person. So that:

$$\mathrm{S} \xrightarrow{\beta} \mathrm{I} \xrightarrow{\gamma} \mathrm{R}$$

The expected duration of individual's infection before recovering is the inverse of the recovery rate γ. We can present the differential equations for the SIR model:

$$\dot{S} = -\frac{\beta}{N} SI \qquad \dot{I} = \frac{\beta}{N} SI - \gamma I \qquad \dot{R} = \gamma I \tag{3.8}$$

and, of course, the conservation constraint applies:

$$S + I + R = N \tag{3.9}$$

Unfortunately, it is not possible to solve the integral in a closed form in practical applications. Nonetheless, we can compute it numerically (see Fig. 3.2)

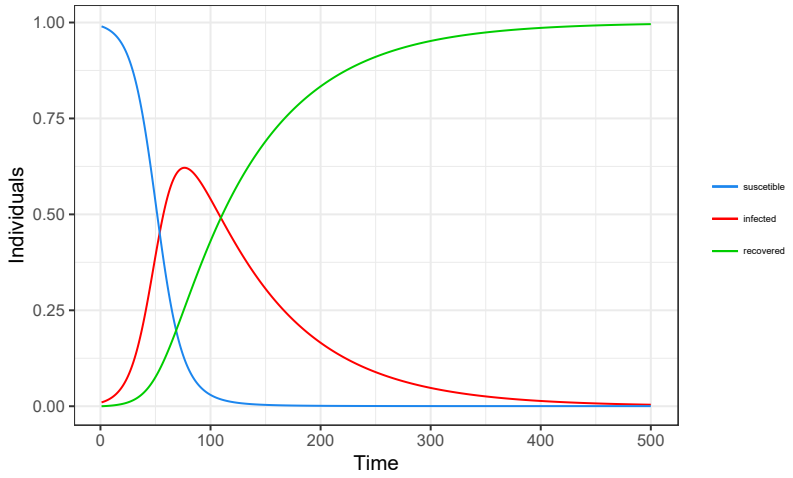


Figure 3.2: Time evolution of the SIR model (β: 0.102, γ: 0.0125)

As the infection spreads, the proportion of susceptible individuals in the population decreases continuously, while the amount of recovered individuals increases continuously. In contrast, the fraction of infected individuals initially increases as more people get infected, then decreases as they recover, and ultimately approaches zero as time moves towards infinity.
It is interesting to note that if $\beta \leq \gamma$, susceptible individuals are less likely to become infected compared to the rate at which infected individuals recover. As a consequence, despite the small number of infected individuals at the beginning, their count decreases instead of increasing, and the disease eventually fades out rather than spreading. The point at which the transition occurs between the non-epidemic and epidemic states is known as the **epidemic transition**, which takes place when $\beta = \gamma$. A quantity that controls this threshold in the study of epidemics is $R_0 = \frac{\beta}{\gamma}$, the **basic reproduction number**. We can better define it as the average number of individuals in a completely susceptible population to whom an infected person transmits the disease before recovering. The epidemic threshold occurs at $R_0 = 1$, which distinguishes the growth and decline patterns of the disease. This point serves as the boundary between the two regimes in which the disease either proliferates or extinguishes.

Concerning SIR models on networks, the equation that governs the evolution of S_i is (roughly) the same as before:

$$\dot{S}_i = -\beta S_i \sum_j A_{ij} I_j \qquad \dot{I}_i = \beta S_i \sum_j A_{ij} I_j - \gamma I_i \qquad \dot{R}_i = \gamma I_i \tag{3.10}$$

Similarly to the SI model, we cannot obtain a precise solution to these equations, but we can extract some valuable outcomes. With these equations, the new epidemic threshold for our model is at $\beta k_1 - \gamma = 0$. Where k_1 is the largest positive eigenvalue of the adjacency matrix.

3.3 SIRS MODEL

We will now briefly define an epidemic model incorporating reinfection, the SIRS model. In this model, individuals recover from the infection and acquire immunity, as in the SIR model. However, the immunity gained is not permanent, and individuals lose it after a specific period of time, making them susceptible again. To represent the average rate at which individuals lose immunity, we introduce a new parameter, δ.

$$\mathrm{S} \xrightarrow{\beta} \mathrm{I} \xrightarrow{\gamma} \mathrm{R}, \quad \mathrm{I} \xrightarrow{\delta} \mathrm{S}$$

So that the differential equations are:

$$\dot{S} = \delta R - \frac{\beta}{N} SI \qquad \dot{I} = \frac{\beta}{N} SI - \gamma I \qquad \dot{R} = \gamma I - \delta R \tag{3.11}$$

and, as always,

$$S + I + R = N \tag{3.12}$$

The SIRS model cannot be solved analytically but can be analyzed using techniques such as linear stability analysis and other non-linear methods.

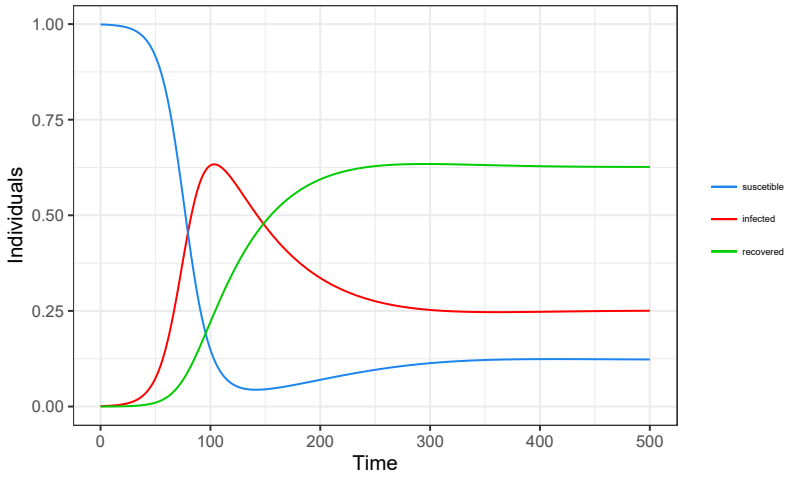


Figure 3.3: Time evolution of the SIRS model (β: 0.102, γ: 0.0125, δ: 0.005)

4

Pathogen evolution

The dynamics of epidemics are influenced by two distinct processes that operate at different scales. At the individual level, the pathogen that has infected a person replicates and may potentially mutate within their body. Subsequently, when the pathogen reaches a critical load, it spreads between hosts, allowing it to circulate in the social network. In this chapter, we are about to discuss the within-host dynamics captured by a random process of mutation and selection.
When the pathogen infects a particular host A, it replicates inside the body. After some replication cycles, one or more offspring of the pathogen infects one of A's social network neighbors, B. During the replication process within the host, the pathogen may undergo mutations, increasing the likelihood that B will be infected with a distinct strain than the one initially contracted by A.
It is important to introduce the parameter f_i that represents the fitness of the pathogen i, which measures its replication rate inside host i. Various factors affect this parameter, such as the host's immune response and body temperature. Strains with higher f_i can multiply more efficiently within the host's body, enabling them to achieve a greater prevalence with each replication cycle [Zhang et al. (2022)].

