UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Non-Markovian epidemic processes in complex networks

José Andrés Guzmán Morán

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP
SÃO CARLOS

José Andrés Guzmán Morán

Non-Markovian epidemic processes in complex networks

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos

August 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

G993n

Guzmán Morán, José Andrés
Non-Markovian epidemic processes in complex networks / José Andrés Guzmán Morán; orientador Francisco Aparecido Rodrigues. -- São Carlos, 2023.
144 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. Non-Markovian processes. 2. Epidemic models. 3. Complex networks. 4. Modular networks. 5. Weibull distribution. I. Aparecido Rodrigues, Francisco, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

José Andrés Guzmán Morán

Processos epidêmicos não-Markovianos em redes complexas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos

Agosto de 2023

To my family and friends.

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to Prof. Francisco Rodrigues for his invaluable supervision during my Masters degree. Especially, for the freedom he gave me to explore my own ideas and enthusiasm to the challenges found in the way. Francisco is an exceptional scientist, his ideas, insight and patience have had a fundamental impact to my development as a graduate student and future researcher. Also, I would like to acknowledge Prof. José Fontanari and Prof. Thomas Perón for the insight and constructive criticism during the qualification.

Special thanks to all my family in El Salvador, in particular, to Celia Morán and José Luis Guzmán my parents who always gave me support and comfort in the difficult moment and cultivated the curiosity that led to the path of academic career. Also thanks to my sister and brother Isabel Guzman and Fernando Vasquez, and cousins Sofia and Fernando Hernández. Furthermore, special thanks to my family in Brazil who have given me the strength to pass difficult moments through past six years: Lucas Perossi, Emanuel Alcantara, Fabio Pires, Gustavo Sabbag, William Arruda, Vinicius Moraes, and Roberto Flores.

The journey would have no meaning without the people we meet. Special thanks to my "Bandejão" friends who shared with me the joys and pains of academia, in particular Clara Vigor, Julia Marcolan, Rafael Garcia, Juliana Pirolla, Gustavo, and for particular love and support to Julia Jaccoud and Guilherme "Kibinho" Correr. Also, I can not forget the support of my hose friends Rafaela Fernanda, Sabrina, and especially Isadora Maria for her companionship. All my gratitude for the laughs and care of my friends outside USP João Vitor, Gabriel Silva, Matheus Vellosso, and especially to Ana Claudia and her family for the support and love in 2022.

The day in and day out (and sometimes late nights) would have been very painful without my lab and research friends Paula, Juniormar, Nastaran, Poppoff, Andressa, Aquisson, Rafael, Beatriz, and in particular Ricardo Tetti for the constant discussions and help with the project. Thanks also to Diogo da Silva, researcher of the group of Compelx Systems, for the insight and advice in last year of the Master's project. Thanks to Leonardo Martinussi, technician of LMACC, for the technical support.

I am deeply grateful for all the love and support of my Salvadoran friends, who have been with me every step of the way despite the physical distance between us: Franco, William, Evelyn, Damaris, Michele, Ana, Natalia, Nathalia, Pablo, Fernando and Dani.

Throughout these two years, I had the opportunity to be with a lot of people and make many friends that directly or indirectly contributed to my development. As this acknowledgment section cannot extend for more pages than the actual dissertation, I would like to express my gratitude to everyone that helped me during this period.

For the knowledge and patience, I would like to thank all my professors at ICMC-USP and IFSC-USP. Also, thanks to all the employees and staff of the University of São Paulo, especially to the cleaning and maintenance staff who suffered difficulties in 2023. Without you, none of this would be possible.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. I acknowledge FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) who also participated in the the financial support of the Masters degree. Additionally, the simulations were conducted utilizing the computational resources provided by the Center for Mathematical Sciences Applied to Industry (CeMEAI), which is also funded by FAPESP.

“Vivimos en un mundo maravilloso pero no necesariamente vemos las maravillas. Los despertadores son fundamentales, porque a veces da la sensación de que el ser humano pasa al lado de la belleza y no la ve”

(José “Pepe” Mujica)

RESUMO

GUZMÁN J.A. **Processos epidêmicos não-Markovianos em redes complexas**. 2023. 144 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Um dos pilares mais importantes na modelagem matemática de processos epidêmicos é considerar os processos de infecção e recuperação como Markovianos. Isto significa que, os tempos em que um indivíduo infectado irá contagiar os seus vizinhos e, o tempo em que irá recuperar-se são exponencialmente distribuídos. No entanto, a propagação de doenças depende diretamente do comportamento humano e de períodos de incubação de doença, que em geral são fenômenos descritos por distribuições não-exponenciais. Na presente pesquisa, objetivamos explorar diferentes problemas relacionados a processos epidêmicos não-Markovianos em redes complexas. Especificamente, realizamos simulações numéricas para estudar o modelo SIR e SIS não-Markovianos com um processo de infecção Weibull em diferentes tipos de redes: (i) redes regulares aleatória descritas pelo modelo Erdós e Renyi, (ii) redes sem escala feita com o modelo Barabási Albert e (iii) redes com caraterística de pequeno mundo descritas pelo modelo Watts Strogatz. Foi mostrado que considerar um processo de infecção não-Markoviano pode mudar significativamente o tamanho da epidemia e o valor da taxa crítica de transmissão. Quanto maior o parâmetro de forma α, associado ao envelhecimento da probabilidade de infecção, epidemias com menor tamanhos e maiores taxas críticas de transmissão são encontradas. Também, estudamos processo SIR não-Markovianos em redes modulares. Para um processo de infecção Weibull foi encontrado que estruturas de comunidades mais fortes e envelhecimentos positivos tendem a gerar epidemias com maiores tamanho. Além disso, nossos resultados sugerem que enquanto a taxa crítica de transmissão estiver abaixo da taxa efetiva de transmissão, envelhecimentos positivos podem enfatizar o efeito de estruturas de comunidades. Isto se deve ao fato de que o envelhecimento gera uma transmissão mais lenta, o que permite que as comunidades sejam mais efetivas em conter a doença, evitando a contaminação de comunidades saudáveis.

Palavras-chave: Modelos epidêmicos, Porcessos não-Markovianos, Redes complexas, Redes modulares, Simulações de epidemias.

ABSTRACT

GUZMÁN J.A. **Non-Markovian epidemic processes in complex networks**. 2023. 144 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

One of the cornerstones of mathematical epidemic modeling consists in assuming that infection and recovery can be described by Markov processes. This assumption implies that the inter-event times follow a negative exponential distribution. However, real-world epidemics are influenced by complex factors like human behavior and non-exponential incubation periods. As a result, there has been a growing interest in exploring non-Markovian epidemic processes. in the last decade. This work has the goal of exploring different problems concerning non-Markovian epidemics in complex networks. In particular, numerical simulations were conducted to study SIR and SIS non-Markovian epidemics using Weibull infection processes on different network models: (i) regular random networks (Erdos and Renyi model), (ii) scale-free networks (Barabási Albert model), and (iii) small-world networks (Watts Strogatz model). Our results reveal that considering a non-Markovian infection can significantly alter the epidemic size and threshold values. Increasing the shape parameter α, associated with aging in the probability of infection, leads to smaller epidemic sizes and higher critical effective rates. Our investigation also extends to the study of SIR non-Markovian processes in modular networks. For a Weibull infection process, we verify that strong community structures and positive aging contribute to larger epidemic sizes. Furthermore, the results suggest that as long as the critical transition rate remains below the effective rate chosen, positive aging processes can hence the role of communities in hindering the disease propagation to the entire network. This is caused by the slower propagation associated with the positive aging process, which prevents the disease from reaching healthy communities.

Keywords: Epidemic models, Non-Markovian processes, Complex networks, Modular networks, Epidemic simulations.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BA	Barabási Albert network model
ER	Erdős and Rényi network model
GA	Gillespie Algorithm
HMF	Heterogeneous mean-field approach
LFR	Lancichinetti–Fortunato–Radicchi benchmark algorithm
QMF	Quench mean-field approach
SF	Sale-free networks
SIR	Susceptible - Infected - Recovered
SIS	Susceptible - Infected - Susceptible
WS	Watts Strogatz network model

LIST OF SYMBOLS

σ — Transmission or infection rate

γ — Recovery rate

R_0 — Basic reproduction number

G — Network or Graph

N — Total number of individuals or total number of nodes in the case of network models

L — Total number of edges or vertices of a network

k_i — Degree of node i

$\langle k \rangle$ — Average degree of a network

$\langle k \rangle$ — Degree distribution

A — Adjacency matrix

ν — Exponent of the degree distribution of a scale free network

ξ — Rewiring probability of the Watts Strogatz model

λ_c — Critical transmission rate for epidemics on networks

λ_c^{HMF} — Heterogeneous mean-field approximation for the critical transmission rate

λ_c^{QMF} — Heterogeneous mean-field approximation for the critical transmission rate

χ — Susceptibility of a system

Δ — Vulnerability of a system

λ_c^{NIMFA} — NIMFA approximation for critical transmission rate for a non-Markovian SIS epidemic with Weibull infection

λ_c^{UB} — Upper bound of the critical transmission rate for a non-Markovian SIS epidemic with Weibull infection

T_{inf} — Random variables that describe the interarrival times of infection for a non-Markovian

epidemic process

T_r — Random variables that describe the interarrival times of recovery for a non-Markovian epidemic process

τ_{inf} — Random time at which the susceptible node of the active link will get infected

τ_r — random times at which a node spontaneously recovers

$\psi_{inf}(\tau_{inf}$ — Probability distribution of τ_{inf}

$\psi_{inf}(\tau_{rec}$ — Probability distribution of τ_r

γ — Recovery rate of the non-Markovian epidemic process with Weibull infection and Poisson recovery

α — Shape parameter of the Weibull distribution that describes the infection process for a non-Markovian epidemic

β — Scale parameter of the Weibull distribution that describes the infection process for a non-Markovian epidemic

$\Psi_{inf}(\tau)$ — Survival probability

$\lambda(\tau)$ — Hazard rate of a non-Markovian epidemic process

λ_{eff} — Effective transmission rate of a non-Markovian epidemic process

Υ — Exponent of the power law distribution of degree in the LFR benchmark model

Ω — Exponent of the power law distribution of community sizes in the LFR benchmark model

s_{max} — Maximum community size in the LFR algorithm

s_{min} — Minimum community size in the LFR algorithm

μ — Mixing parameter in the LFR algorithm

M — Number of communities in a modular network

T_M — Time when the disease has infected at least one individual in each community of a modular network

T_l — Epidemic life.

$\Delta\rho$ — Difference in prevalence between two epidemic processes.

CONTENTS

CHAPTER

1

INTRODUCTION

Throughout the course of human history, infectious diseases have been a significant threats to society, particularly when they spread rapidly and give rise to epidemic outbreaks [Anderson and May 1992]. In recent times, the world has witnessed the devastating consequences of an uncontrolled pandemic caused by the SARS-CoV-2 virus. The far-reaching impact of the COVID-19 pandemic has highlighted the urgent need to develop effective strategies to combat such infectious crises, leading to a surge in research projects exploring epidemics and disease propagation [Cai, Fry and Wagner 2020]. Although the scientific community has given particular attention to epidemic processes in the last few years, the interest in propagation phenomena is not new.

One of the main approaches used to understand epidemics is the study of mathematical models. In this context, one of the first and most relevant works was developed by Kermack and McCormick in the 1920s [Kermack, McKendrick and Walker 1927], which was followed by a golden age in mathematical modeling of epidemics in mid-20th century [Pastor-Satorras *et al.* 2015]. These first models were based on the idea of classifying individuals in compartments according to their state related to the infection, for example, infected, recovered, exposed, etc. Unfortunately, the majority of these models used the homogeneously mixed assumption, which implicated that all individuals can interact with each other. This assumption is far from reality, where spatial limitations determine how people interact.

Independently, in the second half of the 20th century, the study of complex systems emerged as a new field of research. This kind of system is characterized as being hopelessly

complicated, as it can not be understood just as a sum of the individual behavior of its components [Barabási and Pósfai 2016]. The concept of complex networks, as a combination of graph theory and probability theory, is one of the most useful tools to face the challenges of complex systems. The idea of nodes or individuals connected is very powerful, and network science can help to describe fundamental aspects of many real systems like human interaction, computer systems, species interaction, and many more [Vespignani 2009, Butts 2009, Jackson 2008]. The power of complex networks becomes particularly evident in the analysis of spreading phenomena, which often manifest in systems such as those aforementioned. Take, for instance, the spread of epidemics, which relies heavily on human interactions; the proliferation of computer viruses, facilitated by the internet; or the dissemination of fake news across social media networks. In each case, the underlying structure of complex networks plays a crucial role in shaping the dynamics of these spreading processes.

In the first decades of the new century, a second wave of interest in epidemic modeling appeared, mainly focusing on combining the concept of complex networks with epidemic models [Pastor-Satorras *et al.* 2015]. With this approach, some spatial characteristics could be added to further deepen the understanding of real epidemics. The first work in this field was elaborated in 2001 by R. Pastor Satorras and A. Vespigani [Pastor-Satorras and Vespignani 2001], where a mean-field approximation proved that heterogeneity in complex networks can help spreading processes. From this point forward, epidemics in networks have been a growing research area, and it has been characterized by a variety of approaches: from mean field approximations to simulation-based research [Keeling *et al.* 2011, Keeling and Eames 2005, Pastor-Satorras *et al.* 2015, Boguñá *et al.* 2014, Masuda and Rocha 2018]. Also, a lot of interest in these topics comes from the versatility of these models, as they can be applied to describe other dynamical processes. For example, these epidemic models can be used to study information diffusion, rumor spreading, and even the adoption of social behaviors [Bikhchandani, Hirshleifer and Welch 1992, Goffman and Newill 1964].

One of the main simplifications used in epidemic modeling is considering that the processes involved are well described by Markovian processes. The central idea of this approximation is to consider that the probabilities of the characteristic inter-event times of an epidemic (infection, recovery, etc) come from an exponential probability distribution, and have constant transition rates. These models are memory-less, which means the probability of future events does not depend on the past.

This simplification has led to many analytical and numerical results that have helped deepen the understanding of real epidemic processes. However, this approximation has been challenged in recent years. The increase of big databases of human interactions has revealed that homogeneous temporal processes are not always adequate to describe networks of empirical data [Gonzalez, Hidalgo and Barabasi 2008, Stouffer, Malmgren and Amaral 2005, Brockmann, Hufnagel and Geisel 2006]. Many times, these empirical studies have shown inter-event times with heavy-tailed distributions, instead of exponential ones [Barabási 2005]. For this reason, the study of non-Markovian processes, which are in general described by non-exponential probability distributions, has been increasing in popularity [Vazquez *et al.* 2007, Boguñá *et al.* 2014, Starnini, Gleeson and ná 2017, Feng *et al.* 2019].

Some of these researches have shown that epidemic processes can drastically change when considering non-Markovian transition processes and indicating that in many cases the memory effect is not just a correction of the Markovian models [Mieghem and Bovenkamp 2013, Boguñá *et al.* 2014]. One of the most popular non-Markovian models studied is the non-Markovian SIS process with Weibull infection, where the inter-event times of the infection process are described by a Weibull distribution. In this model, as infected nodes spend more time in the infected state, the likelihood of infecting others changes. This consideration has been reported to significantly modify key characteristics of the process like the total number infected individuals at the end of the process, the time until the steady state is reached, and the threshold values associated with process [Cator, Bovenkamp and Mieghem 2013, Boguñá *et al.* 2014, Mieghem and Liu 2019].

Although the study of non-Markovian epidemics has increased in the past years, there are still many answered question. Unfortunately, many techniques and results of Markovian theory do not have a direct translation to non-Markovian systems. As mentioned, nowadays it is well known that considering non-Markovian aspects can alter the outcome of an epidemic, but how this happens for more complex epidemiological models, like a SIRS or SEIR model, or the effects in network with specific characteristics, like modularity, clustering and temporal changes, are still open problems.

The present dissertation is dedicated to the exploration of some of this problems of non-Markovian epidemic processes in complex networks. Specifically, we focused on the simulation of a SIS and SIR non-Markovian process with Weibull infection in different network models with different topological characteristics.

1.1 Chapter organization

This text is organized into seven chapters. The first one (the current chapter) contains the introduction and chapter organization. The next five chapters are organized into topics with the goal of constructing a logical sequence. Chapter 2 called “Fundamental Concepts: epidemic models and Complex Networks” has e goal of presenting the basic concepts needed to develop this research. This chapter is divided into two parts, the first one, in section 2.1 presents the SIS and SIR compartmental models and their main characteristics. The second one, in section 2.2, presents a brief revision of Complex networks, from basic properties to a quick review of the most popular models.

Once we establish the basics of epidemic processes and complex networks we will combine them in chapter 3. To do this we make a revision of stochastic processes in section 3.1, which are a fundamental tool used in describing epidemics on networks. Sections 3.1.4 and 3.1.4.1 are particularly relevant for this project as they present the idea of Renewal processes, which are the heart of non-Markovian epidemics. Once we described the most important ideas of the stochastic process we present the main results and approaches used in epidemics in complex networks in section 3.2.

In Chapter 4 we start with the main topic of the text: non-Markovian epidemics in Complex Networks. Using the ideas, concepts, and methods described in the previous chapters, we present the standard approach used in the literature for describing non-Markovian epidemic processes. The rest of the chapter is a literature revision of the most important works in this field 4.3.

Followed by this, chapter 5 and 6 present the main results of the research associated with this dissertation. In the first of them, chapter 5, we explore the SIS and SIR non-Markovian epidemic for three network models: regular random networks described by the Erdős and Rényi model, heterogeneous scale-free networks, made by the Barabási Albert model, and the Watts Strogatz networks with small world characteristic. We verify some results present in the literature and further explore some characteristics associated with the critical rate of these systems. In chapter 6, we present the results of the study of a SIR non-Markovian process with Weibull infection in networks with community structure.

The last chapter 7 contains the conclusions, final considerations, limitations and possible future works.

CHAPTER

2

FUNDAMENTAL CONCEPTS: EPIDEMIC MODELS AND COMPLEX NETWORKS

In this section, we cover some of the fundamentals and basic concepts needed to develop this research project. These basic concepts are divided into two big areas: basic epidemic models and fundamentals of complex networks. The first part is a brief revision of the traditional homogeneously mixed SIS and SIR compartmental models, and the second part is a revision of basic concepts used in network science: properties and measurements of complex networks, models used to construct them, and modularity in networks. The next chapter 3, covers a second part of fundamental concepts but focused on epidemics in complex networks, which is a combination of the ideas presented in the current section.

2.1 Basic Epidemic Models

Mathematical epidemic models are a fundamental approach to understanding and controlling infectious diseases [Pastor-Satorras *et al.* 2015]. Develop a mechanism of how diseases spread between individuals is the first important step in modeling [Kiss, Miller and Simon 2017]. In this context, the work of Kermack and MacKendrick [Kermack, McKendrick and Walker 1927] defined the modern approach to epidemic modeling. This is, the compartmental models, which are a very general technique that consists in assigning to each individual a label that describes its status in relation to the disease. For example, an individual could be assigned the label of “Infected” if it has a pathogen and can infect other people.

This basic idea can be used in numberless ways to create models that mimic real epidemics. Most of the models studied in the last century where characterized for being homogeneously mixed deterministic models [Kiss, Miller and Simon 2017]. In this approach, the homogeneously mixed characteristic implies that all individuals can interact with each other, and as consequence we don't need to care about which compartment each individual belongs, and rather we can just look to the overall fraction of individuals in each compartment. This models are deterministic, and differential equations are used to study the dynamics of the total number of individuals that belong to a specific group (Infected, Susceptible, etc). Now we will present two of the most popular compartmental models in the lens of this deterministic approach: the SIS and SIR model. Although this project centers its attention in stochastic simulations for epidemic processes, this deterministic approach in homogeneously mixed systems is important as it can be used to introduce the basic concepts around epidemic models.

2.1.1 SIS model: Susceptible - Infected - Susceptible

The Susceptible - Infected - Susceptible (SIS) model considers that an individual that has recovered from a disease can get infected again. This means that in an ongoing epidemic a person can go trough a cycle where they get infected then recover, then infected again and so on. It is a very simplified model, but can be used as a basic approach to describe STDs or viral diseases characterized by viruses that mutate constantly over time [Keeling and Eames 2005].