It is important to highlight that evolutionary dynamics operate on populations, meaning that it's not genes, cells, or individuals that evolve, but rather the populations they belong to. In this regard, the **quasispecies equation** is used to describe the behavior of population genetics in complex adaptive systems:

$$\dot{x_i} = \sum_j x_j f_j Q_{ij} - \left(\sum_j x_j f_j \right) x_i \tag{4.1}$$

where $\dot{x_i}$ is the change in proportion of strain i, the parenthesis is the average fitness of the entire population and Q_{ij} is the evolutionary leap matrix (assuming it is an upper triangular matrix, in order to avoid backward evolutions), which is a row stochastic matrix, so that

$$\sum_{j=1}^{n} Q_{ij} = 1 \qquad 1 \leq i \leq n. \tag{4.2}$$

In other words, the matrix Q_{ij} represents the probability of mutating from a strain i to a strain j. The Q_{ij} coefficients are proportional to the size of the evolutionary leap (the further the jump, the lower its probability to happen). Note that definition 4.2 leads to the conservation law for the total population.
The replicator equation is the first game dynamics studied from an evolutionary and biological perspective. The aim is to predict the evolutionary outcome of population behavior without involving genetic.
Fitness is used as payoff or reproductive success, and the equation describes the per capita rate of change of a strategy based on the difference between its expected payoff and the average payoff of the population. If the payoff for each strategy remains constant, regardless of its frequency, the end result of evolution will be that all individuals adopt the strategy with the highest payoff. In other words, it can be described as Darwin's principle of natural selection favoring the survival of the fittest [Cressman and Tao (2014)].
It is possible to generalize the replicator equation since frequency-dependent selection is a crucial factor in evolutionary game theory: an individual's fitness can depend on the prevalence of other

strategies in the population [Page and Nowak (2002)]. In this context, the fitness of an individual strain is a function of the overall distribution of strategies in the population $x = x_1, ..., x_n$, so that $f_i = f_i(x)$ (adding the dependency of the fitness on the frequency of the species we obtain the so-called **replicator-mutator equation**). In this perspective, competition models emerge if we suppose that resources diminish proportionally to the population densities of two rival species and that the growth rates rise proportionally with the resources [Bomze (1983)].

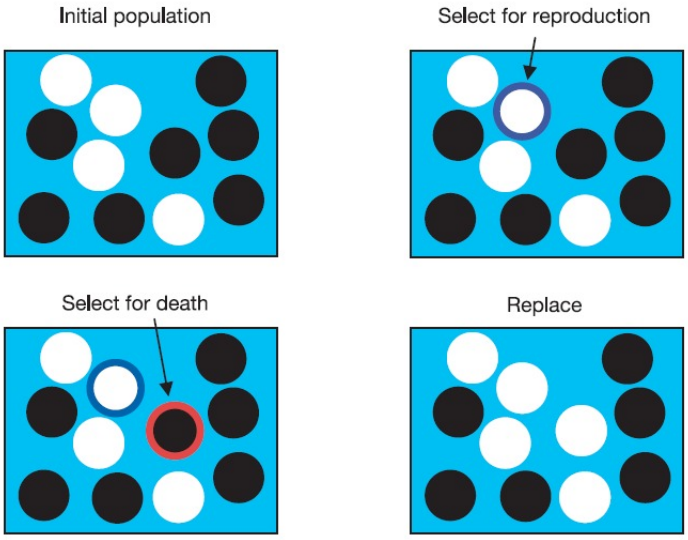


Figure 4.1: A constant-sized homogeneous population undergoes a stochastic evolution where, in each time step, an individual is selected for reproduction based on their fitness level. At the same time, another individual is chosen for death. The second individual is then replaced by the offspring of the first individual. This is known as the *Moran process*. [Lieberman et al. (2005)]

One could argue that the survival of the fittest is not the only option, and predator-pray models are not the only ones that can describe the evolutionary outcome of a population. The goal of ecological research on species coexistence is to explain how various species with different competitive abilities manage to coexist in nature, despite the scarcity of resources. Researches [Levine et al. (2017)] have indicated that when the strengths of competitive interactions are randomly sampled from a distribution, communities with higher levels of diversity are less likely to be stable. This suggests that increasing the number of species reduces opportunities for coexistence, but intransitive competitive structures (in which three or more species compete with each other in a way that does not follow a clear hierarchy of superiority) can alter this expectation.
It has been investigated the specific competitive relationships that can alter the outcome of competition between two species when additional competitors are introduced. By including a third competitor, coexistence can be facilitated by having an equalizing or stabilizing effect. For instance, a superior competitor can encourage the coexistence of two others by differentially harming the fitter of the two.

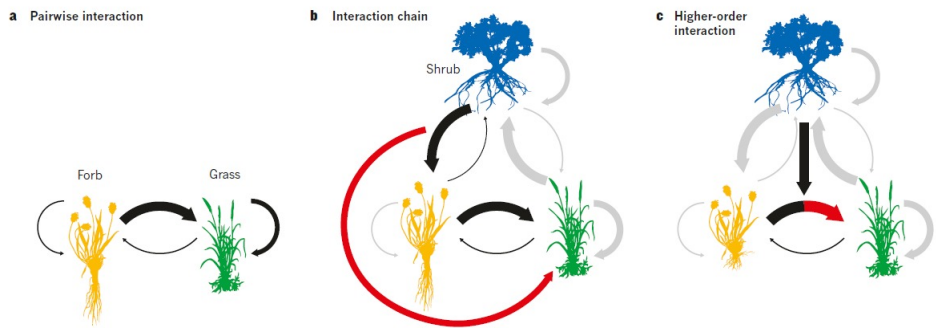


Figure 4.2: Coexistence mechanisms that emerge with more than two competitors. [Levine et al. (2017)]