In this model, the a population of N individuals is divided into two groups: "I", which are the infected ones, and "S" which are the susceptible ones. An infection occurs between a susceptible individual and an infected individual. In the case of a continuous time approach we can consider a constant transmission or infection rate (σ), , which can be understand as the number of infection-transmitting contacts infected individuals make per unit of time. Another way to understand this constant is as the inverse of the mean time between successfully infections of an infected person. With the same logic, we consider a constant recovery rate (γ) [Keeling and Eames 2005]. In equation (2.1) we present a diagram of the infection cycle.

$$S \xrightarrow{\sigma} I \xrightarrow{\gamma} S \tag{2.1}$$

In a population the total rate of infectious contacts is σI, but just a fraction S/N of these contacts will occur with susceptible individuals. The recovery does not depend on the contacts,

so the rate is the same independently of the number of infected or susceptible individuals. Considering a fully mixed population (everyone can contact everyone), and without considering birth and death, which implies N is a constant value, we can construct a model by writing a set of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\sigma\frac{S}{N} + \gamma I \\ \frac{dI}{dt} &= \sigma I\frac{S}{N} - \gamma I \\ S(t) + I(t) &= N \end{aligned} \tag{2.2}$$

As N is just a constant we can perform a normalization considering new variables: $s = S/N$ and $i = I/N$

$$\begin{aligned} \frac{ds}{dt} &= -\sigma i s + \gamma i \\ \frac{di}{dt} &= \sigma s - \gamma i \\ s(t) + i(t) &= 1 \end{aligned} \tag{2.3}$$

We can substitute $s = 1 - i$ and obtain an ordinary differential equation that involves just i and t:

$$\frac{di}{dt} = -\sigma i(1-i) + \gamma i \tag{2.4}$$

We won't go through all details of the solution, but Equation (2.4) can be easily solved using variable separation, a complete procedure can be found in [Kiss, Miller and Simon 2017]. However, the specific equation that describes i as a function of time t isn't always useful, and for most complex models, it is even impossible to express this relation in an analytic form. Because of this, we will focus on the qualitative behavior by describing the steady states. In this case, we have two steady states that depend on the relation of the transmission rate σ and recovery rate. For this, we define the basic reproduction number $R_0 = \sigma/\gamma$, which can be understood as an effective transmission rate. If $R_0 = \sigma/\gamma < 1$ the recovery process "wins" over the transmission process, and for large time $i \to 0$, on the other hand if $R_0 = \sigma/\gamma > 1$ we have an endemic steady state $i \to (1 - \frac{\gamma}{\sigma})$. In the second case, the steady state occurs when in average the same number

of individuals that get infected are recovering per unit of time, causing the fraction of infected and susceptible individuals to stay constant [Kiss, Miller and Simon 2017] In figure 1 we can see a numerical solution of equations (2.3) for different values of R_0.

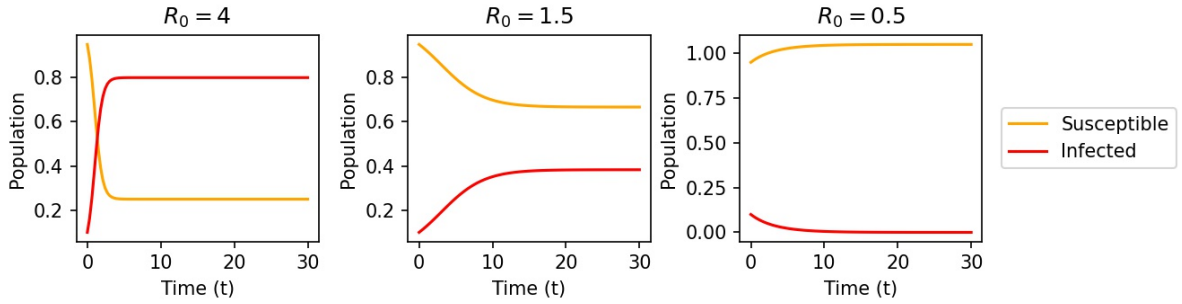

Figure 1 – Numerical solution for the deterministic SIS model. The graphs show the fraction of infected and susceptible individuals as a function of time for different values of R_0.

Source: Elaborated by the author.

In $R_0 = 1$ a phase transition occurs where the behavior of the epidemic when $t \to \infty$ changes dramatically.

2.1.2 SIR model: Susceptible - Infected - Recovered

The Susceptible - Infected - Recovered (SIR) model considers that after some time of infection, the individual infected recovers from the disease and can't get infected again. This idea can be applied for diseases where individuals develop lifelong immunity [Pastor-Satorras *et al.* 2015].

For the deterministic continuous time approach to solve the SIR model we consider a transmission rate σ and a recovery rate γ. These rates have essentially the same meaning as the rates described for the SIS model, with the exception that in the recovery process, infected individuals don't return to the susceptible group, but go to a new group of recovered individuals.

$$S \xrightarrow{\sigma} I \xrightarrow{\gamma} R \tag{2.5}$$

Again, in a population of size N with I infected , S susceptible and R recovered individuals the total rate of infectious contacts is σI, but just a fraction S/N of these contacts will occur with susceptible individuals. Considering the assumption of a fully mixed population with a constant

number of individuals, we can write a system of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\sigma I \frac{S}{N} \\ \frac{dI}{dt} &= \sigma I \frac{S}{N} - \gamma I \\ \frac{dR}{dt} &= \sigma I \\ S(t) + I(t) + R(t) &= N \end{aligned} \tag{2.6}$$

If we perform a normalization where $s = S/N$, $i = I/N$ and $r = R/N$ it can be obtained:

$$\begin{aligned} \frac{ds}{dt} &= -\sigma i s \\ \frac{di}{dt} &= \sigma i s - \gamma i \\ \frac{dr}{dt} &= \gamma i \\ s(t) + i(t) + r(t) &= 1 \end{aligned} \tag{2.7}$$

This is still a vary simple model, but the addition of a third class of individuals complicates the solution of the differential equations causing the impossibility of finding an analytic solution. As a result, in this problem a qualitative analysis of the steady states appears to be best way to understand the behavior of the model.

There is only one steady state for the fraction of infected people: $i \to 0$ as $t \to \infty$. So in contrast to the SIS model, the SIR model does not allow an endemic solution where i stays constant. This makes sense because in the worst case scenario where all individuals get infected at some point the lifelong immunity guarantees that, for a time far from the first contagion, there won't be any more individuals that can get infected. Although there is just one steady state for i, there are other aspects of en epidemic that can be studied. For instance, we can look at the size of the epidemic ρ. For the SIR model this measures is defined as the fraction of recovered people as $t \to \infty$. Considering a small population of infected people at the beginning of the epidemic we can obtain an implicit equation for ρ [Kiss, Miller and Simon 2017]:

$$\rho = 1 - e^{-\frac{\sigma r_\infty}{\gamma}} \tag{2.8}$$

For $R_0 = \sigma/\gamma > 0$ there are two solutions for (2.8): a trivial solution $\rho = 0$ that corresponds to case wheres $i(0) = 0$, and a non trivial solution $\rho > 0$. In contrast, if $R_0 \leq 1$ we have

that $\rho \to 0$ is the only solution, and corresponds to the case where i will just decrease over time. This means that, just as in the SIS model, $R_0 = 1$ correspond to a phase transition. Figure 2 shows some numerical solutions of (2.6):

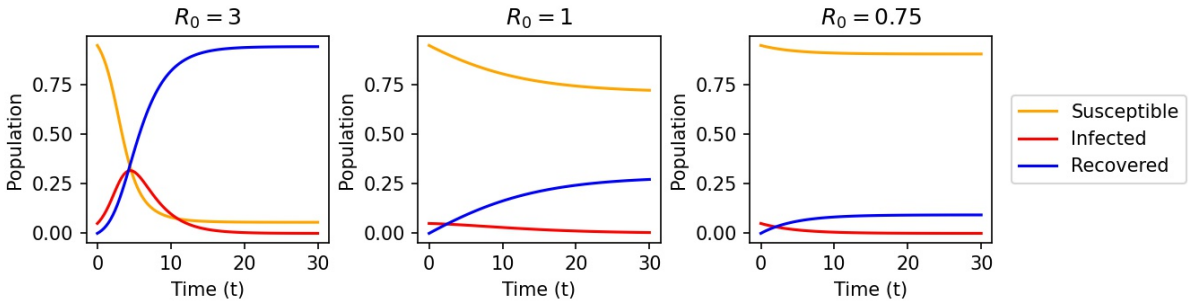

Figure 2 – Numerical solution for the deterministic SIR model. The graphs show the fraction of infected, susceptible and recovered individual as a function of time for different values of R_0.

Source: Elaborated by the author.

The homogeneously mixed models have been very important for epidemic modeling, especially for their simplicity. Unfortunately, the lack of spatial and structural characteristics is one of the main disadvantages of the homogeneously mixed assumption as, in real life, humans are constrained to spatula limits.

2.2 Complex Networks: Basic Concepts

In this subsection, we cover the basic concepts and ideas needed to understand the use of complex networks for this project. First, we present a brief idea of what a complex network is, what are its characteristics and how can we measure them. Then we go through some classes of networks and how we can construct them. And finally present some ideas related to modularity in networks, which is a topic of interest for this project.

A network can be understood as a collection of points joined together in pairs by lines [Newman 2018]. Usually, when these networks have a large number of individuals and don't have trivial topology we called them Complex Networks [Barabási and Pósfai 2016]. This is a simple definition, but captures the essence of complex networks, and with this basic idea is easy to see how this concept can be used to describe biological, social, and physical systems composed by a large number of interacting individuals [Barabási and Pósfai 2016].

2.2.1 Terminology and basic parameters of networks

Although networks and graphs are rigorously different concepts, for this project we use this terms interchangeably, and in general use G to refer to a generic network . A network is an abstraction of a system composed by nodes or vertices, which are the components of the system that often interact, and links or edges, which are the explicit relation or interaction between two components [Barabási and Pósfai 2016]. We can characterize a network by its size, generally described by$N$, total numbers of nodes, and LN , total number of edges.

In general, each node of the system has a label that could be any relevant information about the node. However, it usual to use natural numbers as labels, as it facilities the construction of other concepts, like the adjacency matrix. In this sense, we could have a network G with N nodes characterized by a labels: $1,2,3...,i,...,N$. With this logic, edges are usually represented as pair of numbers according to the label of the nodes it connects, for example, if node i is connected to node j than the edge is represented by: (i, j).

Networks can be directed if the links have a direction, in other words, a connection from node i to node j does not imply that a connection from j to i exists. Naturally, we define an underacted network as one where links don't have directions, so the existence of an edge from i to j doesn't mean that exists a connection the other way around. Also, networks can be weighted if there are different kinds of links, characterized by weights, or unweighted if all links represent the same kind of relation. These concepts are important because for this project we work exclusively with underacted unweighted networks, and this is fundamental as it can change how we define some basic parameters and properties.

To characterize a network we need to define some concepts [Barabási and Pósfai 2016, Newman 2018]:

1. **Degree:** For a node i, we define the degree as the total number of links between this node and any other node. We use the notation k_i as the degree of node i.

2. **Neighbors:** For a node i, we define the neighbors as the collection of nodes j that are connected with i by one link. This concept is directly related to the concept of the degree: if $k_i = 3$ it means node i has three links and, as we are considering underacted networks, it also means that i has 3 neighbors.

3. **Average Degree**: This is a fundamental property that can characterize and differentiate

networks. Is defined as the sum of all degrees from every node divided by the total number of nodes. We denote this quantity with the character: $\langle k \rangle$:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{2.9}$$

4. **Degree distribution:** Is a function that returns the probability that a randomly chosen node in the network has degree k. We use the notation p_k or $p(k)$. This is a fundamental concept to understand how some networks are constructed.

5. **Adjacency matrix**: For mathematical purposes, we usually represent a network using its adjacency matrix. This is a matrix A with N rows and N columns (each corresponding to a specific node), and its elements for an unweighted network are defined as:

$$A_{ij} = \begin{cases} 0, & \text{if there isn't a link between i and j.} \\ 1, & \text{if there is a link between i and j.} \end{cases} \tag{2.10}$$

For underacted networks the adjacency matrix is symmetric, so $A_{ij} = A_{ji}$

There are many more parameters and concepts associated with networks such as: connectedness, centrality, clustering coefficient, paths, distances, etc. But, for this project, these are the essential definitions, if some other concept emerges later in the text it will be defined in its own context.

2.3 Network Models

So far we were concern with defining the essential terminology of network theory. But, haven't talk about how can these networks be generated. Network models help us develop algorithms and strategies for creating synthetic networks that captures characteristics of real systems. In this subsection we will present some of the most popular ones.

2.3.1 Random Networks or Erdős and Rényi model

Paul Erdős and Alfréd Rényi, two Hungarian mathematicians, are considered to be the founders of modern network science. They did this by introducing probability theory in graph theory [Fienberg 2012]. In this context, they first introduced the Erdős and Rényi network model (ER) in 1959 [Erdös and Renyi 1984, Erdös and Rényi 1959], which is one of the most

fundamental models. In honor to this mathematicians this type of networks is often referred as Erdős and Rényi network or *ER* networks.

Although Erdős and Rényi receive a lot of credit for the Random Network model, Edgar Gilbert, an American mathematician, introduced the same concept independently and contemporaneously [Fienberg 2012]. As a consequence, we can define random network in two ways [Barabási and Pósfai 2016]:

- $G(N,L)$ Model: N nodes are connected with L randomly placed links. This is the original definition used by Erdős and Rényi.

- $G(N,p)$ Model: Each of the N nodes are connected to the rest $N-1$ other vertices with probability p. Definition introduced by Gilbert.

The two definition generate random networks and are equally valid. However, many times the second one is chosen because real networks rarely provide a specific number of links [Barabási and Pósfai 2016]. The algorithm for generating random networks given N and p is:

1. Begin with N disconnected nodes.

2. Randomly select two nodes i and j. Generate a random number u between 0 and 1, if the $u > p$ then add a link between i and j.

3. Repeat step 2 for the N(N-1)/2 node pairs.

The random characteristic of this type of network causes that two realizations of the same algorithm generate different networks. However, if the parameters stay the same, $\langle k \rangle$ will also stay constant trough different realizations. It is can be prove the average degree is given by [Newman 2018]:

$$\langle k \rangle = p(N-1) = \frac{2\langle L \rangle}{N} \tag{2.11}$$

The strategy of taking a random number and compare it to the probability p corresponds to a binomial process. So the degree distribution is given by:

$$p_k = \begin{pmatrix} N-1 \\ k \end{pmatrix} p^k (1-p)^{N-1-k} \tag{2.12}$$

In the thermodynamic limit, $N \to \infty$, this distribution can be approximate to a Poisson distribution:

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{2.13}$$

Figure 3 displays some ER networks with different values of p. For small p the network is spare and as this parameter increases the network gets denser till $p = 1$ where a completely connected network is generated.

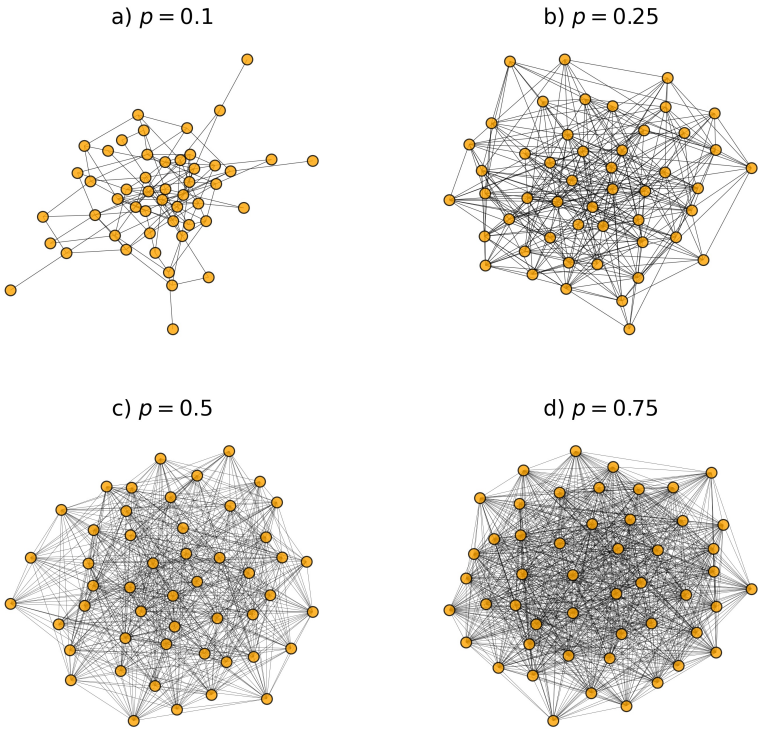

Figure 3 – Visual representation of different Erdős and Rényi random networks with different values of p.

Source: Elaborated by the author.

This model has been extensively studied, specially because of its simplicity. Unfortunately, it has been shown many times that real networks don't behave like random networks [Albert, Jeong and Barabasi 1999, Redner 1998]. For example, this model doesn't have hubs (nodes with large degree), which are a fundamental characteristic of social networks [Watts and Strogatz 1998].

2.3.2 Scale-free Network and Barabási–Albert (BA) model

When we try to fit data from real social networks with a random network model, in general, the results are very disappointing. A lot of real networks don't have e Poisson's degree distribution, and the presence of nodes with large degree is frequent [Barabási and Pósfai 2016]. One classic example is the study of the World Wide Web and networks of scientific citation [Albert, Jeong and Barabasi 1999, Redner 1998].

As a solution it is possible to build networks that have a power law degree distribution, which allows the existence of nodes with large degrees. This networks are also called Sale-free networks (SF), and equation (2.14) shows a generic form of its power law distribution:

$$p_k \propto k^{-\nu} \tag{2.14}$$

Where ν is the degree exponent and is typically in the range $2 < \nu \leq 3$

The presence of nodes with large degrees in scale-free networks compared to ER networks can be traced to the differences in the degree distribution. Power law distributions have "heavy tails", so the value of p_k for k far from the average degree $\langle k \rangle$ is orders of magnitude bigger than its counterpart in an ER network. The result of this characteristic is the presence of a few nodes with very large degrees and many nodes with few links. This property makes scale-free networks heterogeneous networks, for the large diversity of degrees, in contrast with the regular networks, like the ER network, which are called homogeneous networks. We can see a compression of these distributions in figure 5 and in image **??** we can appreciate an example of a heterogeneous network with the presence of 1 hub and a regular network, generated by the ER model.

The first model proposed to reproduce networks with power law distributions was developed by Albert-László Barabási and Réka Albert [Barabasi and Albert 1999]. This model, called Barabási Albert network model (BA), is based on two important assumption: the network are not treated as static objects, at each step of the generating process new nodes are constantly

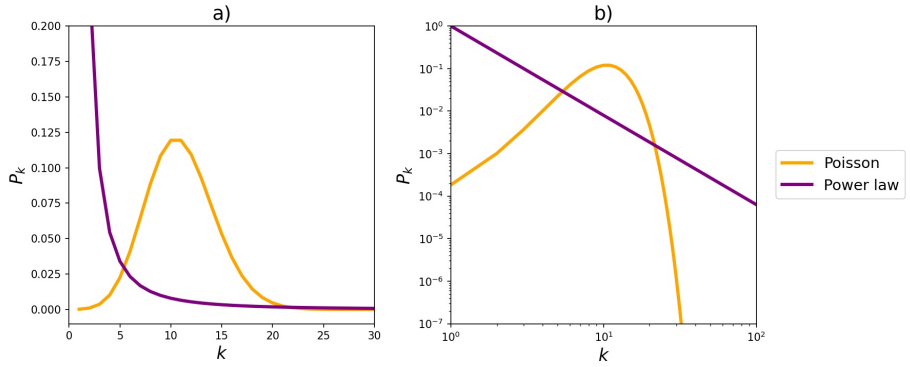

Figure 4 – Graphical representation of the Poisson and Power law distribution. In a) on regular scale, and in b) with logarithmic scale.

Source: Elaborated by the author.

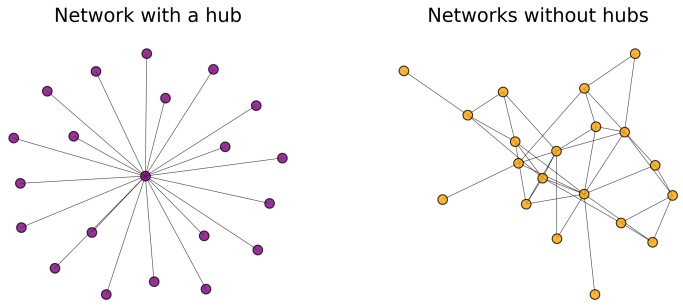

Figure 5 – Visual representation of two networks: on the left a heterogeneous network with the presence of 1 hub, and on the right a regular network generated by the ER model.

Source: Elaborated by the author.

being added; new nodes prefer well connected nodes. The second assumptions is known as the "preferential attachment", and follows the idea of "rich get richer", which means that nodes with already high degree have a bigger probability of getting new links. A simple algorithm idea for this model follow:

1. Begin with a connected network of n_0 nodes.

2. We add a new node to the network. This node is connected to m existing nodes with a probability proportional to the degree of the original nodes. So the probability of connecting a new node with an existing node j is:

$$p_j = \frac{k_j}{\sum_i k_i} \tag{2.15}$$

3. The second step is repeated until we reach a network with properties we desired.

This model generate a scale-free network with power law exponent ν near 3. This formulation is considered a breakthrough for network science, but there were still some characteristics of real systems that were being left out, specifically the small world characteristic.

2.3.3 Small World Network and Watts-Strogatz model

The idea of "six degree of separation" first appeared in pop culture at the beginning of the 1920's [Fienberg 2012]. This concept stands that to reach any person in the world we just need six or fewer social connection. Although this might not be completely true, it captures an interesting characteristic of real human networks: the small world property.

Stanley Milgram, an American social psychologist, conducted many experiments in the mid 20th century that tested the the small world concept in real social networks. The results suggested social network are characterized by short path-lengths and high clustering, this is, every individual can reach any other individual in a small number of steps [Travis and Milgram 1977]. Networks that present this behavior were called "small-world networks".

Again, just like the lack of hubs in real networks, the small-world characteristic wasn't captured by ER model. At the end of the 1990s Duncan Watts and Steve Strogatz proposed a new model to characterized and generate small world graphs, also called Watts Strogatz network model (WS) [Watts and Strogatz 1998].

The remarkable characteristic of the algorithm is that we "tune" the small world characteristic using just one parameter ξ . The algorithm proposed by Watts and Strogatz follows:

1. We start with a ring network. This is, nodes are placed in a one-dimensional ring with periodic boundary conditions.

2. Connect every node with its k nearest neighbors in both directions.

3. Rewire each node with probability ξ.

If ξ=0, we stay with a ring network. In the other extreme case if p=1, we rewire all the links, and end up with a random network. But, for intermediate values of ξ we generate networks with small world characteristics. In figure 6 there are some networks generated with the WS model for different values of ξ.

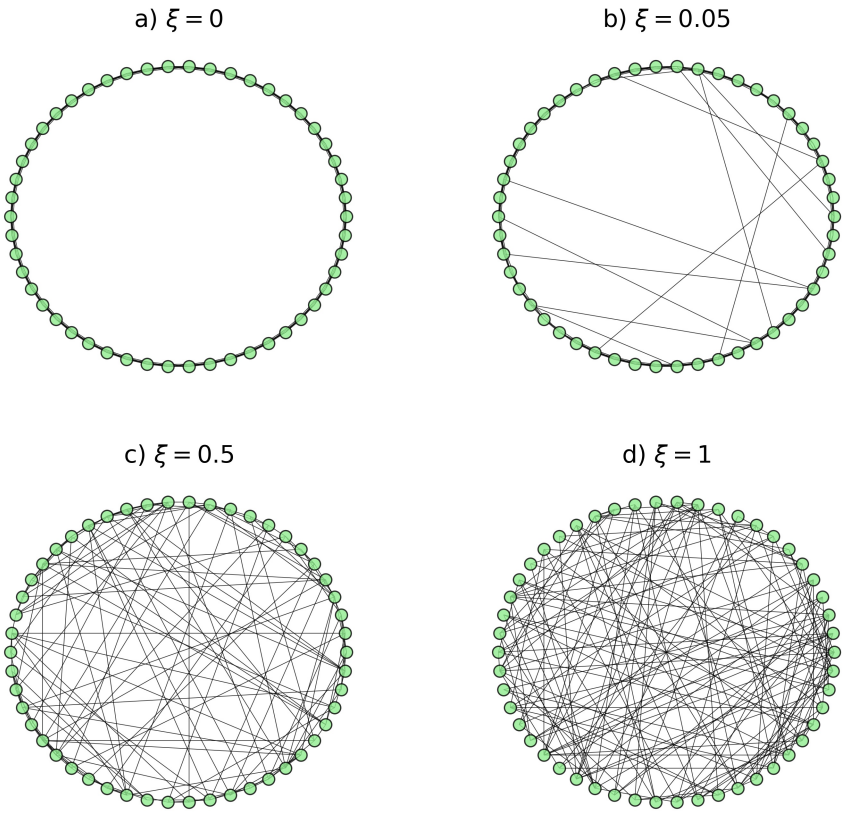

Figure 6 – Visual representation of different Watts Strogatz network, each one with a different values of ξ.

Source: Elaborated by the author.

The main disadvantage of this model is it does not capture heterogeneity in the degree distribution. Just like in the ER network, we end up with nodes having too similar degrees, and having a small probability for hubs.

2.3.4 Modular networks

In many social systems, individuals don't have the same probability of connecting to any other individual. Humans tend to group with other people who have similar social characteristics. None of the models studied so far take this property into account, so we need to introduce a new concept: communities

Networks with community structures, also called modular networks, are graphs where

nodes have a higher likelihood of connecting with other nodes within the community than with nodes outside the community [Barabási and Pósfai 2016]. This generates networks with groups of nodes densely connected, and spare connections between nodes form different groups. This modular structure has been proven to be relevant for modeling many social and biological phenomena [Palla *et al.* 2005, Spirin and Mirny 2003, Ravasz *et al.* 2002, Watts, Dodds and Newman 2002].

Not all modular networks have the same structure so it's important to introduce the idea of strong communities and weak communities. Consider a node i that belongs to a community M in a modular network, we define k_i^I as the internal degree of i, this is, the total number of connections from i to any other node that belongs to M. Also, we define k_i^O as the external degree of i, in other words, the total number of links between i and any other node that is outside M. With this consideration we define strong and weak communities, based on [Barabási and Pósfai 2016], as follows:

- Strong community: A module M is strong if each of its nodes has more links to other nodes in the same community than to nodes that belong to other communities. This is:

$$k_i^I > k_i^O, \text{For every } i \in M \tag{2.16}$$

- Weak community: A module M is weak if the total internal degree exceeds their total external degree. This is:

$$\sum_{i \in M} k_i^I > \sum_{i \in M} k_i^O \tag{2.17}$$

It also might be useful to establish the concept of *Very weak community*, not presented in [Barabási and Pósfai 2016], but that comes as a logical follow-up from the strong and weak community idea.

- Very weak community: A module M is very weak if the total external degree exceeds their total internal degree. This is:

$$\sum_{i \in M} k_i^I < \sum_{i \in M} k_i^O \tag{2.18}$$

In this sense, networks can have a stronger community structure if more of their communities are strong, or a weaker community structure if have more weak communities. The specific

criteria used to classify the structure of a network as strong or weak can vary, but for now, it's important just to have a basic understanding of these concepts.

A lot of research in modular network has been dedicated to develop algorithms to successfully detect communities [Hu *et al.* 2008, Duch and Arenas 2005, Schaub *et al.* 2017, Tasgin, Herdagdelen and Bingol 2007]. This is important, as it helps us capture an accurate description of real-world systems. There is also an interest in generating and producing modular networks [Lancichinetti, Fortunato and Radicchi 2008, Pasta and Zaidi 2016, Muscoloni and Cannistraci 2018]. This is especially relevant for the study of dynamic processes on networks, like epidemic processes.

2.4 Summary

In this chapter we covered the fundamental and basic concepts associated with the research project. In section 2.1, we talked about the SIS and SIR deterministic homogeneously mixed compartmental models. This is one of the most popular approaches to study epidemic modeling and consists in classifying individuals into groups according to their state related to the disease. For the SIS model, infected individuals recover but can get infected again so there are two compartments: Susceptible and Infected. However, for the SIR model infected individuals can not get infected again, so an additional compartment is added. Both models have two types of behavior depending on the value of the basic reproduction number $R_0 = \sigma/\gamma$.

The homogeneously mixed assumption is considering that all individuals can interact with each other, which is not realistic. One way to consider some spatial property is using complex networks as the structure where an infectious disease propagates. In section 2.2, we cover basic concepts and definitions related to complex networks. We also saw three different models, being the first the ER network, which is the most basic network model with the characteristic of being regular and random. Many real networks have properties that are not captured by the ER network, so the other two models the WS and BA model help include two important properties: small world characteristic and heterogeneity in degree distribution, respectively.

We this basic ideas established, we now can combine them to have epidemic models in complex networks. The next chapter 3 will be focused on covering this topic, together with a revision of stochastic process, which are fundamental to integrate this two models.

CHAPTER

3

EPIDEMICS IN COMPLEX NETWORKS AND STOCHASTIC PROCESSES

At the beginning of the last chapter, we presented the compartmental epidemic models in a fully mixed population. This is one of the most popular approaches to simulating epidemic disease, but one of its biggest downsides is the lack of a spatial dimension. In the real world, human systems are far from being fully mixed, so it's important to take into account some spatial property to have a more realistic model. As mentioned at the end of last chapter, one way to solve this issue is to use networks as the social structure in which a disease spreads. This approach has been increasing in popularity in the last few decades [Pastor-Satorras *et al.* 2015].

When we consider epidemics on networks the complexity of models increases and many times deterministic approaches are not viable. Because of this stochastic approaches have been fundamental to model this kind of systems. Before diving into some approaches on how to study epidemics in networks in section 3.2, we must present a brief review on stochastic processes in the first section of this chapter 3.1. Special attentions is dedicated to the "Memory-less" property of Markov processes as it will be a fundamental concept of the models developed in the proceeding chapters.

3.1 Revision of Stochastic Processes

The idea of stochasticity can be tracked to the origin of the word in old Greek "stokhastikos" which can be translated as "to guess", which reflects the idea of randomness and chance involved

in these kind of processes [Rinne 2008]. In its most general definition a stochastic process can be understood as any family or collection of random variables [Karlin and Taylor 1975]. Random variables are not variables in the traditional sense, they are functions of possible outcomes of a sample space to a measurable space [Karlin and Taylor 1975].

We consider a probability space $[\Omega, A, P]$ and a non-random index set T, called parameter space, which is a measurable. The probability space is defined by a sample space Ω, a $\sigma-$algebra A, and a probability measure P. A stochastic process, $\{X_t\}_{t\in T}$, will be a collection of random variables X_t that maps $\Omega \times T$ into a measurable space S (an equivalent notation of X_t is $X(t)$). The stochastic process has a state space which is the set of all possible values of the random variables X_t. A fundamental aspect that many time distinguishes one stochastic process from another are the probability distribution function associated with random variables involved.

With these concepts it is possible to elaborate simple classifications for stochastic processes. If the state space is countable we call the stochastic process a point process, and if the state space is $\mathbb{R}$ or any compact subset of $\mathbb{R}$ we call it a diffusion process. On the other hand, if the parameter space is discrete, we call it a stochastic series or a stochastic chain, and if it is non-countable we will deal with continuous processes. In the context of spreading and epidemic modeling in general t represents the time, and the characterizations of continuous or discrete parameter space implies a continuous or discrete time model. For this text, we will center on continuous parameter space.

3.1.1 Markov processes

Markov chains, for discrete times, and Markov processes, with continuous time, are within the most famous and widely used stochastic processes. They were proposed by Andrey Markov and have what is usually called the "memory-less property". There are many formulations of this property, but one of the most common is considering that given the value of X_t, the values of X_s for $s > t$, are independent of the values of X_u for $u < t$. This characteristic can be understood as independence of past values and is what is called the "memory-less property", as we only need the present state to know the probability of future outcomes [Rinne 2008]. This idea can be translated in the equation (3.1)

$$P(X_t = x | X_{t_1} = x_1, X_{t_2} = x_2, X_{t_3} = x_3, \ldots X_{t_n} = x_n) = P(X_t = x | X_{t_n} = x_n) \tag{3.1}$$

Where the times have the relation $t_1 < t_2 < t_3 < t_4 ... < t_n < t$.

3.1.2 Point processes and counting processes.

Outside these classification, we can also talk about two important types of stochastic processes: point processes and counting processes. A point process is characterized by a sequence of events occurring in continuous time, like birth, death, failure of a system, etc. In this kind of processes the random variables involved are related to the time when the events will happen. For example, a variable T_n could be the instance of occurrence of the n-th event of the system [Ross 2010]. Many times, we are interested not just on the time of occurrence but also in the total number of events that have happened, for this we can define the total number of events in the interval $[0,t]$ as the random variable N_t, that follows the characteristic described in (3.2).

$$N_t = n \quad if \quad T_n \leq t < T_{n+1} \tag{3.2}$$

If we consider a family of random variables $N = (N_t, 0 \leq t < \infty)$ we will end up with a counting process. Some characteristics of this kind processes follows:

- N is an increasing process such that $N_t \geq N_s$ if $t \geq s$
- $N_t \in 0,1,2,3...$
- Consider two times s and t such that $0 \leq s \leq t$. Then $\Delta N = N_t - N_s$ will be equal to the events occurred in the interval $[s,t]$.

There is a direct relation between N_t and T_n. For instance, if $N_t < n$, meaning that at time t there has been a number of occurrences less than n then $T_n > t$, which means that the n-th event occurs after time t. Similarly, if $N_t \geq n$ then $T_{n+1} > t$. From these ideas we can conclude that:

$$P(N_t < n) = P(T_n > t) \tag{3.3}$$

$$P(N_t \geq n) = P(T_{n+1} > t) \tag{3.4}$$

3.1.3 Poisson processes

The homogeneous Poisson process is a counting process widely used to model social and biological phenomena. A collection of random variables $\{N_t, t \geq 0\}$ will be a Poisson process if the following conditions are held [Ross 2010]:

1. $N_0 = 0$
2. For all $t_0 = 0 < t_1 < t_2 < t_3 < ... < t_n$ the increments $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, $...N_{t_n} - N_{t_{n-1}}$ are independent random variables. Which mean that the number of events in disjoint time intervals are independent.
3. The distribution of the number of events in given interval depends just on the length of the interval. This means that the probability distribution of $N_{t+h} - N_t$ will be the same regardless of the value of t. Also if we define a random variable $X_h := N_{t+h} - N_t$ and take $h \to 0$ then:
   - $P(X_h = 1) = \lambda h + o(h)$
   - $P(X_h \geq 2) = o(h)$

Where λ is said to be the rate of the Poisson process. The idea of a function f being $o(h)$ means that $\lim_{h\to 0} \frac{f(h)}{h} = 0$. The condition 3, which also called the stationary increment assumption, is a fundamental characteristic that make the Poisson process a Markov process. For a Poisson process the probability of an event happening in the future time does not depend on the time already passed, which is the memory-less property characteristic of Markov processes.

Now we will turn our attention to the probability distribution associated to the Poisson Process. Let us use the notation $P(\{X_t = m\}) = P_m(t)$. The probability of having no events in the time interval $t + h$ can be written as:

$$\begin{aligned} P_0(t+h) =& P\{N(t+h) = 0\} \\ P_0(t+h) =& P\{N(t = 0, N(t+h) - N(t) = 0\} \end{aligned} \tag{3.5}$$

From the condition 2 disjoint times are independent random variables, so we can write:

$$P_0(t+h) = P\{N(t=0\} P\{N(t+h) - N(t) = 0\} \tag{3.6}$$

If we call $P\{N(t=0\} = P_0(t)$, and consider that $P_0(t) = 1 - (\sum_{m=1}^{\infty} P_m(h))$, we can express $P_0(t+h)$ as:

$$P_0(t+h) = P_0(t)[1 - \sum_{m=1}^{\infty} P_m(h))] \tag{3.7}$$

From condition 3 if $h \to 0$ we know that $P_1(h) = \lambda h + o(h)$, and we can consider $\sum_{m=2}^{\infty} P_m(h)) = o(h)$ so:

$$\begin{aligned} \lim_{h\to 0} P_0(t+h) &= \lim_{h\to 0} P_0(t)[1 - \lambda h - o(h)] \\ \lim_{h\to 0} P_0(t+h) &= \lim_{h\to 0} [P_0(t) - P_0(t)\lambda h - P_0(t)o(h)] \\ \lim_{h\to 0} \frac{P_0(t+h) - P_0(t)}{h} &= -\lambda P_0(t) - \lim_{h\to 0} \frac{P_0(t)o(h)}{h} \end{aligned} \tag{3.8}$$

Note that $\lim_{h\to 0} P_0(t)o(h) = P_0(t) \lim_{h\to 0} o(h) = 0$, and the left side of the equation is the derivative of $P_0(t)$, so:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) \tag{3.9}$$

This differential equation has a well known solution:

$$P_0(t) = ce^{-\lambda t} \tag{3.10}$$

It is not relevant for this work, but it can be proven that $P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}; m = 0, 1, 2, 3..$, which is the Poisson distribution and is the origin of the name of this kind of process [Rinne 2008].

The probability of no occurrence in an interval time $[0, t]$, is equivalent to the probability of occurrence after a time t, this is:

$$P(T_1 > t) = P_0(t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} \tag{3.11}$$

Where T is time of the first occurrence. From this idea and using condition 3 we can take the point t_0 to be any moment, and then we can define the inter-arrival time Y between two

successive events, which also has a negative exponential probability distribution. This means that the cumulative probability distribution and the expected value of Y are:

$$\begin{aligned} P(Y < t) =& F_Y(t) = 1 - e^{-\lambda t} \\ E(Y) =& \frac{1}{\lambda} \end{aligned} \tag{3.12}$$

There also exist non-homogeneous Poisson processes where the independent increment property condition is relaxed. For this cases the rate depends on time $\lambda(t)$. We just need to be careful because other processes, like renewal process that will be presented in next subsection, also have rates that depend on time, but they are not necessarily non-homogeneous Poisson processes.

3.1.4 Renewal processes and memory-less property

One way we can understand renewal processes is as a generalization of a Poisson process. A renewal process is a counting process for which the inter-arrival times (even though they are still independent identically distributed random variables) are not constrain to an exponential distribution, as the Poisson process, and can be any arbitrary distribution [Mieghem 2006]. A classical renewal processes is a failure replacement processes, where events occur when elements of a system fail and are immediately replaced, like light bulbs or components of a circuit. We call $\{T\}$ a family of random variables that give the inter-event times for failure/replacement. For instance, T_1 will give the time τ_1 for the first failure/replacement, and T_2 will give a random time τ_2 for the second failure/replacement elapsed since τ_1. In figure 7 we can see a schematic representation of this process where $N(t)$ are the total number of failures by time t.

If the distribution of $\{T\}$ are negative exponential distribution then we recover the Poisson process, which is the only renewal process with the Markov property, this is, a memory-less property.

3.1.4.1 Markov assumption: How can we interpret the memory-less property?

Let us consider random variable X which gives the inter-event times of a renewal process. Then the probability of the event happening in a time grater then t can be written as $P(X > t)$. If we consider two instant of time τ_1 and τ_2 with $\tau_2 > \tau_1$, and we want to know the conditional

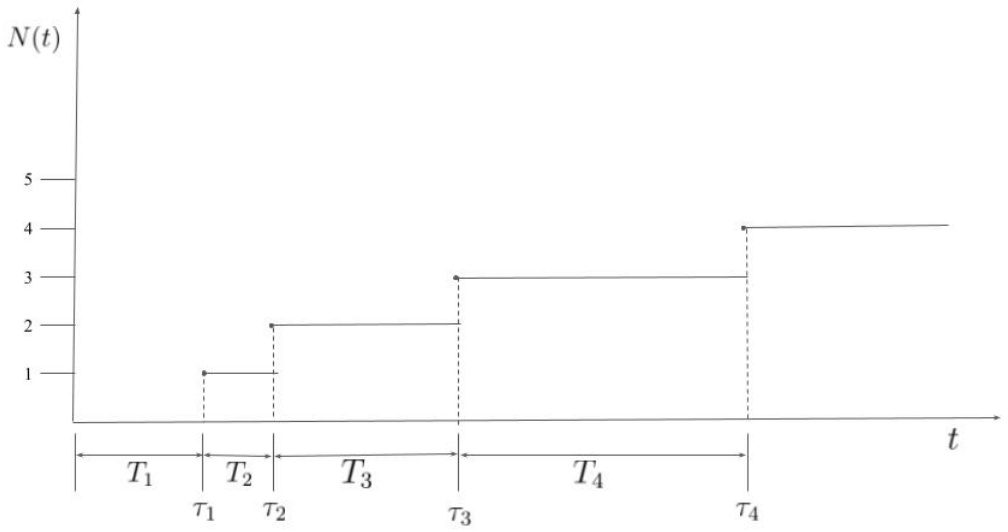

Figure 7 – Representation of a renewal process described by a $\{T\}$ a family of random.