In order to accurately analyze the probability of mutations in a DNA sequence, we need to consider a parameter known as the **mutation rate**, denoted by η. Each base in the DNA sequence has a certain probability of mutation, which is represented by η_0. Since the mutation rate is closely related to the size of the genome, as shown in Fig. 5.5, we define η as a function of the total number of bases, denoted by L. This initial approximation for η can be expressed as:

$$\eta = 1 - (1 - \eta_0)^L \tag{4.3}$$

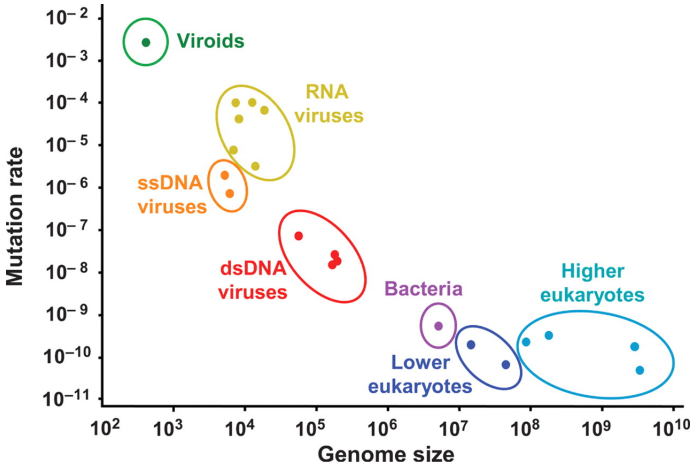


Figure 4.3: Per-site mutation rate versus genome size for some biological entities [Gago et al. (2009)]

Traditionally, the dynamics of evolution have been investigated within the framework of homogeneous or spatially extended populations. However, in this study, we expand upon population structure by organizing it into a graph, where each vertex corresponds to an individual.
Recent theoretical research has indicated that the framework of these networks can significantly impact evolutionary dynamics outcomes. Social network structures can affect how genes change over time, and if the network is dense, genes may be lost more quickly. In other words, networks less dense preserve different alleles more effectively since they connect distinct sections of the population distantly. In contrast, networks more densely connected link the population more closely, leading to a higher likelihood of losing genetic diversity [Kurvers et al. (2014)].

5

Coupled dynamics models

In this chapter, we will present and discuss the results obtained from Monte Carlo simulations of an epidemic spreading on top of networks while the pathogen evolves and mutates, and the mathematical model built in order to describe the phenomenon. We assume there are 20 possible mutations of the pathogen each with a potentially different transmission probability β and fitness. Despite the fact that the fitness is a fundamental characteristic of the different strains, in order to simplify the model, the fitness is not considered neither in the Monte Carlo simulations nor in the mathematical model.

5.1 Monte Carlo simulations

A disease spreads into a network starting from a patient zero, infected with the original species of the virus. Each susceptible individual, linked through the network with the patient zero, has a probability β of being infected itself.

Specifically, our code simulates the spread and mutation of a virus by defining rates. Firstly, it creates a network and randomly chooses an individual to be the patient zero with the original strain.

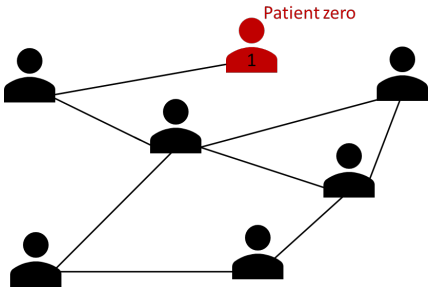


Figure 5.1: Network at t=0

Then, at each time step, the code identifies all susceptible individuals and their neighboring infected individuals with a certain strain. Using a fixed probability β (that could be different for each species but, for the purpose of this thesis, we assumed to be the same for all species), the susceptible individuals become infected with the same strain that infected them.

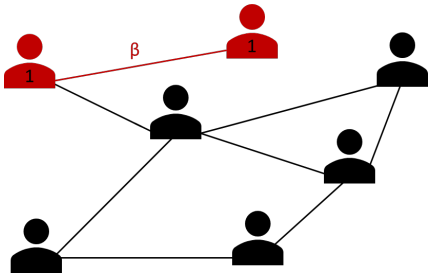


Figure 5.2: After some time steps the virus starts to spread

The infected individuals then undergo possible mutations (with a probability η) that change the initial strain. Consequently, at successive time steps, instead of passing the original species, it infects its

susceptible neighbors with the mutated strain.

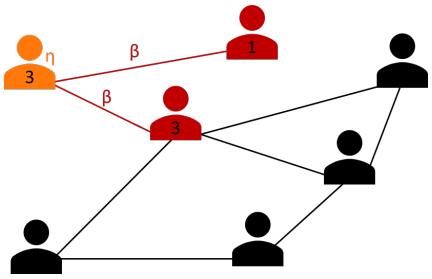


Figure 5.3: As time passes the virus can mutate within an individual

We built four different networks of 100 nodes that represents 100 individuals in a social network where edges are stable contacts between individuals.

5.1.1 SI Model

We now present the average results obtained for 20 Monte Carlo simulations, a total of N=100 individuals and 20 possible mutations of the virus (Fig. 5.7)

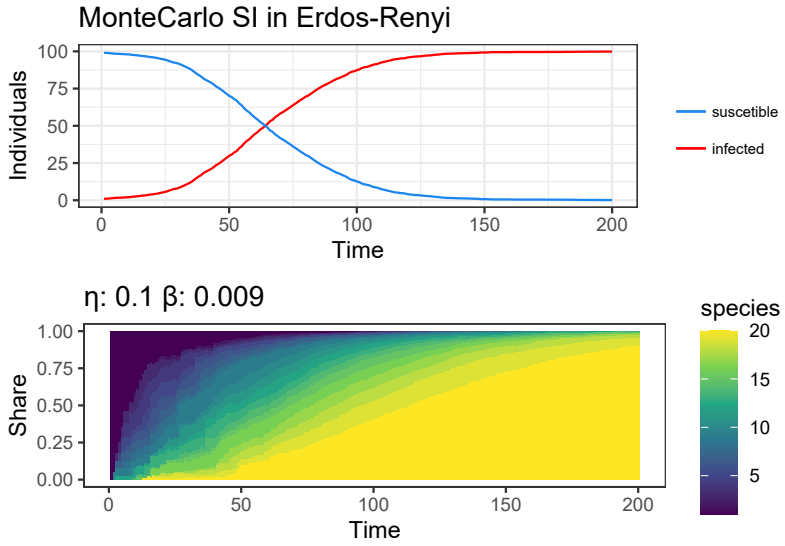


Figure 5.4: The top picture shows the spreading dynamics, while the bottom one shows the average frequency of the strains with respect to the evolutionary dynamics.

As we can notice, since mutations occur within infected individuals, and there is no recovery, over time the last strain will outnumber the others.
Our focus will be mainly on the SIR model but from the above plots we can grasp the behavior of the strains in absence of recovery.

5.1.2 SIR Model

Our code is implemented to include the recovery process. At each time step, there is a chance γ for an infected individual to recover. For further details on the code, see Appendix in 7.

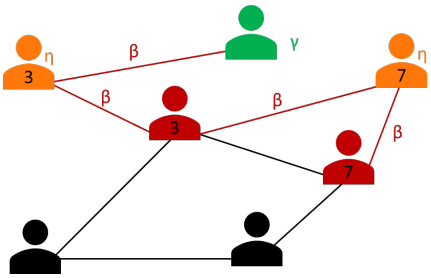


Figure 5.5: Infected individuals can recover and never being subjected to the virus anymore

In Fig. 5.6 we show the epidemic spreading dynamics obtained by averaging over 20 Monte Carlo simulations:

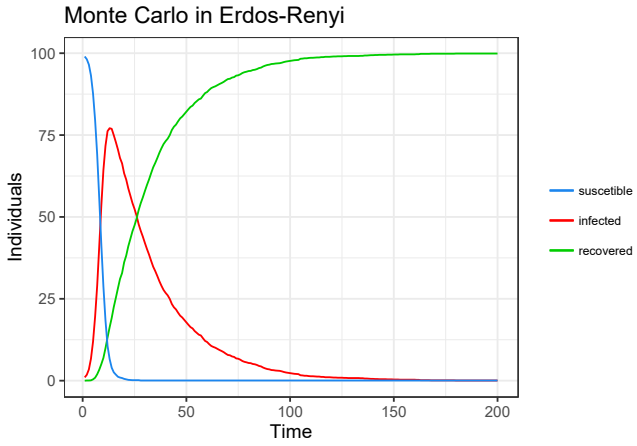


Figure 5.6: Average results of 20 Monte Carlo simulations for the epidemic dynamics with $\eta : 0.1$, $\beta : 0.09$, $\gamma : 0.04$

Since there is a recovery, at a certain point there will be no infected individuals left, resulting in a heat map partially empty:

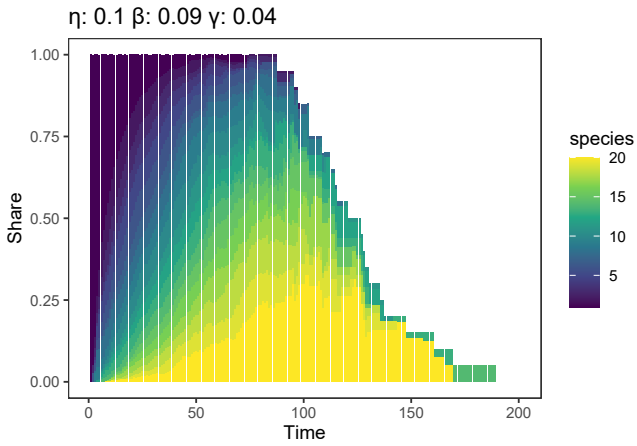


Figure 5.7: Half empty heat map for 20 Monte Carlo simulations in Erdos-Renyi. The value of the parameters used for the simulations is reported on the top.

To prevent this, the code has been improved. Firstly, a simulation starts to gather the number of individuals infected for each strain at a certain time step. This vector is then normalized with the

total number of infected individuals at that time. If there are no infected individuals left during a simulation, an abundance value of zero is returned for each strain. After all simulations have been conducted, the code calculates the average frequency of each strain for each time. However, it only considers simulations that still have infected individuals at that time step to ensure a fully filled heat map and enable comparisons. Of course this method results increasingly precise as we raise the number of simulations.

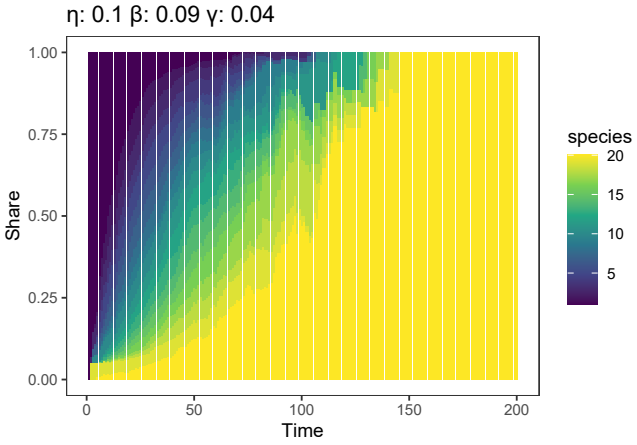
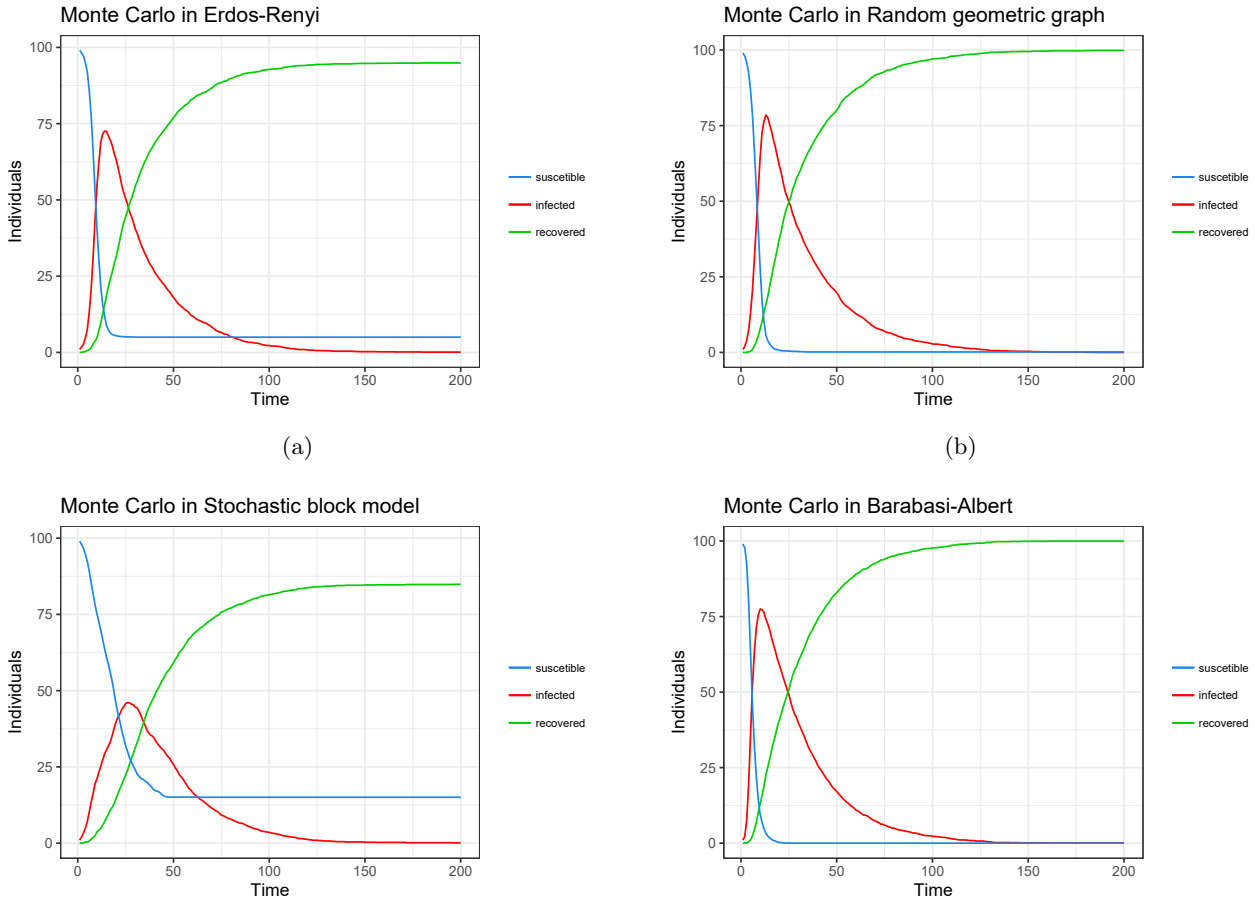