Source: Elaborated by the author.

probability of the event taking place in a time grater then τ_2 given that the event will occur in a time after τ_1, we can write:

$$P(X > \tau_1 + \tau_2 | X > \tau_1) = \frac{P(X > \tau_1 + \tau_2 \cap X > t)}{P(X > \tau_1)} \tag{3.13}$$

But the $P(X > t + s \cap X > \tau_1) = P(X > \tau_1 + \tau_2)$, because if $\tau_2 > \tau_1$ the only cases where $X > \tau_1$ and $X > \tau_1 + \tau_2$ is when $X > \tau_1 + \tau_2$. Then we have:

$$P(X > \tau_1 + \tau_2 | X > \tau_1) = \frac{P(X > \tau_1 + \tau_2)}{P(X > \tau_1)} \tag{3.14}$$

The memory-less assumption is encoded in considering that $P(X > \tau_1 + \tau_2) = P(X > \tau_1)P(X > \tau_2)$. In this case, equation (3.14) will be:

$$\begin{aligned} P(X > \tau_1 + \tau_2 | X > \tau_1) &= \frac{P(X > \tau_1 + \tau_2)}{P(X > \tau_2)} \\ P(X > \tau_1 + \tau_2 | X > \tau_1) &= \frac{P(X > \tau_1)P(X > \tau_2)}{P(X > \tau_1)} \end{aligned} \tag{3.15}$$

$$P(X > \tau_1 + \tau_2 | X > \tau_1) = P(X > \tau_2) \tag{3.16}$$

Equation (3.16) means that the knowledge of the event nos happening until a time τ_1, does not influence the probability of waiting an additional τ_2 time. This is the memory-less property for renewal processes. If $P(X > \tau_1 + \tau_2) = P(X > \tau_1)P(X > \tau_2)$ is not true then the process will be non-Markovian and $P(X > \tau_1 + \tau_2 | X > \tau_1)$ will depend on τ_2 and τ_1.

The function $P(X > \tau_1)$ is the complementary cumulative distribution function, also called Reliability function, and can be computed as $P(X > \tau_1) = 1 - \int_o^{\tau_1} P(X = t)dt = F_X^c(\tau_1)$. So the condition of a renewal process having the Markov property is:

$$F_X^c(\tau_1 + \tau_2) = F_X^c(\tau_1)F_X^c(\tau_2) \tag{3.17}$$

The only function that satisfies this property is the exponential function [Mieghem 2006]. So any other renewal process with an inter-event time distribution different from the exponential distribution will be non-Markovian.

Now that we establish some fundamental aspects of stochastic processes, let us discuss the combination of epidemic modeling and complex networks.

3.2 Epidemic Processes in Complex Networks

At the beginning of the last section 2.1 we highlighted a key limitation of homogeneously mixed models, namely their failure to consider contact structures between individuals. At the end of the 1990s the study of complex networks really took over with. As a result, novel epidemic models emerged that utilized networks to incorporate the structural and contact properties of real systems. This approach gained immense popularity in the study of epidemic processes, and new methodologies continue to emerge constantly [Pastor-Satorras *et al.* 2015].

The general framework employed in studying epidemics in complex networks involves considering a network comprising N individuals, represented by an adjacency matrix A, and an epidemic process characterized by c compartments. At any given time t, the system's state is described by N random variables denoted as $X_i(t)$, representing the state of node i and assuming values from a state space corresponding to the disease compartments. For example, in the SIS model, we can assign $X_i(t) = 1$ to indicate that node i is infected, while $X_i(t) = 0$ represents node i being susceptible. For SIR model, we would just need a new state, and with the same logic any number of compartments could be added. A key aspect of epidemic modeling involves

utilizing Markov chain theory to analyze this framework. Under this assumption, transitions between compartments occur according to a Poisson process [Pastor-Satorras *et al.* 2015] with constant rates. In line with the notation used in the previous chapter, we denote the transmission rate as σ and the recovery rate as γ.

Although the approach described above generates an exact model, it is very limited as few exact results have been obtained. When considering additional compartments, non-Poisson transition processes, or complex interactions such as the simultaneous occurrence of rumor spreading and an epidemic spreading, the exact approach becomes impractical. Consequently, two alternative approaches have gained considerable popularity in the study of epidemics on networks.

The first approach involves mean-field approximations, which employ reductionist assumptions to derive simplified models. These models are typically analyzed using deterministic equations, aiming to extract relevant information about the overall system. The mean-field approach provides a more analytical approach to studying these systems, despite still relying on numerical methods. The second approach is based on stochastic simulations, utilizing numerical techniques such as the Gillespie algorithm to simulate the epidemic process. Stochastic simulations serve as valuable tools, particularly when exact analytical solutions are lacking. In many cases, both approaches complement each other because, although both of them need numerical methods, mean-field approach is a more analytical way to study this systems, and the simulations base research occupies almost an experimental role.

These approaches are used with the main goal investigating how network structure may affect parameters like epidemic spreading velocity and epidemic sizes. Also, for the SIS and SIR models, it has been observed that as time approaches infinity ($t \to \infty$), the order parameter (in general the epidemic size) exhibits a second-order phase transition depending on the values of the transmission and recovery rates. Typically, an effective transmission rate is utilized, denoted as λ, which is defined as the ratio of the transmission rate (σ) to the recovery rate (γ) in the Markov approach. A critical value of the effective transmission rate λ_c serves as a boundary separating two regions: an active region and an absorbing region. When the effective transmission rate surpasses λ_c, a global pandemic unfolds. In the case of the SIS model, the active region gives rise to an endemic steady state, where infected nodes persist indefinitely. For the SIR model, this implies an epidemic that affects a significant portion of the population. A schematic representation of this behavior can be found in Figure 8

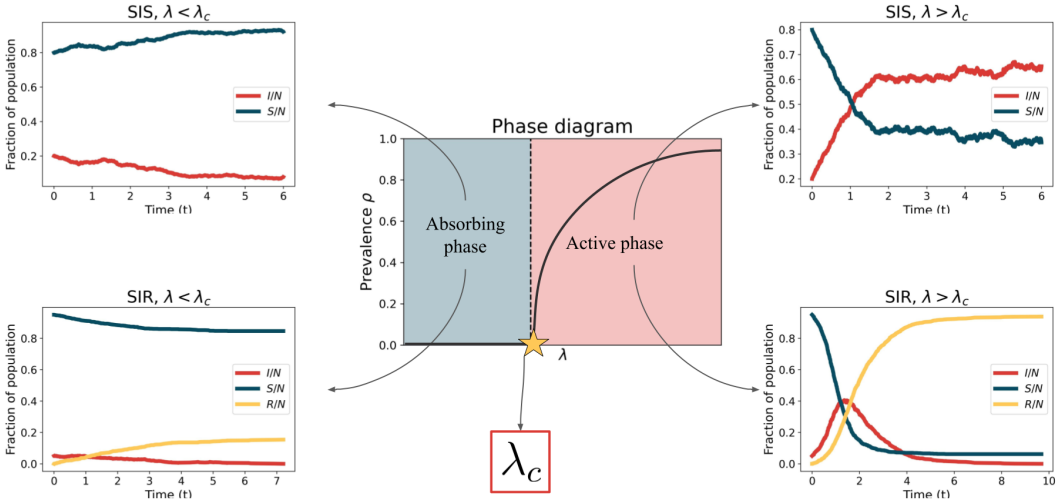

Figure 8 – Representation of the phase transition observed in epidemic processes in complex networks. At the center, an image of the phase diagram is displayed, where the green region ($\lambda < \lambda_c$) represents the absorbing phase and the orange region the active phase ($\lambda > \lambda_c$). The plots around the phase diagram represent the dynamical behavior of the compartments for the SIS model (top two images) and SIR model (bottom images).

In the subsequent part of this section we will explore some of the most relevant results in the classical Markov approach. Our specific focus will be on understanding the various techniques employed and the key results that arise from incorporating network complexity into epidemic models. This discussion will serve as a foundation for the upcoming chapter, where we will delve into non-Markovian models, building upon the concepts introduced in this section.

3.2.1 Mean-field approximation

3.2.1.1 Heterogeneous mean-field (HMF) approximation or Degree Based mean-field approach

The first work that used complex networks as a way to describe contact between individuals in an epidemic model was made by Romualdo Pastor-Satorras and Alessandro Vespignani [Pastor-Satorras and Vespignani 2001] in 2001. In this paper they developed a mean-field approximation to study a SIS process in networks constructed with the Watts Strogatz and Barabási Albert models. Nowadays we call this approximation Heterogeneous mean-field approach (HMF), and has been widely used for different researches [Pastor-Satorras and Vespignani 2001, Morita 2016]

This approach is based in the assumption that all nodes with the same degree are

statistically equivalent [Wang *et al.* 2016]. For this reason, this theoretical framework is also called degree-based mean-field approach. It is very popular because of its applicability to a diverse types of dynamical processes [Barrat, Barthélemy and Vespignani 2008]. For a SIS process the primary magnitude used is $\rho_k^I(t)$, which is the fraction of nodes with degree k that are infected at a specific time t. The evolution of the system is given by equation (3.18):

$$\frac{d\rho_k^I(t)}{dt} = -\gamma\rho_k^I(t) + \sigma[1-\rho_k^I(t)]\Theta_k(t) \tag{3.18}$$

The parameter γ is the recovery rate, similar to the one described in section 2.1 , and σ is the transmission rate for each contact between a susceptible and an infected individual. Also, we define Θ_k as the probability that a node with degree k is linked to an infected node, and is given by $\Theta_k = \sum_k' P(k'|k)\rho_{k'}^I(t)$, where $P(k'|k)$ is the probability that a node with degree k points to a node with degree k'. For uncorrelated networks $P(k'|k) = \frac{k'P(k')}{\langle k\rangle}$. In this case Θ_k does not depend on k anymore so we can just write Θ.

Equation (3.18) basically says that the rate of change of the fraction of infected individuals with the degree k depends on the fraction of infected nodes with degree k that recover, which is expressed in the first term of the right side of the equation, and on the fraction of new infected nodes with degree k, which is expressed in the second term of the right side of the equation. A trick to find the threshold of for this approach is multiplying both sides by $kP(k)/\langle k\rangle$, and then summing over all values of k. By doing this we obtain:

$$\sum_k \frac{kP(k)}{k}\frac{d\rho_k^I(t)}{dt} = -\gamma\sum_k \frac{kP(k)}{k}\rho_k^I(t) + \sigma\Theta_k(t)\sum_k \frac{kP(k)}{k}[1-\rho_k^I(t)] \tag{3.19}$$

Simplifying the equation and noting that the left side is the derivative of Θ we can express:

$$\frac{d\Theta}{dt} = -\gamma\Theta + \frac{\sigma\Theta}{\langle k\rangle}\left(\langle k^2\rangle - \sum_k k^2\rho_k^I kP(k)\right) \tag{3.20}$$

If we consider ρ_k^I goes to zero asymptotically then $\frac{d\Theta}{dt} \approx \Theta\left(\sigma\frac{\langle k\rangle}{\langle k^2\rangle} - \gamma\right)$. With this linearization we can establish a threshold:

$$\lambda_c^{HMF} = \frac{\sigma}{\gamma} = \frac{\langle k\rangle}{\langle k^2\rangle} \tag{3.21}$$

In the case of Scale-Free networks with an exponent $\nu \leq 3$, like the Barabási Albert model described in 2.3.2, there is a vanishing epidemic threshold as $N \to \infty$ as the second moment of the probability distributions $\langle k^2 \rangle$ diverges.

The approach for the SIR model is nearly identical to the one described above for the SIS model. However, in this case, we consider three different partial densities, each corresponding to a specific compartment. These partial densities can be denoted as ρ_k^I, ρ_k^R, and ρ_k^S, representing the infected, recovered, and susceptible compartments, respectively. While we will not provide a detailed derivation as we did for the SIS model, as mean-field approximations are not the primary focus of this project, a comprehensive development of the first HMF approach for the SIR model can be found the work of Moreno Y. and colleagues [Moreno, Pastor-Satorras and Vespignani 2001]. For the purpose of this discussion, we will only highlight the epidemic threshold which coincides with the one found for the SIS model: $\lambda_c^{HMF} = \frac{\sigma}{\gamma} = \frac{\langle k \rangle}{\langle k^2 \rangle}$. However, an improvement can be done for the SIR model in which it is possible to take into account that a vertex cannot propagate the disease to the neighbor who originally infected it. With this consideration, a correction is made for he probability distribution and a new critical value is found, which is:

$$\lambda_c^{HMF} = \frac{\sigma}{\gamma} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{3.22}$$

Again, for scale free networks with $\nu \leq 3$ this threshold has a vanishing value, and for $\nu > 3$ a finite value.

3.2.1.2 Quench mean-field (QMF) approximation or Individual Based mean-field approach

In 2008, another mean-field approach was proposed by Chakrabarti and collaborators [Chakrabarti *et al.* 2008]. In this approach the reduction made was not focused on the structure of hte network, as in the case of the HMF approach, but rather in the dynamical system. Specifically, this method focuses on the Adjacency matrix A of the network, and is commonly called Quench mean-field approach (QMF). In this case, it is used the variable $\rho_i(t)$, which is the probability that node i is infected at a time t. Using the same recovery and infection rate mentioned before (γ and σ) the evolution equation of this variable is given by (3.23) [Wang *et al.* 2016].

$$\frac{d\rho_i(t)}{dt} = -\gamma\rho_i(t) + \sigma[1 - \rho_i(t)] \sum_{j=1}^{N} A_{ij}\rho_j(t) \tag{3.23}$$

The first term of the right side of equation (3.23) is the probability of node i recovering from the disease, and the second term is the probability of the node getting infected at time t. Note that the second term vanishes for the matrix values which are equal to zero, which is obvious as nodes are in risk of infection just by their neighbors.

Using this formulation we can make a stability analysis of the system. This can be done by finding the signal of the Jacobian's leading eigenvalue evaluated at the fixed pint. The Jacobian matrix elements are given by:

$$J_{ij} = \frac{\partial \dot{\rho}_i}{\partial \rho_j} = A_{ij}\sigma(1-\rho_i) - \delta_{ij}\left(\gamma + \sum_k A_{ik}\sigma\rho_k\right) \tag{3.24}$$

We can make a linearization for $t \to 0$ where $\rho_i \to zero$ and obtain a matrix equation of the form:

$$J = \sigma A - \gamma \mathbf{I} \tag{3.25}$$

Where $\mathbf{I}$ is the identity matrix with size N. From this equation the following threshold is obtained:

$$\lambda_c^{QMF} = \frac{\sigma}{\gamma} = \frac{1}{\Lambda_A} \tag{3.26}$$

Where Λ_A is the largest eigenvalue of the adjacency matrix A. For heterogeneous networks with scale-free property with $\nu < 2.5$ $\lambda_c^{QMF} \approx \langle k \rangle / \langle k^2 \rangle$ and for $\nu > 2.5$ then $\lambda_c^{QMF} \approx 1/\sqrt{k_{max}}$, which in both cases indicates a vanishing threshold in the thermodynamic limit [Wang *et al.* 2016].

One important aspect of the QMF approach is that it diverges from the HMF result for scale-free networks with $\nu > 3$. For this cases the QMF still holds a vanishing threshold for $N \to \infty$, which is different from the results of HMF which finds a finite value of λ_c.

For the SIR model the QMF approach results in the same threshold value of the SIS model $\lambda_c^{QMF} = \frac{1}{\Lambda_A}$ [Youssef and Scoglio 2011]. Again we will not detail the derivation of this threshold value.

3.2.2 Simulation focused approaches

As shown above, the HMF and QMF approach lead to different results for the threshold values. While these approaches are undeniably valuable, it is essential to acknowledge that they rely on approximations which can potentially overlook crucial details in more complex scenarios. For this reason, the use of stochastic computational simulations has been of great use [Pastor-Satorras *et al.* 2015].

One way to tackle this problem is using a discrete time Monte-Carlo simulation. In this approach, random numbers are generated to determine the events that will occur in the next time step. Although this method is straightforward to implement, continuous time models are more commonly used.A popular approaches for continuous time is a Monte Carlo method called Gillespie Algorithm (GA), which was originally proposed to perform a stochastic simulation of chemical reaction systems with known reaction rates [Gillespie 1976]. The next subsection briefly describes this algorithm and its application to epidemic processes following the framework of Marian Boguñá and others in [Masuda and Rocha 2018].

3.2.2.1 Gillespie algorithm

Consider a collection of N statistically independent stochastic processes, each one with its own rate λ_i. Ir operates by generating a sequence of events that specify, at each step of the simulation, the time τ until the next event and which event i will occur. For the Markovian approach, where all processes involved are Poisson processes, this can be accomplished using two distributions: $\Pi(i)$ for selecting the event and $\psi(\tau)$ for determining the timing of the next event. These distributions are given by $\Pi(i) = \frac{\lambda_i}{N\langle\lambda\rangle}$ and $\psi(\tau) = N\langle\lambda\rangle e^{-N\langle\lambda\rangle\tau}$, where $\langle\lambda\rangle = \sum_{j=1}^{N}\lambda_j$.

This approach can be applied to epidemic process where transitions between compartments treated as the stochastic independent processes. In such case, we would define $\langle\lambda\rangle = \sigma M_{SI}(t) + \gamma N_I(t)$, where M_{SI} are the number of links between susceptible and infected individuals, and N_I the number of infected individuals. Using this average, we generate an exponential time τ using $\psi(\tau)$, and then decide whether to perform a new infection with probability $\frac{\sigma M_{SI}}{\langle\lambda\rangle}$ or a recovery with probability $\frac{\gamma N_I}{\langle\lambda\rangle}$. If an infection is chosen, we select one of the existing S-I links and randomly choose one of its ends to initiate the infection. On the other hand, if a recovery is chosen, we randomly select an infected node for recovery.

3.2.2.2 Finding the threshold value of the system

Considerable computational efforts have been dedicated to studying the threshold behavior of epidemics on networks. As demonstrated, mean-field approaches can provide approximations of critical rate value. However, different methods can yield to varying results, and for more complex models mean-field approximations turned out to be limited.

The simulation based approach to study threshold of the system focuses on calculating the first and second moments of the order parameter of the system. For s SIS-like model, the order parameter is given by the density of infected nodes at the steady state. The most basic approach is calculated using an average of the fraction of infected nodes before reaching the absorbing state or after reaching the endemic state. These averages are calculated after many simulations on different networks. However, this basic method incurs a high computational cost. To address this limitation, the quasi-stationary state analysis was developed [Ferreira, Castellano and Pastor-Satorras 2012]. In this method, simulations are constrained to the active state by replacing the system when the epidemic reaches the absorbing state. Essentially, during the simulation there is a store of previous active states, and when the absorbing state $(I = 0)$ is reached, the system is replaced by one of the previous active state with a probability proportional to the time the system was in that given configuration. With this technique the calculations of the order parameter are not just more efficient but also more reliable for Markovian epidemics.

Once we have a method to obtain the first and second moments of the order parameter, in the case of SIS-like models, the standard procedure involves studying the susceptibility of the system [Ferreira, Castellano and Pastor-Satorras 2012]. The susceptibility is defined as follows:

$$\chi = N\frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle} \tag{3.27}$$

The specific magnitude of the susceptibility is not the primary focus; rather, our aim is to identify the transition rate at which this measure reaches its maximum value, as it corresponds to the critical threshold for sufficiently large systems [Ferreira, Castellano and Pastor-Satorras 2012].

For SIR-like processes the order parameter is the final size of the epidemic, this is, the density of recovered nodes after the end of the epidemic [Pastor-Satorras *et al.* 2015]. For this model, the calculation of the first and second moment are a little bit easier, and some times more refined methods like the quasi-stationary state are not necessary [Shu *et al.* 2015].