Figure 5.8: Filtered heat map for 20 Monte Carlo simulations in Erdos-Renyi.

We present here the results obtained from the simulations conducted using each of the networks mentioned in Chapter 2.



(a) (b) (c) (d)

Figure 5.9: Average results for 20 Monte Carlo simulations of the spreading dynamics $\eta : 0.1, \beta : 0.09, \gamma : 0.04$

The networks we analyzed have an average degree of $\langle k \rangle = 10$. Showing all the networks helps to better understand the role of their structure in the simulated dynamics. It is evident from the plots that distinct networks respond differently to the spread of infection. The major difference is seen in the stochastic block model where the epidemics has more difficulties to take off since the network is more sparse.
However, there are minor discrepancies in their evolutionary behavior so we decided to omit the majority of them. As evidence, we present the heat map of the SBM model in Fig. 5.10 :

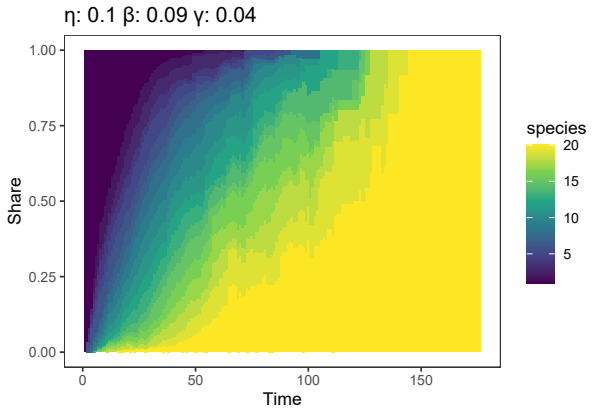
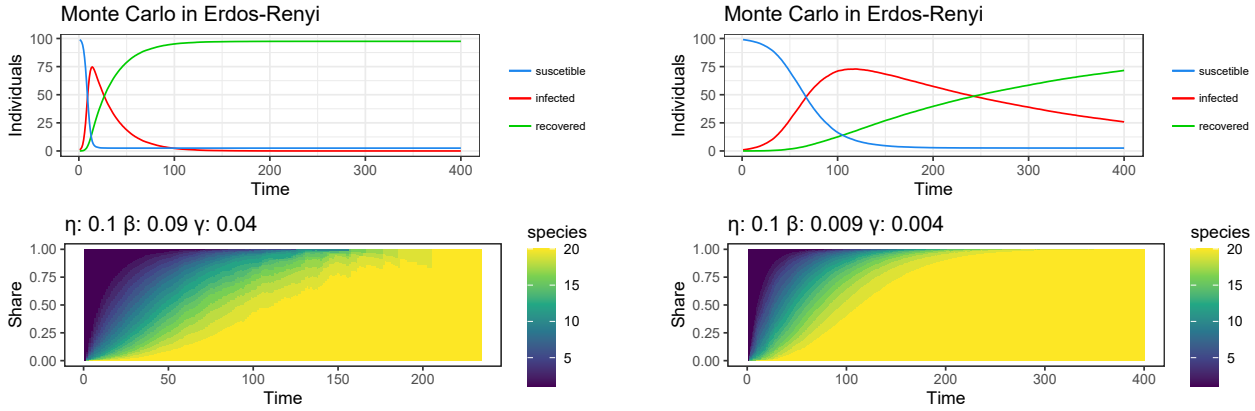


Figure 5.10: Heat map for 20 Monte Carlo simulations in Stochastic block model. The trend is the same of the heat map in Erdos-Renyi in Fig. 5.8

Regardless of the network used, a triumph of the last strain occurs as we approach stationarity.
From now on, we will only consider the Erdos-Renyi network because it is the most similar to a simple network, where each node has an average degree of $\langle k \rangle$. This decision will be useful in the next section of this chapter, where we will discuss the mathematical model.

Let's now discuss how the choice of parameters affects the spreading dynamics in networks. To illustrate this, let us consider the two epidemic spreading graphs in Fig. 5.11 a - b.



(a) Average results for 200 Monte Carlo simulations with 1° set of parameters

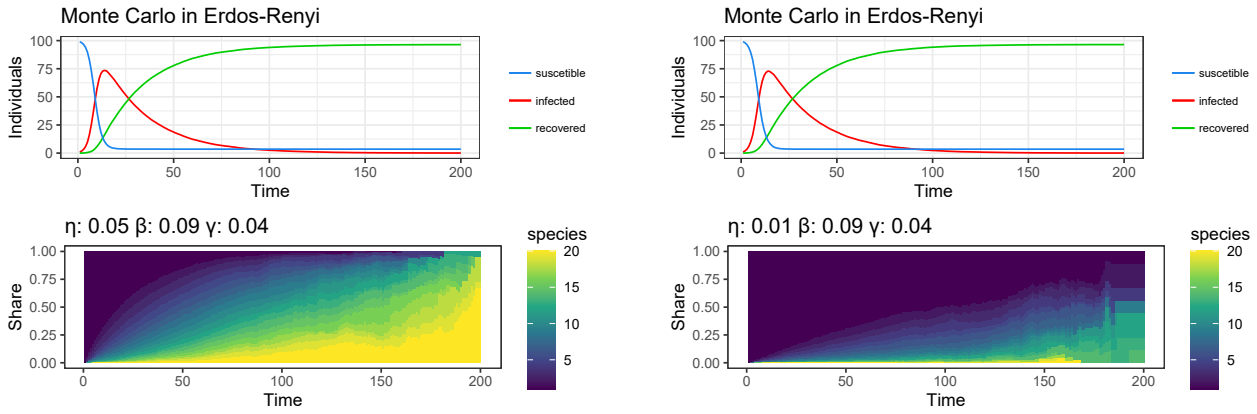
(b) Average results for 200 Monte Carlo simulations with 2° set of parameters