In this case, it has been found that susceptibility is not a most reliable measure for

determining the critical transition rate. Typically, it provides higher transition rates compared to those obtained from mean-field approximations. Instead, the vulnerability measure, which is commonly used in identifying critical points in the equilibrium phase of magnetic systems, has been demonstrated to be a more accurate measure for the SIR model [Shu *et al.* 2015]. The vulnerability is defined as follows:

$$\Delta = \frac{\sqrt{\langle \rho^2 \rangle - \langle \rho \rangle^2}}{\langle \rho \rangle} \tag{3.28}$$

Again, just like in the case of the susceptibility, in the critical transition rate λ_c the vulnerability has a maximum value.

3.3 Summary

In this chapter we cover two main topics: stochastic processes and epidemics on networks. The goal to cover this topics together in this chapter was because stochastic processes are a fundamental tool needed to understand the models and methods used in epidemics on networks. In section 3.1, we talked about Markov processes, Point processes, Counting processes, Poisson processes, and Renewal processes, which are commonly used in constructing epidemic models. We take spacial attention to the section 3.1.4.1, where we present what is the Markov assumption for renewal processes, which is a concept that will be fundamental for the rest of the text.

In the rest of the chapter, in section 3.2 we present the main approaches, results and concepts needed to study epidemics on networks. One of the most important results discussed is the critical behavior of epidemics on networks. Models like SIS and SIR epidemic processes in networks present critical behavior, where using a control parameter λ (in general associated with the recovery and infection rate) we see two phase: an absorbing one where the disease does not propagate in the system, and an active one where an epidemic outbreak is observed. This has been studied with mean-field approaches, presented in section 3.2.1, and simulation approaches 3.2.2.

With this chapter, we have constructed the bases of this research field. Now we will dive in the main topic of this text: Non-Markovian epidemic models. In chapters 4, we will present the most popular approach to describe non-Markovian epidemic processes and the most relevant literature in this area.

CHAPTER

4

NON MARKOVIAN EPIDEMIC PROCESSES IN COMPLEX NETWORKS

When studying complex phenomena, incorporating new features or details into our models can lead to more realistic conclusions. However, this comes at the cost of increasing the difficulty of studying such models. It is essential, therefore, to always question why we are including new characteristics that make our models more intricate. In section 4.1, we address this question for non-Markovian epidemic processes. We will establish the importance of working with this kinds of models and why its a feature that cannot be ignore.

Next, we will present the standard approach to non-Markovian epidemic processes in 4.2. This step is fundamental as the term “non-Markovian” can encompass a wide range of approaches, and is fundamental to understand where is the dependence of past states for this systems.

Finally, in section 4.3, we will cover the main literature developed in this field, first in a historical point of views, and then focus on different approaches and highlighting the most important results.

4.1 Why bother with non-markovianity in epidemic processes?

A lot of research in epidemic modeling is centered on Markovian processes, mainly because this simplification allowed researchers to develop very useful and relatively straightforward models. Unfortunately, many times human activities are not well characterized by Markovian processes and are better described by heavy-tailed inter-event time distribution [Gonzalez, Hidalgo and Barabasi 2008, Barabási 2005, Brockmann, Hufnagel and Geisel 2006]. Additionally, diseases frequently exhibit incubation periods with strongly non-exponential tails [Blythe and Anderson 1988]. As infectious diseases evolve and interact within the host, it becomes crucial to adopt approaches that account for aging in the infection or recovery process, a characteristic commonly observed in real infectoues diseases.

Adopting a memory-less model, while convenient, can be a big limitation in comprehending real systems. It has been shown that relevant biases arise when using a Markovian model when the actual phenomena are better described by a non-Markovian approach [Kenah 2010].

Moreover, non-Markovian processes are not exclusive to epidemics; they have been identified in various social and biological modeling domains. For instance, models with non-Markovian characteristics have been studied for gen self-regulation and cell processes [Yin, Liu and Wen 2021, Ebadi *et al.* 2016], social systems and voting models [Chen *et al.* 2020, Peralta, Khalil and Toral 2020], heart and Parkinson-related research [Yulmetyev *et al.* 2005, Panischev, Yulmetyev and Demin 2005], and even queuing theory [Kella and Stadje 2006]. This wealth of evidence across diverse fields strongly suggests that epidemics are unlikely to be an exception, reinforcing the need to incorporate non-Markovian elements into epidemic models.

4.2 The standard non-Markovian approach for epidemic in complex networks: the application of renewal processes

In the previous section 3.2 we explored the Markovian approach to epidemic modeling, where infection and recovery are treated as independent Poisson processes. The theorem on Poisson processes, as described in the book "Performance Analysis of Communications Networks

and Systems" [Mieghem 2006] on page 146, establishes that if exactly one event of a Poisson process occurs in an interval [0, t], then the time of occurrence of this event is uniformly distributed. In essence, this theorem indicates that Poisson processes are the “most” random process [Mieghem and Liu 2019]. A generalization of this formulation involves the use of renewal theory, which is briefly described in section 3.1.4. In this framework, events are still independent of each other, but the interarrival time of events can follow any distribution, including distributions different from the exponential distribution [Mieghem and Liu 2019]. As shown in section 3.1.4.1, only the exponential distribution exhibits the memoryless property, making any renewal process with a general distribution a non-Markovian process. While not all non-Markovian approaches for modeling epidemics on networks explicitly employ renewal theory, most of them can be reduced to this approach as they treat the time of recovery and infection as independent processes with general distributions.

Until now, we have have mentioned that the processes involved in non-Markovian epidemics have general distributions, but we have not clarify what exactly does this mean. To address this, let us introduce the notions of infection age and activation age, applicable to both SIS and SIR models. The infection age of a node refers to the time elapsed since the node was initially infected. Similarly, the activation age represents the elapsed time since an active link was established, i.e., when one of two susceptible nodes connected gets infected. In the context of a Markovian epidemic process, these age concepts hold little relevance since the probabilities of recovery or disease transmission remain independent of these ages. However, in the case of non-Markovian processes, the time elapsed since a node’s infection or since the establishment of an active link generally has an impact on the probability of future events.

To gain a deeper understanding of these concepts, let us introduce two random variables associated with the renewal processes: T_r and T_{inf}, which describe the interarrival times of recovery and infection. The random variable T_r will give us a random time τ_r when an infected node will spontaneously recovers in the infinitesimal time interval $(\tau_r,\ \tau_r + dt)$. On the other hand, T_{inf} provides a random time τ_{inf} at which an active link will generate a new infected node in the infinitesimal time interval $(\tau_{inf},\ \tau_{inf} + dt)$. These variables follow probability distributions $\psi_{rec}(\tau_r)$ and $\psi_{inf}(\tau_{in})$. To recover the Markovian assumption we just have to consider this distribution as negative exponential distributions with their respective constant rates. In the general case, $\psi_{rec}(\tau_r)$ and $\psi_{inf}(\tau_{in})$ are any probability distributions, which generate a dependence on the infection or activation age. We can interpret this as an “aging” of the probability of transition

between states.

It is essential to emphasize that in this approach, the non-Markovian characteristic does not rely explicitly on the complete past states of the entire system (i.e., the history of infections and recoveries of all nodes). Rather, the dependency on past states occurs within each individual recovery and infection process. As time elapses, the disease ages within each individual, and this aging process significantly impacts the probability of future events for that particular individual.

Having now established the standard approach to non-Markovian epidemic processes in complex networks, we will proceed with a brief review of the primary works in this field.

4.3 Literature review of non-Markovian epidemic processes: approaches and main results

From a historical perspective, the investigation of memory effects in propagation phenomena has been a topic of interest for several decades. The first models exploring this concept emerged in the late 20th century and early 2000s [Yang 1972, Yulmetyev *et al.* 2004]. Notably, Vasquez and colleagues conducted pioneering research on the exchange of emails among individuals within a university, employing fat-tailed distributions to describe the probability distribution of inter-event times [Vazquez *et al.* 2007]. The presence of non-Poisson inter-event times revealed longer prevalence decay times compared to those obtained with the standard Poisson process.

As the limitations of the Markovian approximation in understanding propagation dynamics became apparent in the first decade of the new century, there was a significant surge in research on non-Markovian epidemics applied to network models, particularly during the 2010s. Among these early works, a notable article by P. Van Mieghem and R. van de Bovenkamp stood out. They modeled and simulated a non-Markovian SIS propagation with a Weibull infection process and discovered that the characteristic threshold significantly changed with the shape parameter of the distribution [Mieghem and Bovenkamp 2013]. Specifically, when the shape parameter α increased, the threshold of the system also increased, and the critical transition rate tended to zero as α decreased. In the same year, these two researchers, along with R. Cator, developed an N-intertwined mean-field approximation (NIMFA), which is a variation of the QMF approach, for the SIS process with Weibull infection [Cator, Bovenkamp and Mieghem 2013]. This mean-field approximation provided a first-order estimation of the threshold value for

large networks. This threshold is:

$$\lambda_c^{NIMFA} = \frac{1}{\Gamma(1+\frac{1}{\alpha})[\Gamma(\alpha+1)]^{1/\alpha}} \frac{1}{\Lambda_A^{1/\alpha}} \tag{4.1}$$

Where Λ_A is the largest eigenvalue of the adjacency matrix.

Another significant contribution was made by M. Boguñá and colleagues in 2014 [Boguñá *et al.* 2014]. This noteworthy work introduced a simple method to simulate non-Markovian processes on networks using a generalization of the Gillespie Algorithm. They also highlighted the dramatic effect of memory in SIS propagation not just in the threshold, but also in the size of the epidemic which increases with the shape parameter when using a Weibull transmission process.

From this point forward, it was clear that the non-Markovian process was not just a correction of the traditional Markov assumption, but needed to be fully understood. For the rest of this subsection, we will leave this historical aspect and focus on the approaches and main results for the research developed in the last 10 years.

4.3.1 Analytical approaches and mean-field approximations

The analytical approaches for studying Markov processes are already quite limited and it is natural that the scope of non-Markovian models is even more constrained. Consequently, some of the analytically solvable models have been focused on simpler contagion dynamics, like the SI model. Specifically, I.Kiss and colleagues worked on a solvable approach to the SI model in infinite systems, considering general inter-event time distributions. They discovered that the late-time dynamics exhibit a slower convergence towards a fully infected state, contrasting the exponential decay observed in Poisson-like processes [Jo *et al.* 2014].

A pairwise model for the non-Markovian epidemic also was proposed with an analytical threshold-like quantity and its relation with the final epidemic size [Kiss, Röst and Vizi 2015]. Furthermore, research that shows functional limit theorems for non-Markovian epidemic models also was developed [Pang and Pardoux 2020].

Similarly to the traditional Markov processes, mean-field approximations serve as a crucial tool for understanding the dynamics of epidemics with memory. As previously mentioned, the work by Piet Van Mieghem, Bovenkamp, and E. Cator **??** explored a generalized SIS model

with Weibull infection employing a NIMFA approach to determine a threshold value, described in equation (4.1). Another significant study by N. Sherborne, J. Miller, and colleagues demonstrated the equivalence between an edge-based compartmental model and message-passing models for general non-Markovian infection and recovery in a SIR model on regular networks [Sherborne *et al.* 2016]

More recently, L. Qiang and P. Van Mieghem made two important contributions regarding mean-field approaches and threshold values of the SIS non-Markovian process. The first one in 2018, proved the existence of an upper bound for the epidemic threshold of a SIS process with Weibull infection [Liu and Mieghem 2018], being this bound:

$$\lambda_c < \lambda_c^{UB} = \frac{1}{\Lambda_A + 1} \tag{4.2}$$

Subsequently a NIMFA mean field approach to find explicit threshold for SIS epidemics with Weibull and Gamma infection where found [Mieghem and Liu 2019]. For the Weibull infection the critical value was:

$$\lambda_c = \frac{1}{\Gamma(1+\frac{1}{\alpha})\Phi^{-1}(\frac{1}{1+\Lambda_A;\alpha})} \tag{4.3}$$

Where Φ is the generating function of the random variable with Weibull probability distribution. Through their investigation, they proved that by considering the first term of the Lagrange series, the threshold previously identified in [Cator, Bovenkamp and Mieghem 2013] and referenced in (4.1) serves as a reliable approximation, particularly for large spectral radii Λ_A and especially when $\alpha < 1$.

Mean field approximations also have been used to study the transient state and steady state of the non-Markovian SIS model [Feng *et al.* 2019].

4.3.2 Simulations and numerical approaches

In response to the analytical limitations, many researchers have directed their efforts towards simulation-centered research. There have been two main approaches on how to accurately simulate non-Markovian processes: non-Markovian Gillespie algorithm, and an event-driven simulation. The first one, is a modification of a Gillespie algorithm, originally proposed for stochastic simulation of chemical reactions [Gillespie 1976], and widely used in epidemic

modeling. This approach has been successfully used to simulate SIS, SIR, and even SEIS propagation models [Boguñá *et al.* 2014, Tomovski, Basnarkov and Abazi 2021, Kiss, Miller and Simon 2017, Masuda and Rocha 2018].

The second popular simulation strategy, the event driven approach, has the advantage of being very efficient and versatile to different distributions and transmission rules [Kiss, Miller and Simon 2017]. Event-driven can accurately produce stochastic epidemic SIR propagation models [Cuypere *et al.* 2013, Kiss, Miller and Simon 2017], and other algorithms like rejection based event-driven have been proposed to further improve these methods [Grossmann, Bortolussi and Wolf 2020].

Additionally, data analysis has also been a fundamental branch of non-Markovian research, especially to validate the analytical and numerical results. The increase of available data due to the COVID-19 pandemic has contributed to this numerical analysis, mainly studying the correspondence of existing models and real data [Lauro *et al.* 2022, Lu *et al.* 2022]. Other epidemics, like the ones caused by food-to-mouth diseases or Zika viruses, also have been studied under the non-Markovian lens [Lauro *et al.* 2022, Ferdousi *et al.* 2019].

4.3.3 Comparison between Markov and non-Markov processes

As there is a better knowledge of epidemics with Markov assumption and relatively small literature on non-Markovianity, comparisons between these two approaches turned out to be a necessity. Some of the most interesting results in epidemic processes with memory have emerged from research that tries to connect Markovian processes with non-Markovian ones. We can highlight an article by M.Boguñá and colleagues who introduced a formalism to reduce non-Markovian processes to Markovian processes and proved that all non-Markovian aspects can be captured in a single parameter, the effective infection rate [Starnini, Gleeson and ná 2017].

Another relevant work already mentioned in 4.3.1, develops a mean-field theory for an SIS model and establishes an equivalence between non-Markovian and Markovian processes that depends on a specific edge activation mechanism [Feng *et al.* 2019]. There are also more particular researches like comparing the endemic behavior of a SEIS non-Markovian dynamic with the known SIS Markovian [Tomovski, Basnarkov and Abazi 2021], and comparisons in sparse network based on a regular network [Kuga and Tanimoto 2022].

4.3.4 Future problems in the field

As shown from the cited research, the interest in non-Markovian epidemic processes has been increasing over time. Nevertheless, several questions remain unanswered. Exploring non-Markovian epidemics in various network models, like modular networks, dynamic networks and multilayer networks is still an open challenge. Also, more complex disease models, such as SIRS or SEIR models, are an ongoing pursuit.

An example of such research can be seen in the works conducted by L. Han and colleagues in 2023. They proposed a general formalism to investigate non-Markovian dynamics in non-Markovian temporal networks [Han *et al.* 2023]. This study holds great significance as it recognizes that the memory effects are not limited to the infection or recovery process alone, but are also characteristics of human interactions. By incorporating these additional factors, a more comprehensive understanding of non-Markovian epidemics can be attained.

4.4 Summary

We started this chapter by answering two fundamental questions: why study non-Markovian epidemic models? What is the standard approach to this kind of model? For the first question, we mentioned some evidence that sustains that in many cases human activity and disease incubation times are not well described by Markovian models. Moreover, considering the non-Markovian effect has been proven to have a drastic influence on the resulting epidemic, proving that non-Markovianity is not just a simple correction of the Markovian case. For the second question, we mentioned that the most popular approach for non-Markovian epidemics is considering infection, or recovery, as renewal processes with general distributions. With this lens, the individual suffers and aging in which the probability of infecting or recovering might be influenced by the time an individual has been in a certain state (recovered or infected). The dependence on past states, characteristic of non-Markovian processes, is encoded in this aging process.

In the rest of the chapter, we mentioned some of the most important literature in the field. Spatial attention is dedicated to the works of P. Van Mieghem and collaborators and Marian Boguñá and colleagues who have made the most influential research articles on the topic. Now that we presented why non-Markovian epidemics are relevant, and what is the standard approach to study them, we will present some of the most relevant results of this research project in chapter

5 and chapter 6

CHAPTER

5

NON-MARKOVIAN EPIDEMIC PROCESS WITH WEIBULL INFECTION IN DIFFERENT NETWORK TOPOLOGIES

In section 4.2 of the last chapter we described the standard approach for non-Markovian epidemic processes in complex networks. For the first section of this chapter 5.1 we will discuss a specific case of this approach, the non-Markovian epidemic process with Weibull infection. We present the interpretation of the aging process in this kind of non-Markovian epidemic. After this, in section 5.2, we cover the simulation methods used in this work for both generating the networks and executing the stochastic simulations.

Section 5.3 shows our results on the SIS non-Markovian epidemic process with Weibull infection in three different network models: an ER model, which generates random regular networks, a BA model, which generates scale-free networks and a WS model which produces networks with the small world property. We show some general results to understand the effect of the shape parameter of non-Markovian epidemics on these systems, then proceed to find the critical transition rate using the susceptibility of the system in section 5.3.1, and finally compare the results with the mean-field approximations of the literature. The same results are shown for the SIR non-Markovian epidemic process with Weibull infection in section 5.4 and 5.4.1.

5.1 Non-Markovian Epidemic process with Weibull infection

In section 4.2, we discussed the standard approach to non-Markovian epidemic modeling. The fundamental idea is to describe the infection and recovery as renewal processes. In the case of the SIR and SIS model these means:

- Recovery: will be described by T_r, a random variable that describes the interarrival times of recovery. For an infected node, T_r will return a random times τ_r at which the node will spontaneously recover, and will be given by the probability distribution $\psi_{rec}(\tau_r)$.
- Infection: for every active link (SI), the random variable T_{inf} will return a random time τ_{inf} at which the susceptible node of the active link will get infected. These time follows probability distributions $\psi_{inf}(\tau_{inf})$.

As stated in section 3.1.4.1, if the probability distribution of the renewal process is a negative exponential distribution we recover the Poisson process. For all renewal processes with probability distribution different than an exponential distribution condition (3.17) is not held, and consequently the process does not hold the Markov property.

For the case of non-Markovian epidemic processes the infection age, the time since a node was infected, and activation age, the time since an active link was formed, will affect the probability of recovery and infection. This characteristic is understood as an aging process.

The focus of this project is the non-Markovian epidemic process with Weibull infection. As we saw in the literature review in section 4.3 this is one of the most popular non-Markovian processes, and is given by an infection described by a renewal Weibull process, and a Poisson recovery process. This choice is convenient as the Weibull distribution is very versatile and maintaining the recovery as the known Poisson process reduces the complexity of the model. The probability distribution of the recovery process is the negative exponential (5.1)

$$\psi_{rec}(\tau_r) = e^{-\gamma\tau_r} \tag{5.1}$$

Where γ is the recovery rate. On the other hand, the probability distribution of the infection process is the Weibull distribution, expressed in equation 5.2 and has an expected value

given in equation 5.3 [Rinne 2008].

$$\psi_{inf}(\tau_{inf}) = \frac{\alpha}{\beta}\left(\frac{\tau_{inf}}{\beta}\right)^{\alpha-1} e^{-\frac{\tau_{inf}}{\beta}^{\alpha}} \tag{5.2}$$

$$E[\tau_{inf}] = \beta\Gamma\left(1+\frac{1}{\alpha}\right) \tag{5.3}$$

The parameter α is called the shape parameter, β is called the scale parameter, and Γ is the gamma function. In general β is left as a constant parameter, as α is the responsible of changing the general behavior of the system.

One of the most interesting characteristics of this distribution is how the shape parameter can change the behavior of the function. We can see in figure 9 that there are three main behaviors.