Figure 5.11: Simulation results in Erdos-Renyi network using different β and γ parameters

Both graphs have the same $R_0 = \langle k \rangle \frac{\beta}{\gamma} = 22.5$, but the epidemic spread is different from a timescale point of view. In the first graph, the epidemic peaks early and then declines rapidly, while in the second graph, the epidemic peaks later and declines more slowly. This suggests that there is another parameter that influences the dynamics of the disease. In this case, we can see that, even though they have the same R_0 value, they have different values of β. The first epidemic process has a higher value of β than the second graph, which means that the disease spreads more easily and rapidly in the

first case. Although R_0 and β are intimately connected, they capture different aspects of the disease dynamics.
However, it's important to note that this observation applies only in network-structured SIR models and does not apply to the fully connected case, where it has been shown [Harko et al. (2014, sec. 2.2)] that the only parameter that affects the dynamics is R_0.

Finally, we examine various values of the η coefficient. One noticeable difference in Fig. 5.12 a below, compared to the previous ones, is that different species coexist for a longer period of time. However, over time, the last species eventually dominates, even though more slowly.



(a) Average results for 200 Monte Carlo simulations with 1° set of parameters

(b) Average results for 200 Monte Carlo simulations with 2° set of parameters

Figure 5.12: Confront between simulations with different parameters

On the other hand, in Fig. 5.12 b, we witness the extinction of several species before they have the chance to dominate the others.

5.2 Modeling via differential equations

We define rates for the SIR model with V strains that evolve hierarchically, where each strain V is fitter than the previous strain V-1. The model's dynamics occur on a network, and individuals are grouped according to their degree. In order to gain a deeper insight into the infection dynamics, we have provided flow charts and rates. Note that I_{ik} represents the number of infected individuals with strain i and degree k, while S_{ik} and R_{ik} represent the number of susceptible and recovered individuals, respectively:

$$I_{ik} \xrightarrow{\eta_{ij}} I_{jk}$$

$$I_{ik} + S_{i\tilde{k}} \xrightarrow{\beta_i} I_{ik} + I_{i\tilde{k}}$$

$$I_{ik} \xrightarrow{\gamma_i} R_{ik}$$

Contrary to the definition provided earlier, η_{ij} is not solely the mutation rate. In this notation, it represents the product of two factors assumed to be independent of each other, which is a good approximation for our purposes. The first factor is the mutation rate, denoted by η, while the other is the probability of transitioning to another strain, denoted by Q_{ij}. Furthermore, we need to redefine β_i so that it can be normalized with the β_j values of the other strains:

$$\beta_i = \beta_{max}(1 - e^{-\alpha(i-1)}) + \beta_0 \tag{5.1}$$

while, since $\beta_i \leq 1$, $\beta_{max} + \beta_0 = 1$

$$\Rightarrow \quad \beta_i = 1 - e^{-\alpha(i-1)}(1 - \beta_0) \tag{5.2}$$

Where β_0 is the transmission rate of the original pathogen species, β_{max} a normalization factor that scales the transmission rate, and α is a parameter that controls the exponential decay multiplied by the evolutionary distance between the i-th strain and the original one.

Finally, we present the deterministic heterogeneous mean-field equations:

$$\begin{cases} \dot{S_{ik}} = -S_{ik} \sum\limits_{\tilde{k}} \lambda_{ik\tilde{k}} \\ \dot{I_{ik}} = S_{ik} \sum\limits_{\tilde{k}} \lambda_{ik\tilde{k}} + \eta \left[\sum\limits_{j=1}^{i-1} Q_{ji} I_{jk} - \sum\limits_{j=i+1}^{V} Q_{ij} I_{ik} \right] - \gamma_{ik} I_{ik} \qquad 2 \leq i \leq V-1 \\ \dot{R_{ik}} = \gamma_{ik} I_{ik} \end{cases} \tag{5.3}$$

and

$$\begin{aligned} \dot{I_{1k}} &= S_{1k} \sum_{\tilde{k}} \lambda_{1k\tilde{k}} - \eta \sum_{j=2}^{V} Q_{1j} I_{1k} - \gamma_{1k} I_{1k} \\ \dot{I_{Vk}} &= S_{Vk} \sum_{\tilde{k}} \lambda_{Vk\tilde{k}} + \eta \sum_{j=1}^{V-1} Q_{jV} I_{Vk} - \gamma_{Vk} I_{Vk} \end{aligned} \tag{5.4}$$

In homogeneous mean-field equations, which apply to a single population without a network structure and for a small value of β, the probability λ for susceptible individuals to be infected with strain i is reduced to the traditional value $\beta_i \frac{I_i}{N}$. However, under heterogeneous mean-field, λ becomes a function of the degrees k and $\tilde{k}$ of two compartments, so that:

$$\lambda_{ik\tilde{k}} = 1 - (1 - \beta_i)^{\frac{I_i}{N} \tilde{k} P(k|\tilde{k})} \tag{5.5}$$

The exponent contains the term $P(k|\tilde{k})$, which depends on the chosen network structure as it quantifies degree correlations. In networks without topological correlations, such as Erdos-Renyi type networks,

where all nodes have the same average degree, $P(k|\tilde{k})$ is simply a Dirac delta function on $\langle k \rangle$. In that limit, we find $\langle k \rangle \frac{I_i}{N}$ in the exponent as in the homogeneous case.

In order to verify the validity of our model, we simplified the equations by assuming the same degree for each node. Indeed, to represent this assumption, we used the Erdos-Renyi network, which we believe is the best model for this purpose. Therefore, we will compare our model to this type of network. Then, since we are assuming the totality of the population is susceptible to each strain, we slightly modify the equations in order to avoid considering every S_{ik} but consider only one S_k as a whole. So the equations in 5.6 become:

$$\begin{cases} \dot{S_k} = -S_k \sum_i \lambda_{ik} \\ \dot{I_{ik}} = S_k \lambda_{ik} + \eta \left[\sum_{j=1}^{i-1} Q_{ji} I_{jk} - \sum_{j=i+1}^{V} Q_{ij} I_{ik} \right] - \gamma_{ik} I_{ik} \qquad 2 \leq i \leq V-1 \\ \dot{R_{ik}} = \gamma_{ik} I_{ik} \end{cases} \tag{5.6}$$

and

$$\begin{aligned} \dot{I_{1k}} &= S_k \lambda_{1k} - \eta \sum_{j=2}^{V} Q_{1j} I_{1k} - \gamma_{1k} I_{1k} \\ \dot{I_{Vk}} &= S_k \lambda_{Vk} + \eta \sum_{j=1}^{V-1} Q_{jV} I_{Vk} - \gamma_{Vk} I_{Vk} \end{aligned} \tag{5.7}$$

In 5.13 we present the results for the ODE deterministic model simulation, under the before mentioned approximations.

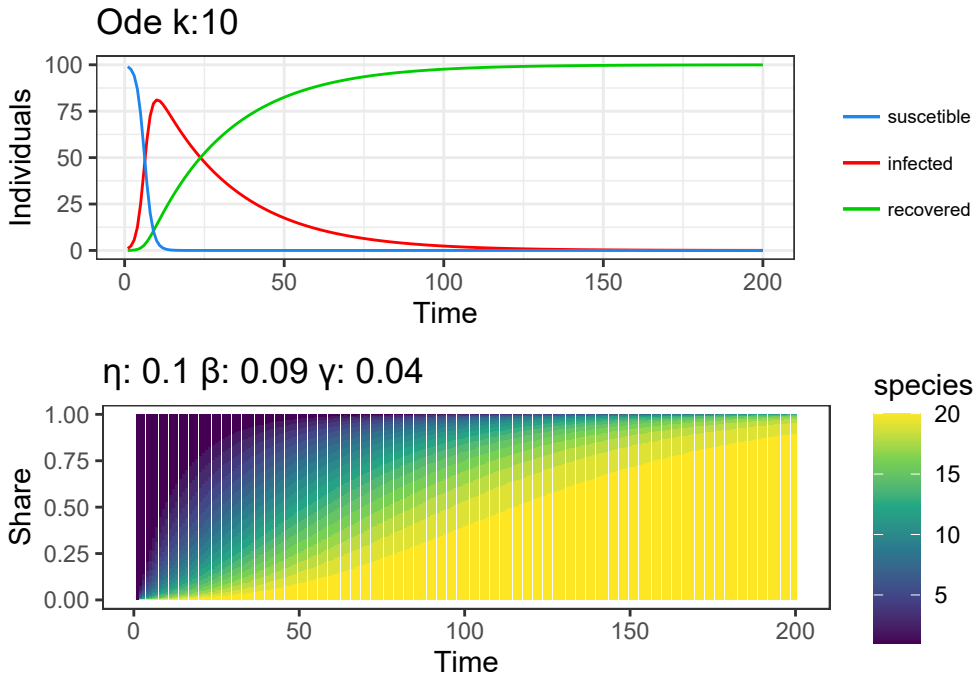


Figure 5.13: Results for the deterministic model using an average degree of 10.

Comparing this with the Monte Carlo simulation with the same parameters in Fig. 5.11 a, we can observe that the two models' evolution dynamics are very similar. In both, we notice the predominance of the last strain on the others as time passes and, generally speaking, the trend is very similar.

While if we want to focus on the infection spreading process:

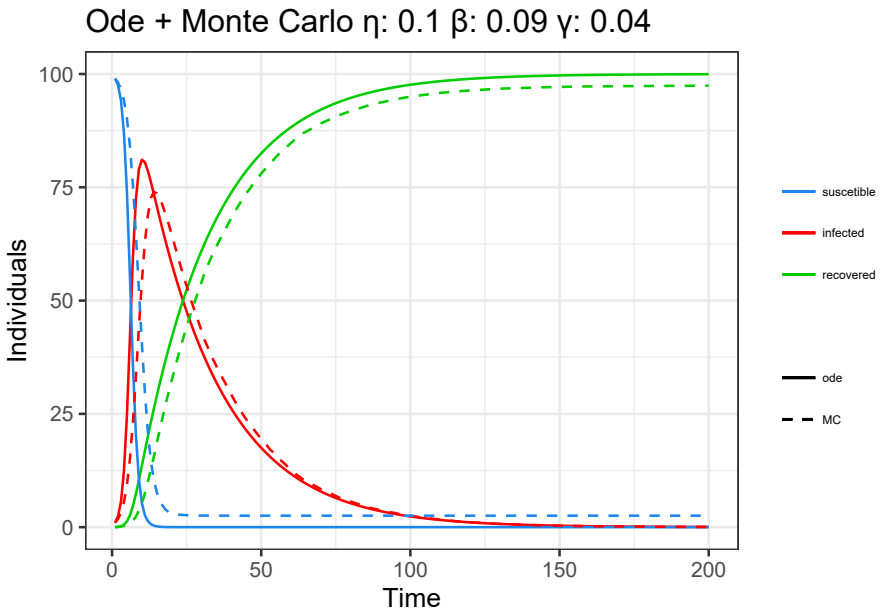


Figure 5.14: Comparison between our Monte Carlo and ODE models

We can assume that our mathematical model is satisfactory despite having several approximations.

6
Conclusions

In conclusion, the study of coupled dynamics between epidemic spreading on networks and pathogen evolution dynamics is a complex and highly interdisciplinary field with significant practical implications. This thesis has highlighted the importance of considering both aspects in order to understand and predict the spread and evolution of infectious diseases. We think that incorporating network structure and pathogen evolution into mathematical models can improve their accuracy and predictive power. Our findings suggest that future research should focus on developing more sophisticated models that better capture the complex interplay between these two dynamics, with the ultimate goal of informing public health policies and interventions to mitigate the impact of infectious diseases.

In the near future, our plan is to expand our study by avoiding the use of average degree approximation. This will enable us to reproduce any network structure that we desire, thanks to the factor $P(k|\tilde{k})$ in eq. (5.5). By incorporating this factor, we will be able to better capture the higher-order connectivity patterns in the network and reproduce more accurate network structures in the ODEs. Moreover it will also be important to investigate the impact of different types of fitness and how they would affect the dynamics if we incorporate them in the model. By studying the impact of different types of fitness and distinct β for each strain on the network dynamics, we can simulate how real-world pathogens spread and evolve over time.