- For $\alpha = 1$ we retrieve an exponential distribution, this is especially important as it can be used as a base for comparing a Markovian and non-Markovian process.
- For $\alpha < 1$ the distribution takes the form of a power law distribution with a heavy tail behavior. There are no extreme values, like maximum and minimum.
- Finally, for $\alpha > 1$ the function rises and then falls, creating a maximum value. For this case, we can divide the system into two sub-cases: if $\alpha < 2$ then there is just one inflection point making the function go from a concave to a convex form, and for $\alpha > 2$ the function has two inflection points making the function going from concave to convex and then returning to concave.

.

This description doesn't give us a physical interpretation of the Weibull infection, and we have not understood where is the aging mentioned before. To understand this we may define the survival probability $\Psi_{inf}(\tau)$, also called a reliability function or complementary cumulative distribution function (CCDFF). $\Psi_{inf}(\tau)$ is the probability that a susceptible node will not get infected before a time τ. This probability can be computed as follows:

$$\begin{aligned} \Psi_{inf}(\tau) &= \int_{\tau}^{+\infty} \psi_{inf}(\tau')d\tau' \\ \Psi_{inf}(\tau) &= e^{-\tau^{\alpha}} \end{aligned} \tag{5.4}$$

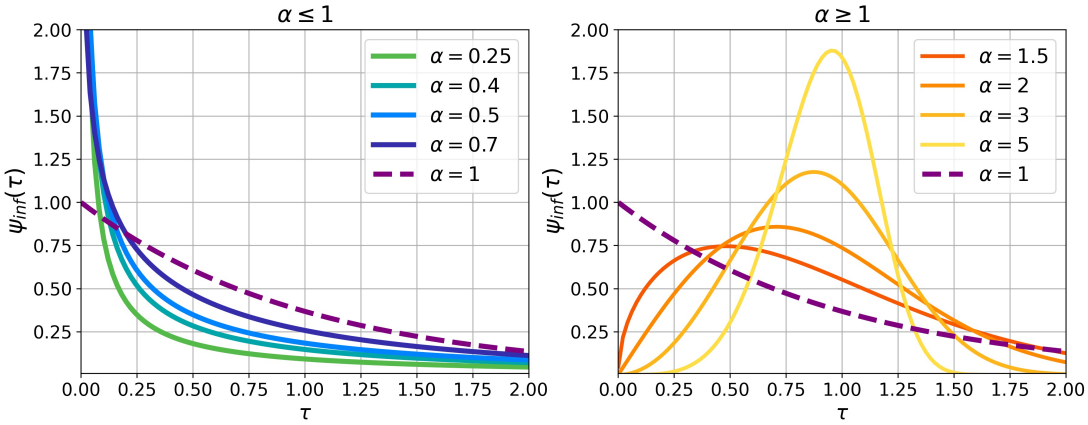

Figure 9 – Weibull probability distribution for different values of shape parameter. In a) we display distributions with $\alpha < 1$. For b), distributions with $\alpha > 1$ are shown. Both a) and b) also display the Markov case as compression in a dashed purple line. For all cases $\beta = 1$.

Source: Elaborated by the author.

From this quantity, we can also define the hazard rate of the distribution function or in the approach of Boguñá and colleagues [Boguñá *et al.* 2014] the instantaneous transition rate:

$$\lambda(\tau) = \frac{\psi_{inf}(\tau)}{\Psi_{inf}(\tau)} \tag{5.5}$$

In figure 10 we can observe the hazard rate behavior for different values of α.

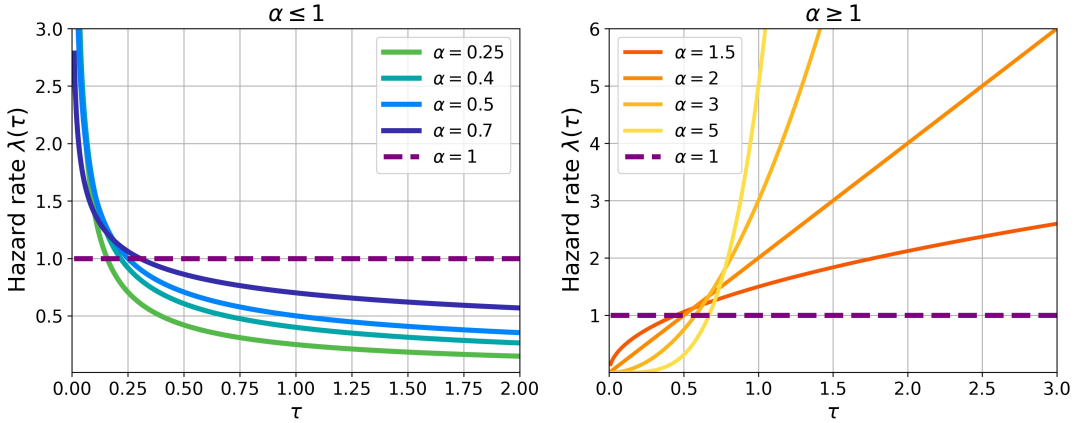

Figure 10 – Hazard rate of the Weibull distribution for different values of shape parameter. In a) we display distributions with $\alpha < 1$. For b), distributions with $\alpha > 1$ are shown. Both a) and b) also display the Markov case as compression in a dashed purple line. For all cases $\beta = 1$.

Source: Elaborated by the author.

This hazard rate can help to interpret the Weibull distribution as it works like an instantaneous infection rate. It is important to note that, although we can define this hazard rate,

using a Weibull infection process is not the same as considering a heterogeneous exponential distribution with a time-dependent infection rate. The interpretation of the Weibull infection process is related to an aging process that depends on the value of the shape parameter [Rinne 2008]. Where does this aging come from? In simple words, in the classical Markov approach the fact that an active edge has not come to an infection until a time τ does not affect the probability of infection happening in the future ($t > \tau$), so the probability of infecting a neighbor in the next 5 units of time is independent if the infected individual has been in this state for 1, 10 or 1000 units of time. Contrary to this, when considering non-exponential distributions, there is a dependency on past times, so the fact that the infection has not happened until a certain time will alter the probability of future infections. This generates an aging process, as the state age increases the probability of infection may increase or decrease depending on the shape parameter α. When $\alpha < 1$ there is a negative aging process where the hazard rate follows a power law distribution, so the probability of infection is high for early age times (τ_{inf}) and gets smaller as time passes. In the case where $\alpha = 1$ there is no aging and the transition rate is constant, and we recover the classical Markov process with exponential distribution. Lastly, for $\alpha > 1$ there is a positive aging process where the hazard rate is low at early times, and gets higher as time passes, indicating that as more time passes the the probability of infection increases.

Another important measure that helps us capture the general behavior of the system is the effective transmission rate as the rate [Starnini, Gleeson and ná 2017], which is the ratio between the average recovery time and the average infection time $\lambda_{eff} = \langle \tau_{rec} \rangle / \langle \tau_{inf} \rangle$. This quantity is useful as it can be measured in real epidemic processes, and can be used to compare with traditional Markov processes where this effective rate is the ratio between the constant infection and recovery rates. In this sense, we choose the shape parameter and effective rate and calculate the scale parameter β as a dependent variable given by $\beta = \frac{\langle \tau_{rec} \rangle}{\lambda_{eff} \Gamma(1+\frac{1}{\alpha})}$. For this case, without loss of generality we can consider a $\langle \tau_{rec} \rangle = 1$.

5.2 Simulation algorithm

Before diving into the main results of this section, let us first talk about the simulation algorithms used to obtain the results presented.

One of the most popular approaches to simulate Markovian stochastic processes is the Gillespie algorithm (GA), which was originally proposed to perform a stochastic simulation

of chemical reaction systems with known reaction rates [Gillespie 1976]. In particular, this algorithm can be used to simulate epidemic processes on networks. The algorithm randomly chooses a specific event type (infection, recovery, etc), and randomly generates the time τ when this event will occur, after each step we update the system. For a Markovian epidemic process, the time τ is generated by an exponential inter-event distribution.

This algorithm can be adapted to accurately simulate non-Markovian processes [Masuda and Rocha 2018, Boguñá *et al.* 2014]. The main difference is that in this approach, we don't choose an event type, like infection or recovery, but we consider each infected node and active link as an independent event. This assumption is necessary, as the transition rates in non-Markovian processes are not constant.

This non-Markovian Gillespie algorithm, as presented in [Boguñá *et al.* 2014], was tested and accurately produced epidemic processes. Unfortunately, this simulation is very time-consuming.

5.2.1 Event driven non-Markovian simulation

This algorithm is Gillespie-style, but the main advantage of this approach is that we avoid the slow step of finding which node becomes infected at each time step. Another interesting aspect is that the transmission and recovery rule can be any general distribution. This event-driven algorithm was taken from [Kiss, Miller and Simon 2017], which has its own Python library called *EoN*.

The base idea of the algorithm is the use of a priority queues, which is a priority structure that allows efficient addition and removal of events. The idea is: we give each event a time that depends on the inter-event probability distribution, and attach this event and its corresponding time to the priority queue. The earliest event in the queue is executed and then we add new possible events. For example, in the SIS model, if the earliest event is the recovery of a node, after the execution of the event we put in the priority queue all activation links between the new susceptible node and its infected neighbor. The author of these algorithms called them *fast non-Markov SIS* and *fast non-Markov SIR* [Kiss, Miller and Simon 2017], so we will refer to them with these names from now on.

As we mention, we will consider a Weibull infection process and Poisson recovery. Following this *fast non-Markov* algorithms, the rules that generate the random times for each

process come from these two distributions.

5.3 Non-Markovian SIS process with Weibull infection

We consider a SIS non-Markovian propagation epidemic with a Weibull infection process and a Poisson recovery process. The simulations were executed using the event-driven algorithm presented in the previous section 5.2.1, and the networks were generated using the networkx python library. First, we will show the classic dynamical curves of infected and susceptible individuals for different values of α as a function of time. This is just to have a general view of how the shape parameter might alter the evolution of a disease.

We consider 3 different networks: a random Erdős and Rényi network, a Watts Strogatz network with small world characteristics, and a scale-free network generated with the Barabási Albert method. All networks have $N = 10^3$ nodes and an approximate average degree of 10. For the BA network, we used $m = 5$ to approximate the average degree to 10, and for the WS we used the parameter $\xi = 0.05$ to guarantee the small world characteristic as explained in 2.3.3.

For these initial simulations, we took three values of $\alpha = (0.5, 1, 1.5)$, an effective transition rate of 0.75 and we establish $\gamma = 1$. Results are shown in figure 11.

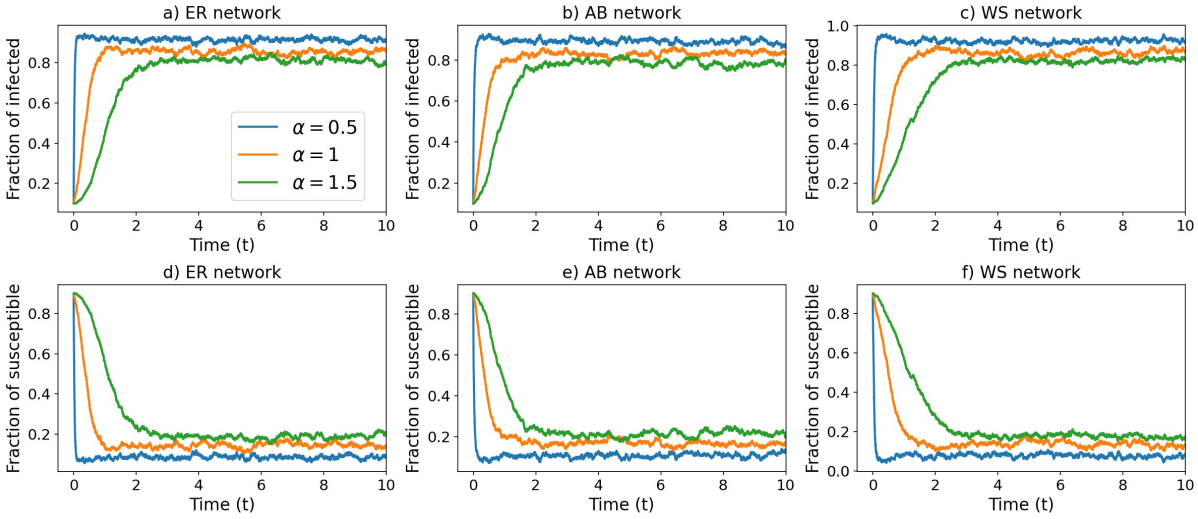

Figure 11 – Stochastic simulation of a SIS process with a Weibull infection. The fraction of infected individuals is shown in a) for Erdős and Rényi network, b) for scale-free networks, and c) a Watts Strogatz networks. The fraction of susceptible individuals is shown in d) for Erdős and Rényi network, e) for Albert Barabási networks, and f) a Watts Strogatz networks. All networks have $N = 1000$ nodes, and $\langle k \rangle = 10$.

Source: Elaborated by the author.

All curves follow the characteristic behavior of a SIS propagation with a transition rate over the critical transition rate: increasing the total number of infected nodes until the system reaches a steady state with i varying around a mean value. However, for all three network classes, the increasing value of the shape parameter results in a smaller fraction of total infected individuals. Also, the time until this steady state is reached appears to be shorter for smaller values of α.

This analysis is very illustrative, but it's not very useful, as it is a very specific case. To have a more general idea of how the shape parameter affects the epidemic processes we construct a prevalence versus effective rate curve. As mentioned in section 2.1.1 the prevalence ρ in a SIS process is given by the fraction of infected individuals at the steady state. We run extensive simulations for values of λ_{eff} between $(0,1)$ and calculated the average prevalence. This process was performed by different values of α. The resulting curve is presented in figure 12.

From these plots, it is evident that the shape parameter can drastically change the outcome of an epidemic process, no matter the network class in which the process takes place. In general, we see that the prevalence tends to increase as α decreases. For small values of α we observe epidemics that contaminate a big part of the population, even for small values of the effective rate. On the other hand, for $\alpha > 1$ the effective rate must be considerably large to reach a relevant fraction of infected individuals.

This behavior can be explained by the idea of aging inside the infected individuals. For $\alpha < 1$ early infections are prioritized as the more time a node is infected smaller the probability of transmitting the disease to its neighbors. For this reason even for smaller effective transition rates, which means that the average recovery times are much smaller compared to average infection times, the early infection is common enough that the propagation reaches more individuals. For $\alpha > 1$ we the positive aging in which more time a node is infected higher the chance of transmitting the disease. For this reason, larger effective rates, which imply in larger average recovery time compared to average infection times, are needed to have an epidemic outbreak.

So far we have just stated qualitative observations, but one quantitative aspect that is of great interest is the change in the characteristic threshold of the SIS processes.

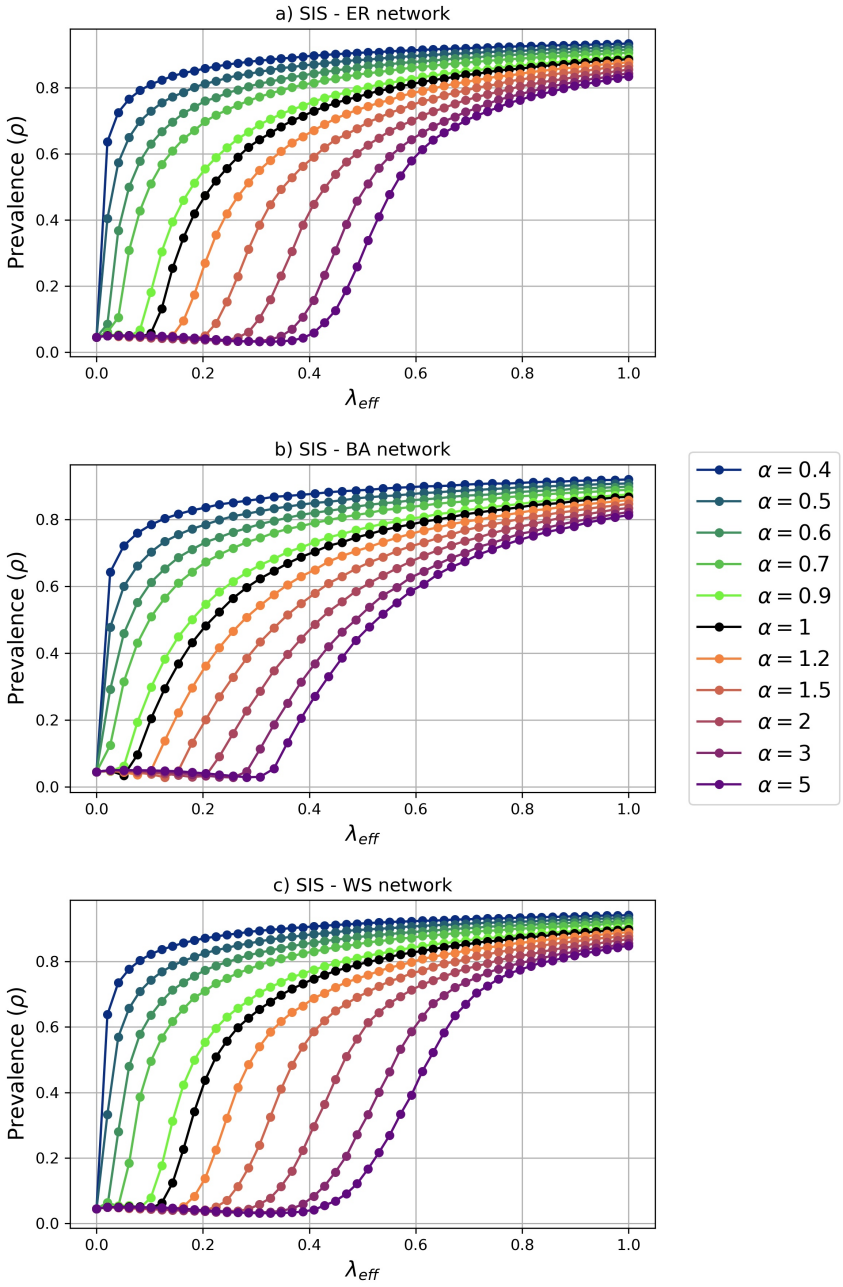

Figure 12 – Curves of average prevalence versus effective (λ_{eff}) for a SIS process with Weibull infection, each color corresponds to a shape parameter α between the interval $[0.4, 5]$. All networks have $N = 1000$ nodes and an average degree $\langle k \rangle = 10$. In a) the simulation is performed on an Erdős and Rényi network, b) on a Scale-free network, and c) on a Watts Strogatz network.

5.3.1 Critical transition for the SIS non-Markovian epidemic processes.

In section 3.2.2.2 we talked about numerical methods for estimating the critical effective transition rate for Markovian epidemic processes. The most popular and in general adequate approach for a SIS process is using the quasi-stationary analysis. Unfortunately, there still isn't a

direct conversion of this analysis for general infection distributions. In principle, this should not represent a big challenge, but as it is not the main focus of this project we used the more "brute force" method. In this case, we made extensive simulations and store the prevalence values for those simulations that reached the steady state or before reaching the absorbing state $(i = 0)$. With these prevalence values we calculate the dynamical susceptibility:

$$\chi = N\frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle} \tag{5.6}$$

For the critical effective transition rate, the susceptibility takes a maximum value.

Based on the works of P. Van Mieghem and his collaborators [Cator, Bovenkamp and Mieghem 2013], we expect that the critical effective rate for this non-Makrovian SIS process with Weibull infection presents the same behavior as the Markovian SIS process for heterogeneous and homogeneous networks. In other words, for the BA network which is scale-free and heterogeneous, we expect a vanishing threshold when the system size is increased. On the other hand, for the homogeneous networks, ER and WS the critical value should not vanish.

To verify this behavior we run extensive simulations in networks with sizes from $N = 10^2$ to $N = 5x10^3$ and for different values of shape parameters and effective transition rate for all the three network models.

In figures 13, 14 and 15 we show the susceptibility (5.6) of simulations performed in the ER networks, BA networks, and WS networks, respectively, for different network sizes, effective transition rates, and shape parameter. Each value of the susceptibility is calculated using the prevalence values of an average of 100 to 500 simulations in 200 to 300 networks, depending on the network's size.

In figures 13, 14 and 15 we can observe the peak behavior expected for the susceptibility. This peak indicates a numerical approximation of the critical effective transition rate. Note that for $N = 10^2$ the peaks are not as well defined as in the other sizes, especially for smaller values of α. Also, for the WS and ER networks, there are fluctuations for the bigger sizes $N = 2 \times 10^3$ and $N = 5 \times 10^3$, but the peak still is well defined.