Hopefully, we will be able to incorporate and validate Covid-19 data, not just number of cases, hospitalizations and deaths but also regarding how the pathogen behave, its fitness and transmission and mutation rates, in order to develop a much more realistic model, that more accurately reflect the real-world dynamics of the virus.
Furthermore, It could be interesting to explore the impact of prevention strategies, such as social distancing, quarantine, non-pharmaceutical interventions and vaccines, in the coupled dynamics by assuming our network structure can change over time.

Finally, the potential benefits of using these types of complex systems are clear. by improving our model, we can better understand the spread of the virus and test the effectiveness of different prevention strategies.

7 Appendix

In this appendix, we provide the details of the SIR model Monte Carlo code in R where infection, recovery and mutation take place:

```
while (t<tmax) {
    sus_prev <- sus_new
    inf_prev <- inf_new
    rec_prev <- rec_new

    #infection
    for(i in sus_prev){
        for (g in neigh[[i]]){
            if(g %in% inf_prev){
                if(i %in% inf_new){}
                else {
                    rdm_num1 <- runif(1, 0,1)
                    if(rdm_num1 <=beta){
                        inf_new<- c(inf_new, i)
                        sus_new <- sus_new[ !sus_new == i]
                        new <- info[info$inf_next==g,]
                        spec_g <- new$spec
                        new_val[spec_g] <- new_val[spec_g]+1
                        info<- rbind(info, data.frame(t_infe=t+1,  inf_next=i,
                                spec=spec_g, passed_from=g))
                    }
                }
            }
        }
    }

    #mutation
    for (i in inf_new){
        new <- info[info$inf_next==i,]
        spec_i <- new$spec
        n_row <- which(info$inf_next ==i)
        spec_new <-  mutation(spec_i)
        info[n_row,"spec"] <- spec_new$Species
        if(spec_new$Species!=spec_i) {
            new_val[spec_new$Species] <-new_val[spec_new$Species]+1
            new_val[spec_i] <- new_val[spec_i]-1
        }
    }

    #recovery
    for (i in inf_prev){
        rdm_num2 <- runif(1, 0,1)
        if(rdm_num2<=gamma){
            new <- info[info$inf_next==i,]
            spec_i <- new$spec
            rec_new<- c(rec_new,i)
            inf_new <- inf_new[ !inf_new == i]
```

```
            new_val[spec_i] <- new_val[spec_i]-1
        }
    }
    t <- t+1

    #filling the dataframe for the infection
    plotta <- rbind(plotta, data.frame(tempo=t, suscetible=length(sus_new),
        infected=length(inf_new), recovered=length(rec_new)))

    #filling the dataframe for mutation
    tot<- sum(new_val)
    if(tot!=0)
        incid <-rbind(incidenza, data.frame(tempo=t, v1=new_val[1]/tot,
            v2=new_val[2]/tot, v3=new_val[3]/tot, v4=new_val[4]/tot,
            v5=new_val[5]/tot, v6=new_val[6]/tot, v7=new_val[7]/tot,
            v8=new_val[8]/tot, v9=new_val[9]/tot, v10=new_val[10]/tot,
            v11=new_val[11]/tot, v12=new_val[12]/tot, v13=new_val[13]/tot,
            v14=new_val[14]/tot, v15=new_val[15]/tot, v16=new_val[16]/tot,
            v17=new_val[17]/tot, v18=new_val[18]/tot, v19=new_val[19]/tot,
            v20=new_val[20]/tot))
    else
        incid <-rbind(incidenza, data.frame(tempo=t, v1=new_val[1],
            v2=new_val[2],v3=new_val[3], v4=new_val[4],
            v5=new_val[5], v6=new_val[6], v7=new_val[7],
            v8=new_val[8], v9=new_val[9], v10=new_val[10],
            v11=new_val[11], v12=new_val[12], v13=new_val[13],
            v14=new_val[14], v15=new_val[15], v16=new_val[16],
            v17=new_val[17], v18=new_val[18], v19=new_val,
            v20=new_val[20]))
}
```

Bibliography

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Immanuel M Bomze. Lotka-volterra equation and replicator dynamics: a two-dimensional classification. *Biological cybernetics*, 48(3):201–211, 1983.

Ross Cressman and Yi Tao. The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences*, 111(supplement_3):10810–10817, 2014. doi: 10.1073/pnas.1400823111.

Valeria d'Andrea, Riccardo Gallotti, Nicola Castaldo, and Manlio De Domenico. Individual risk perception and empirical social structures shape the dynamics of infectious disease outbreaks. *PLOS Computational Biology*, 18:1–19, 02 2022. doi: 10.1371/journal.pcbi.1009760.

Selma Gago, Santiago F. Elena, Ricardo Flores, and Rafael Sanjuán. Extremely high mutation rate of a hammerhead viroid. *Science*, 323(5919):1308–1308, 2009. doi: 10.1126/science.1169202.

Tiberiu Harko, Francisco S.N. Lobo, and M.K. Mak. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. *Applied Mathematics and Computation*, 236:184–194, 2014. doi: https://doi.org/10.1016/j.amc.2014.03.030.

Natalia L. Komarova. Replicator–mutator equation, universality property and population dynamics of learning. *Journal of Theoretical Biology*, 230(2):227–239, 2004. doi: https://doi.org/10.1016/j.jtbi.2004.05.004.

Ralf H.J.M. Kurvers, Jens Krause, Darren P. Croft, Alexander D.M. Wilson, and Max Wolf. The evolutionary and ecological consequences of animal social networks: emerging issues. *Trends in Ecology Evolution*, 29(6): 326–335, 2014.

Jonathan M. Levine, Jordi Bascompte, Peter B. Adler, and Stefano Allesina. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, 546:56–64, 06 2017. doi: 10.1038/nature22898.

Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 2005. doi: 10.1038/nature03204.

M. E. J. Newman. *Networks: an introduction*. Oxford University Press, Oxford; New York, 2010.

H Allen Orr. Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10(8):531–539, 2009.

Karen M Page and Martin A Nowak. Unifying evolutionary dynamics. *Journal of theoretical biology*, 219(1): 93–98, 2002.

Anirudh Sridhar, Osman Yağan, Rashad Eletreby, Simon A. Levin, Joshua B. Plotkin, and H. Vincent Poor. Leveraging a multiple-strain model with mutations in analyzing the spread of covid-19, 2021.

Arie Zackay. Random genetic drift & gene fixation. 2007.

Xiyun Zhang, Zhongyuan Ruan, Muhuaa Zheng, Jie Zhou, Stefano Boccaletti, and Baruch Barzel. Epidemic spreading under mutually independent intra- and inter-host pathogen evolution. *Nature Communications*, 13:1–13, 2022. doi: 10.1038/s41467-022-34027-9.