From this susceptibility analysis, we can plot the critical transition rate as a function of the shape parameters. These values are shown in figure 16.

The ER network and WS are homogeneous networks, a characteristic of this kind of

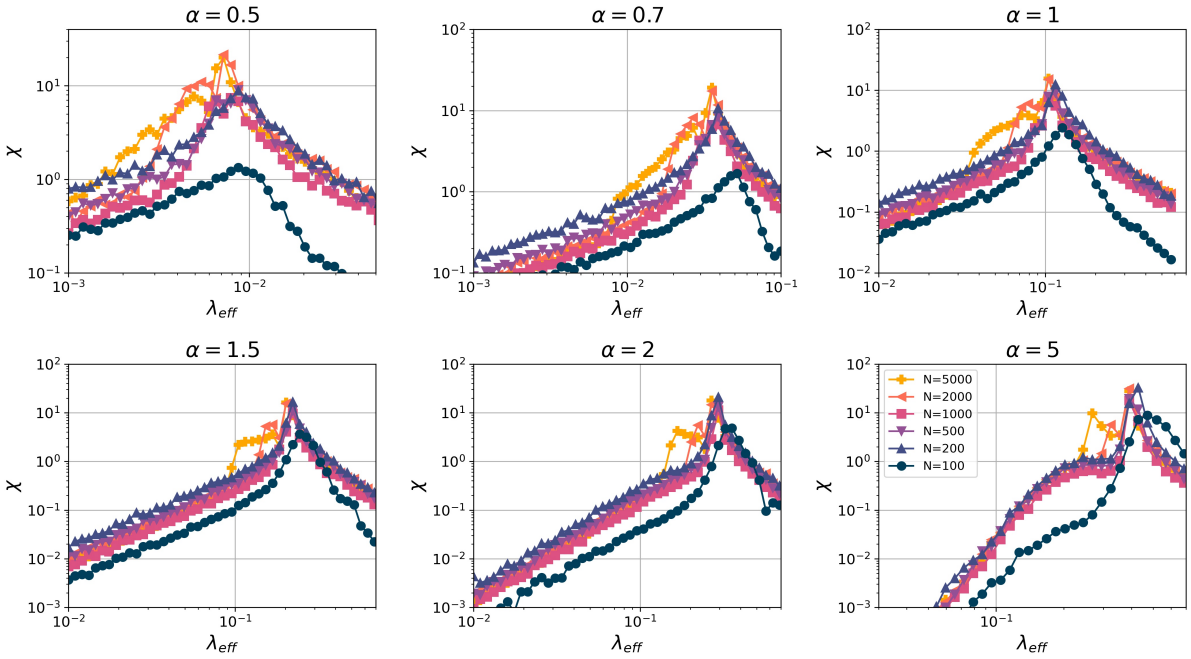

Figure 13 – Susceptibility χ as a function of the effective transition rate λ_{eff} in log scale for Erdős and Rényi networks. Each plot corresponds to a different shape parameter α, indicated at the top, and each color is associated with a different network size in the interval $[N = 10^2, N = 5x10^3]$.

Source: Elaborated by the author.

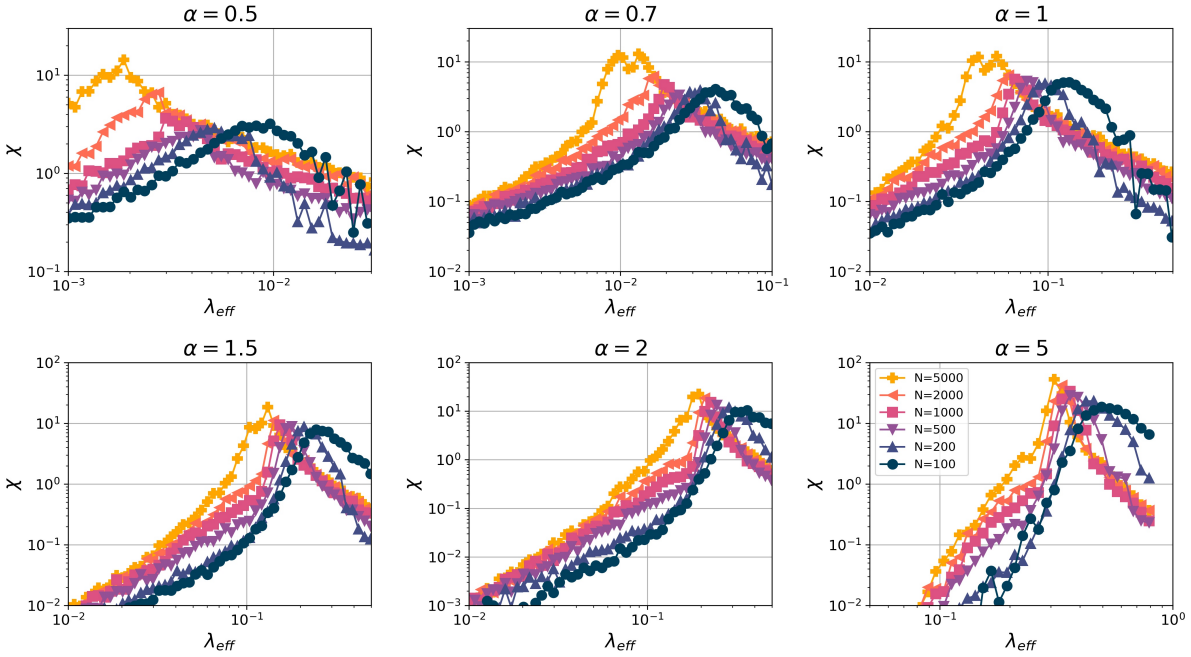

Figure 14 – Susceptibility χ as a function of the effective transition rate λ_{eff} in log scale for Barabási Albert networks. Each plot corresponds to a different shape parameter α, indicated at the top, and each color is associated with a different network size in the interval $[N = 10^2, N = 5x10^3]$.

Source: Elaborated by the author.

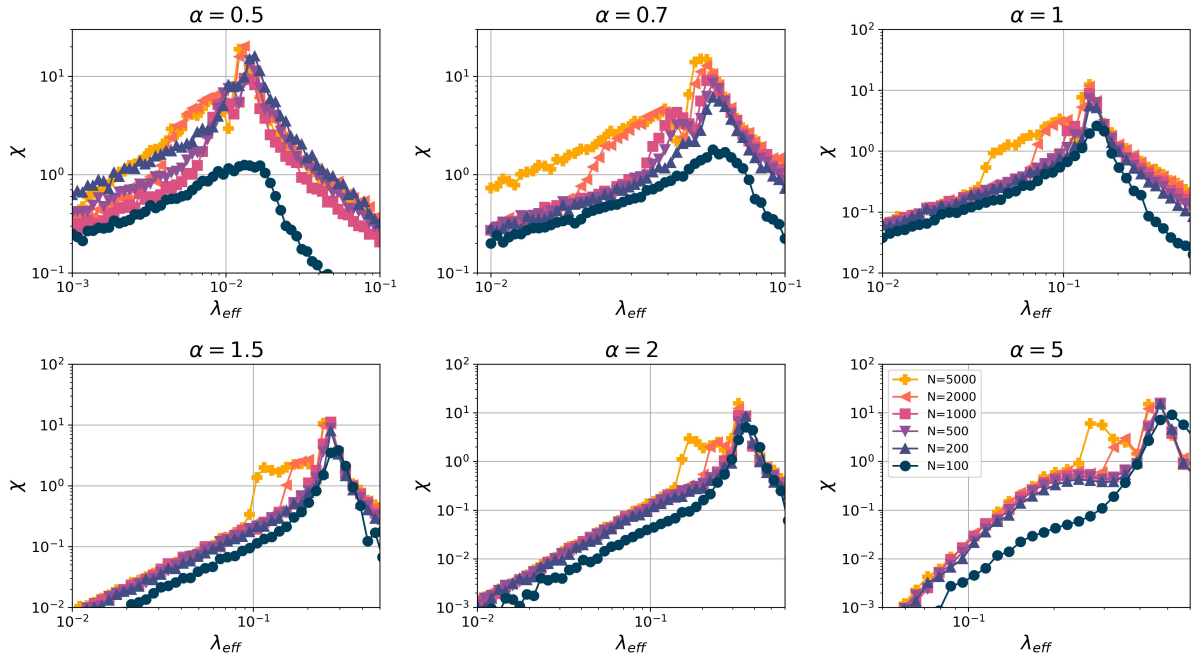

Figure 15 – Susceptibility χ as a function of the effective transition rate λ_{eff} in log scale for Watts Strogatz networks. Each plot corresponds to a different shape parameter α, indicated at the top, and each color is associated with a different network size in the interval $[N = 10^2, N = 5x10^3]$.

Source: Elaborated by the author.

network is that λ_c is approximately the same independent of the size of the system. This can be observed in figure 16a and 16c where each curves that represents different sizes almost overlap. The major out-liner for both cases is $N = 10^2$ where as we discussed the peak values are less defined. On the other hand, for the BA network which is heterogeneous, we know critical rate vanishes in the thermodynamic limit. This also is observed as the critical values decrease as the size of the system increases.

Now let us compare our findings to the models and mean-field approaches described in section 4.3.1. From this point forward we will just focus on bigger networks, as they are the ones with better-defined peaks.

We can compare the critical rate found numerically through the susceptibility analysis λ_c^{χ} with quantities described in section 3.2.2.2: the NIMFA mean field approximation λ_c^{NIMFA} and the upper bound λ_c^{UB} for SIS epidemic threshold, both described by P. Van Meighen and colleagues [Mieghem and Liu 2019, Cator, Bovenkamp and Mieghem 2013, Liu and Mieghem

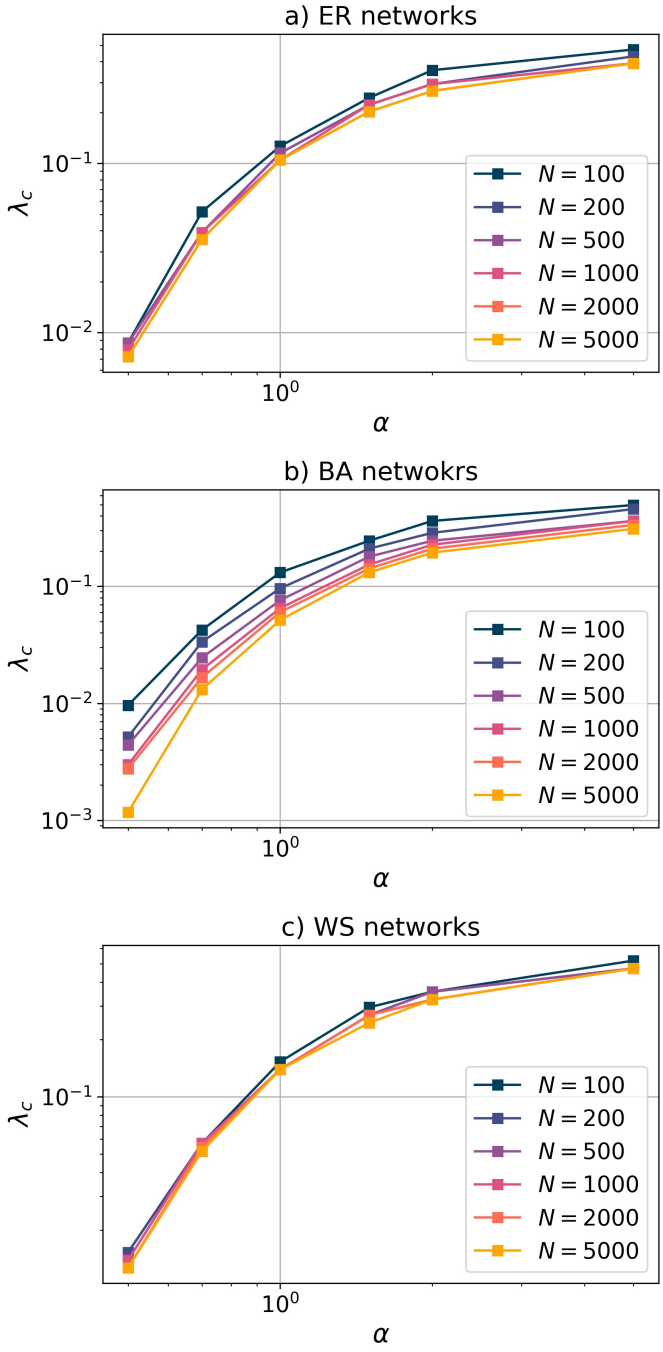

Figure 16 – Critical effective transition rate as a function of the shape parameter α for different network sizes and models for a SIS epidemic process. In a) critical rates for ER model, in b) for the BA model, and in c) a WS model.

Source: Elaborated by the author.

2018]. These expressions are written in equation (5.8) and (5.7) to have them at hand.

$$\lambda_c < \lambda_c^{UB} = \frac{1}{ln(\Lambda_A + 1)} \tag{5.7}$$

$$\lambda_c^{NIMFA} = \frac{1}{\Gamma(1+\frac{1}{\alpha})[\Gamma(\alpha+1)]^{1/\alpha}} \frac{1}{\Lambda_A^{1/\alpha}} \tag{5.8}$$

In figure 17 we compare these three quantities for the three networks models ER, BA, and WS for $N = 10^3, 2\times10^3, 5\times10^3$.

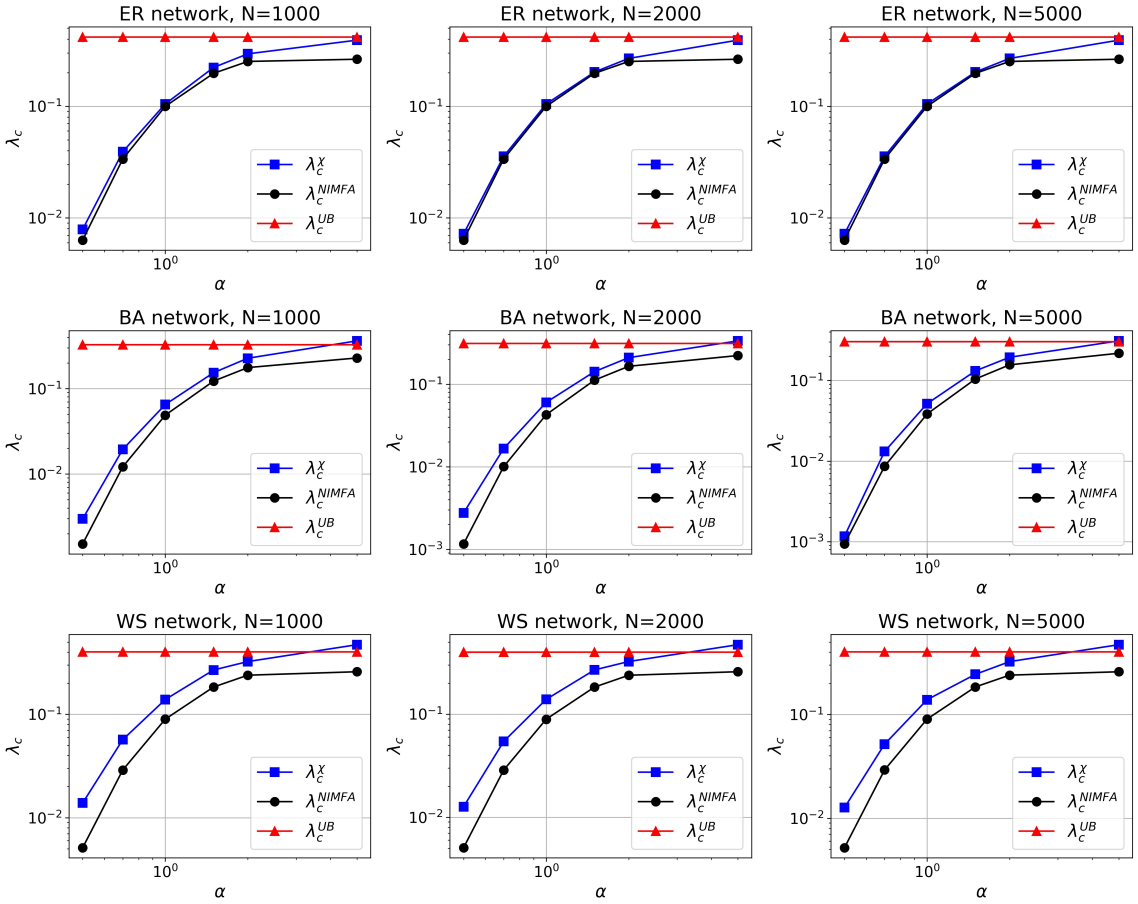

Figure 17 – Compression of the critical effective rate λ_c calculated with the susceptibility analysis with the NIMFA approximation λ_c^{NIMFA} and the upper bound λ_c^{UB} described in section 4.3.1. Each plot correspond to a different size and network model, stated in the title above of each image.

Source: Elaborated by the author.

For the ER networks, in the first row of the plot in figure 17, we see that the numerical values found are very close to the NIMFA mean field approximation for almost all values of α. For $\alpha = 2$ and $\alpha = 5$ the numerical approximation is higher than the mean-field approach. Also, as N increases these quantities coincide even more, which was expected as the mean-field approximation works better for bigger systems [Cator, Bovenkamp and Mieghem 2013]. Additionally, the λ_c^{χ} found are consistently bellow the upper bound λ_c^{UB}.

For the BA network, the calculated critical rate is in general above the NIMFA approxi-

mation but follows the general behavior of the function. For $N = 5000$ the two values start to get closer, so we expect that for bigger systems the threshold would coincide. On the other hand, the upper bound λ_c^{UB} still works well.

In the case of WS networks the numeral approximation is also above the NIMFA approximation. However, in this case, the gap does not seem to be getting smaller as the size increases. The NIMFA approximation for WS networks is almost identical to the ER networks, because they are both homogeneous networks, and the calculation of the largest eigenvalue is directly related to the probability distribution, similar for both models. The WS network is a homogeneous network with high clustering and a small average path length. It has been proven that for homogeneous networks clustering may inhibit the epidemic and even can increase the critical rate [Song, Song and Jiang 2017].

To understand how this affects the compression between λ_c^{χ} and λ_c^{NIMFA} we must remember some properties of the WS network. If $\xi = 1$ the WS network produces a random regular network, like the ER model, and as we increase ξ the clustering increases. In figure 18 we compare the prevalence vs effective rate λ_{eff} for WS networks with different values of ξ.

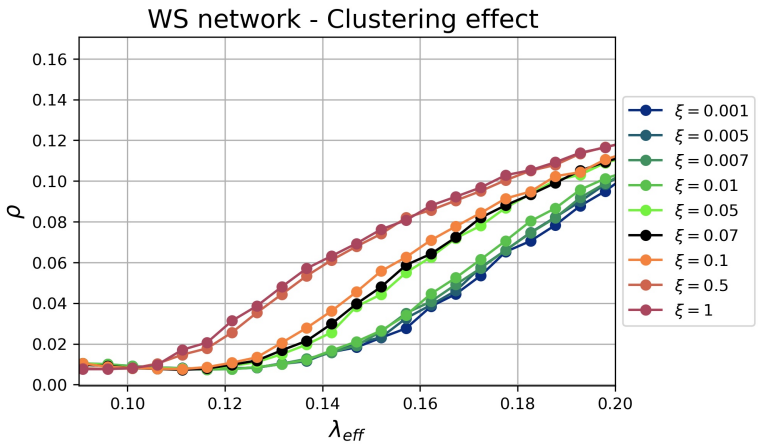

Figure 18 – Prevalence curves for a SIS epidemic process as a function of the effective transition rate λ in a Watts Strogatz model with different values of ξ.

Source: Elaborated by the author.

As ξ decreases, and clustering increases, we see that the critical rate increases. This clustering is lost in the NIMFA approximation so is expected that the mean-field approach is a little smaller than the values found with the numerical method.

Lastly, we will compare the there models with each other. In figure 19 we see a plot for $N = 5 \times 10^3$ that compares the critical rate found for the 3 models.

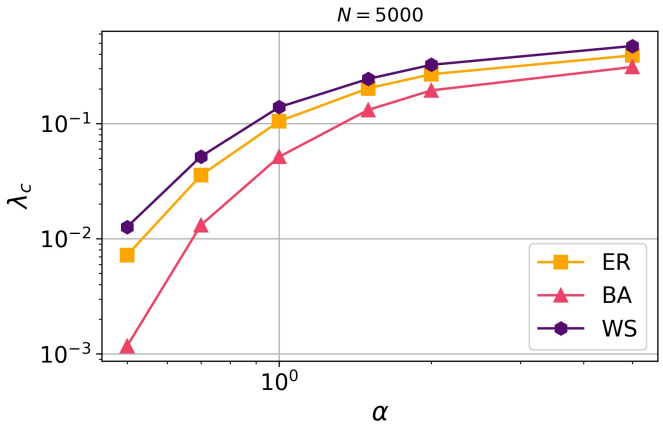

Figure 19 – Compression of the critical transition as a function of the shape parameter α between the three models: Erdős and Rényi, Barabási Albert, and Watts Strogatz.

Source: Elaborated by the author.

The results are coherent with the expectations. The critical rate for BA networks, which are heterogeneous, is considerably smaller than the homogeneous networks WS and ER. For bigger systems, this difference would increase as the heterogeneous networks have a vanishing threshold for $N \to \infty$. On the other hand, the WS networks have higher critical rates than the ER networks, despite both being homogeneous and having the same average degree. This happens because, as explained before, the clustering in WS networks helps inhibit epidemics and enlarge the critical rate, as shown in figure 18.

5.4 Non-Markovian SIR process with Weibull infection

For the SIR propagation process, we expect a similar outcome to the SIS model. First, we simulated some specific cases just to have a general view on how a Weibull infection process can modify an epidemic.

Again, we used three network models: Erdős and Rényi networks (ER), a Watts Strogatz network (WS), and a scale-free network constructed with the Barabási Albert algorithm (BA), all with a total number of nodes $N = 10^3$, and an approximate average degree of $\langle k \rangle = 10$. Each network simulated three different cases, each one with different values of α. In figure 20 we can see the dynamical curves of both the fraction of infected and the fraction of recovered individuals

For the smallest value of the shape parameter, we see a rapid epidemic process that reaches the maximum fraction of infected much quicker than the other two cases. Also, this

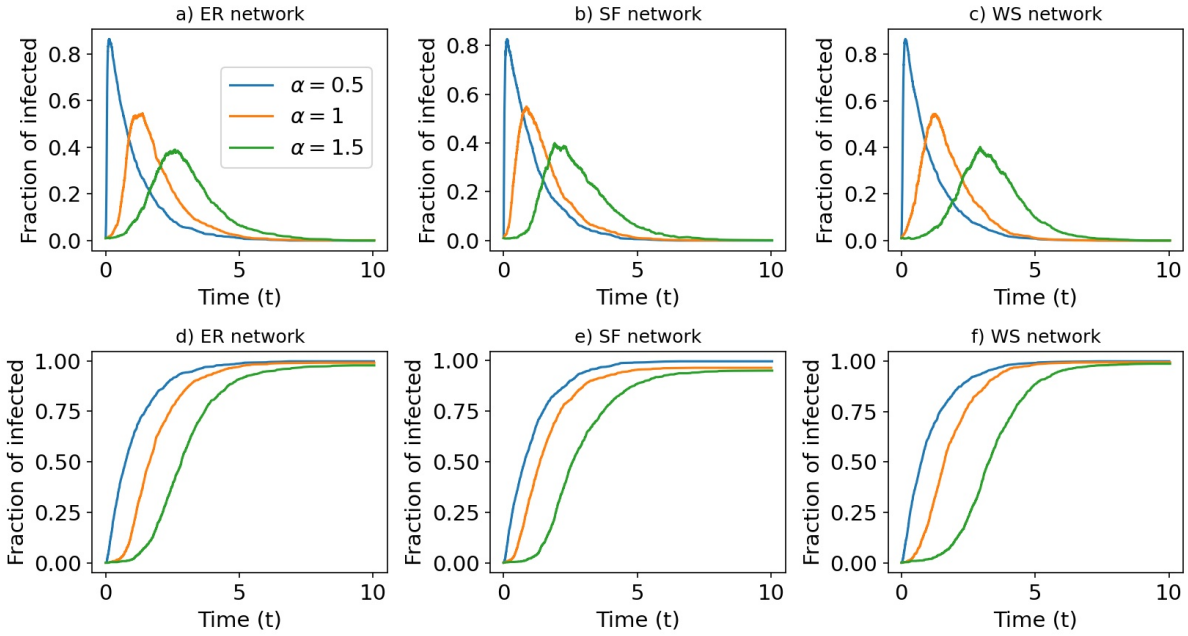

Figure 20 – Stochastic simulation of a SIR process with a Weibull infection and a Poisson recovery. Each plot is simulated on different networks: a) and d) for an Erdős and Rényi networks, b) and e) for a scale-free network, and c) and f) a Watts Strogatz networks. All networks have $N = 1000$ nodes, and an average degree of 10.

Source: Elaborated by the author.

maximum decreases and the epidemic lifetime lengthens as α increases. As expected, the size of the epidemic, given by the total number of recovered, decreases with α, which means a bigger fraction of the population was contaminated before the end of the epidemic.

For a more robust result, we may construct curves of prevalence e as a function of the λ_{eff} parameter. Note that in this case, we define the prevalence as the total number of recovered individuals, which is the same as the epidemic size. This effective rate follows the same logic as the one used for the SIS model, in other words, is defined as $\lambda_{eff} = \langle \tau_{rec} \rangle / \langle \tau_{inf} \rangle$. Also, it serves as a more intuitive variable as it can be compared with the effective parameter of the Markovian case. We perform 600 hundred simulations for each different value of λ_{eff} and α and calculate the average infection size. The results of these simulations are shown in figure 21.

These results sustain the claims made for the specific case presented in figure 20 and show that a SIR propagation with non-Markovian Weibull infection can radically change the resulting epidemic. As the shape parameter increases, larger values of λ_{eff} are necessary to reach a considerable fraction of infected nodes. The critical effective rate also seems to increase with larger α. To verify this behavior and make a more quantitative analysis we will repeat the critical rate analysis made in the last section with the SIS model.

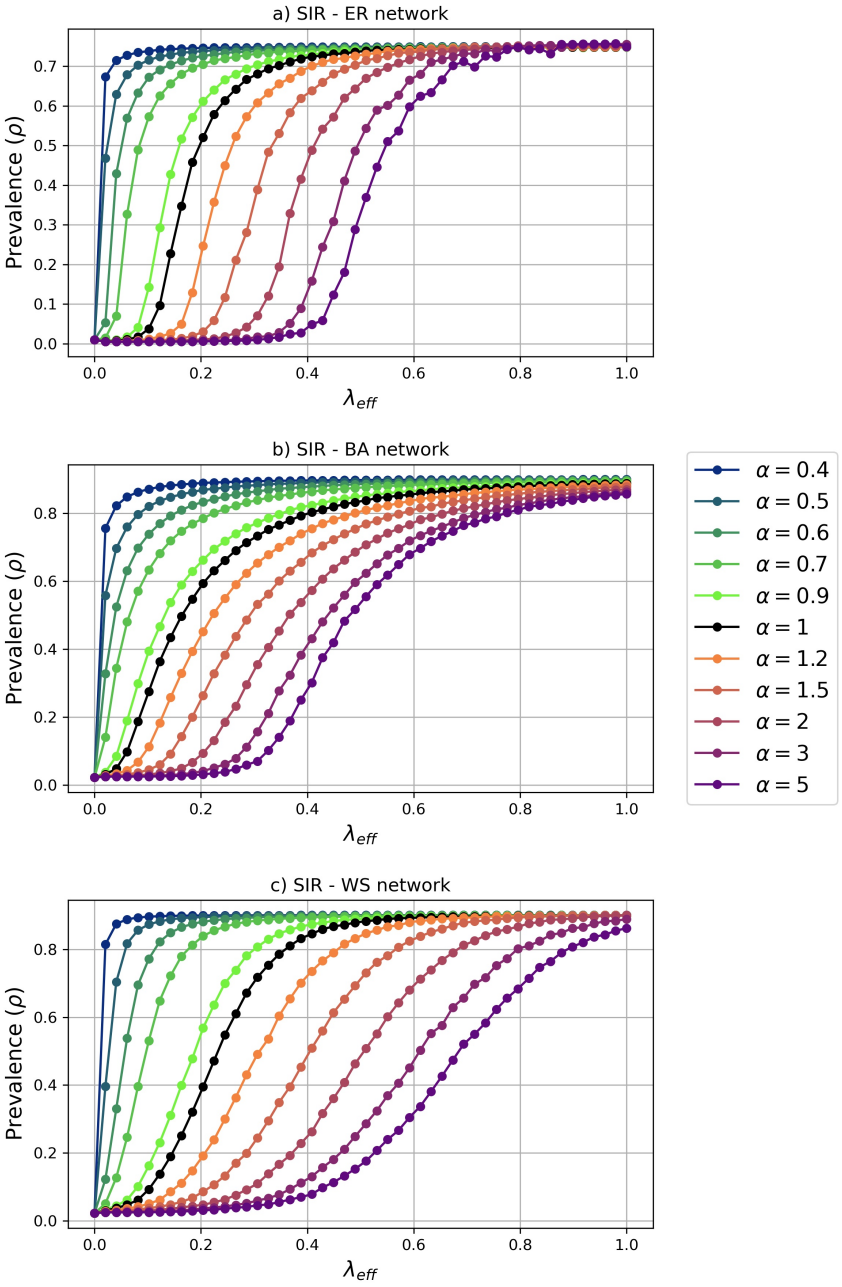

Figure 21 – Curves of Prevalence versus effective rate for a SIR process with Weibull infection, each color corresponds to a shape parameter α between the interval $[0.4, 5]$. All networks have $N = 1000$ nodes and an average degree $\langle k \rangle = 10$ and prevalence values are calculated over 600 repetitions. In a) the simulation is performed on an Erdős and Rényi network, b) on a Barabási Albert network, and c) on a Watts Strogatz network.

Source: Elaborated by the author.

5.4.1 Critical transition for the SIR non-Markovian epidemic processes

The procedure is similar to the one described in the previews section. We use the epidemic size or prevalence as the order parameter ρ, which for the SIR process is the total number of

recovered individuals at the end of the simulation. Extensive simulations were carried out to calculate the first and second moments of the order parameter. For the SIR process, as explained in section 3.2.2.2, a better measure to find the critical transition rate is the vulnerability instead of the susceptibility used in the SIS model. The vulnerability is defined as:

$$\Delta = \frac{\sqrt{\langle \rho^2 \rangle - \langle \rho \rangle^2}}{\langle \rho \rangle} \tag{5.9}$$

The vulnerability Δ has a maximum value at the critical transition rate λ_c.

We will go a little faster in this section as the results presented for the SIR model repeat many the same behavior as the SIS model discussed before.

From the works done on Markovian epidemics, we expect heterogeneous networks, like the BA network, to have a vanishing threshold at the thermodynamic limit. However, for the WS and ER networks that are homogeneous, we expect a constant critical rate despite the size of the system. To verify this behavior we run extensive simulations in networks with sizes ranging from $N = 10^2$ to $N = 5 \times 10^3$.

In figure 22, 23 and 24 we show the vulnerability measure as a function of effective transition rate for the different network models: ER networks, BA networks, and WS networks, respectively.

From figures 22, 23 and 24 we can calculate the critical rate by taking the effective rate that gives the maximum value of Δ. From this critical rate, we can make a plot of the λ_c as a function of the shape parameter for three models. These plots are shown in figure 25.

In this case, different from the SIS model, peaks are less well defined, especially for small system sizes like $N = 10^2$ and $N = 2 \times 10^2$. Nevertheless, we can still see the difference between the homogeneous networks ER and WS with the heterogeneous BA network. For the homogeneous networks, ER and WS, despite points from different sizes do not coincide as well as in the SIS model, this variation does not follow any particular tendency. However, for the BA network critical rates calculated from bigger networks are consistently below the critical sizes of smaller networks. This is a sign of the vanishing threshold expected for heterogeneous networks.

For the SIR non-Markovian process with Weibull infection, the development of a mean-field approximation for the critical effective rate remains an ongoing challenge. However, as stated by P. Van Mieghem in [Mieghem and Liu 2019], the NIMFA mean field approximation

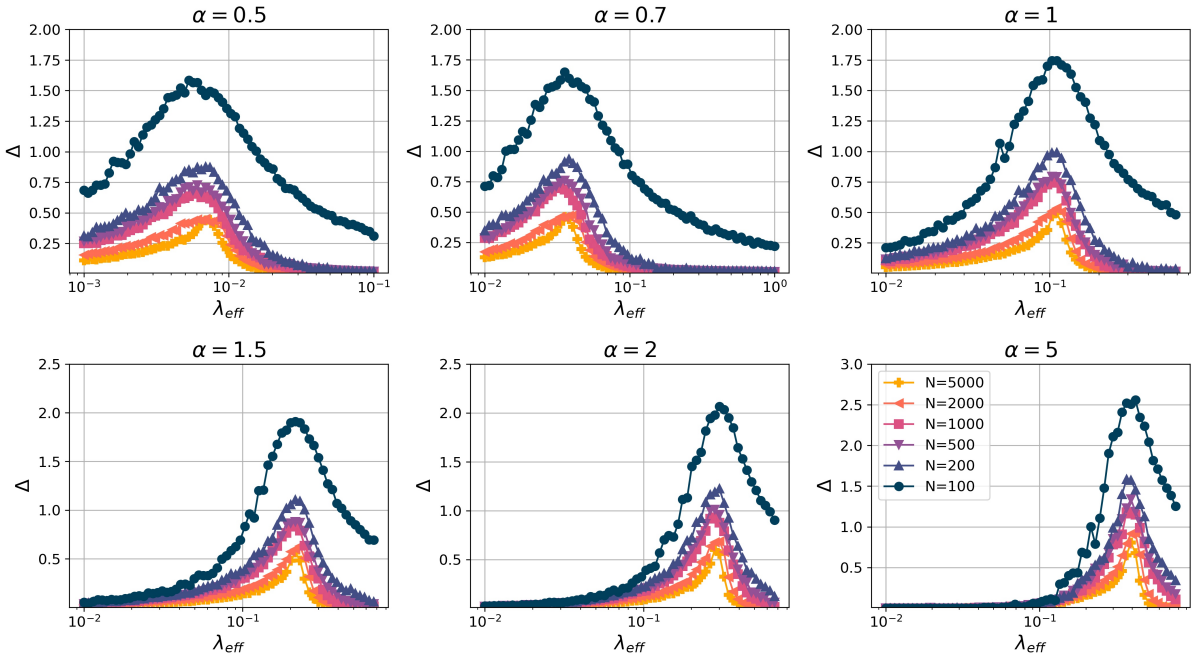

Figure 22 – Vulnerability Δ as a function of the effective transition rate λ_{eff} in log scale for Erdős and Rényi networks. Each plot corresponds to a different shape parameter α, indicated at the top, and each color is associated with different network size in the interval $[N = 10^2, N = 5x10^3]$.

Source: Elaborated by the author.

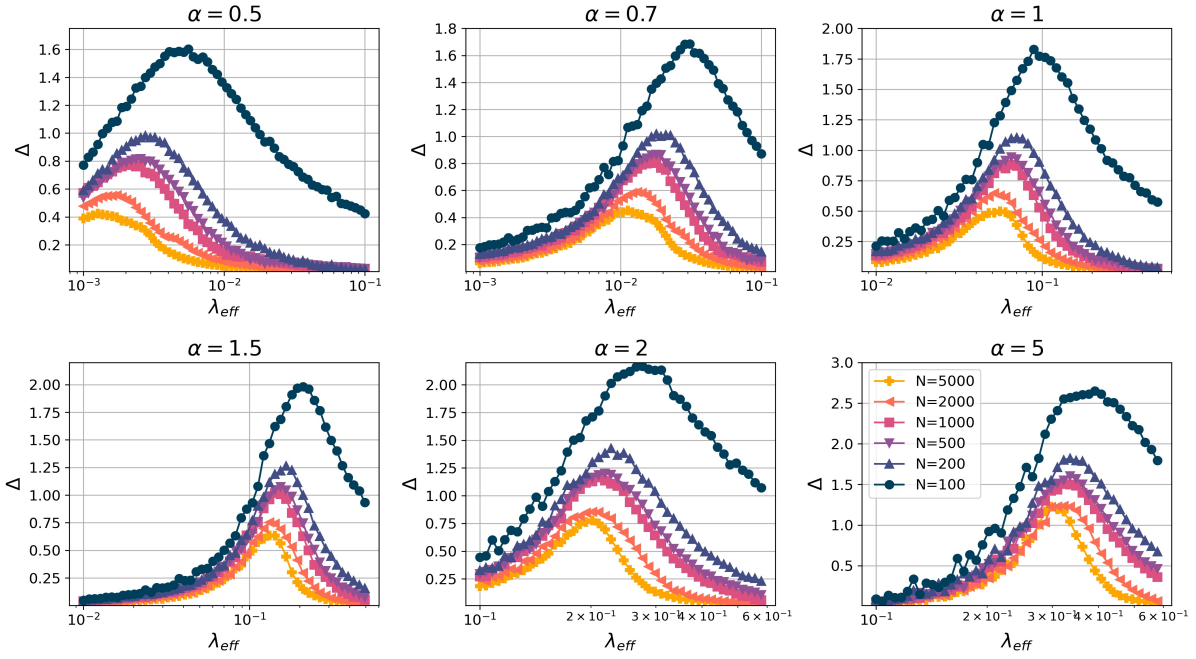

Figure 23 – Vulnerability Δ as a function of the effective transition rate λ_{eff} in log scale for Barabási Albert networks. Each plot corresponds to a different shape parameter α, indicated at the top, and each color is associated with different network size in the interval $[N = 10^2, N = 5x10^3]$.

Source: Elaborated by the author.

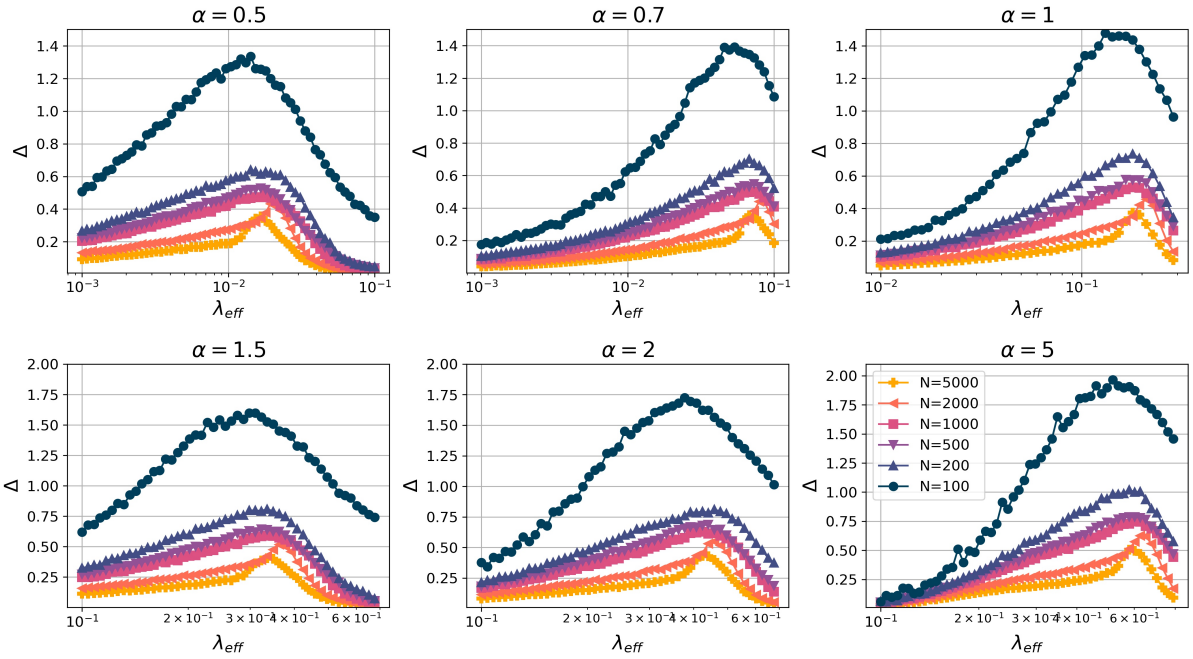

Figure 24 – Vulnerability Δ as a function of the effective transition rate λ_{eff} in log scale for Watts Strogatz networks. Each plot corresponds to a different shape parameter α, indicated at the top, and each color is associated with different network size in the interval $[N = 10^2, N = 5x10^3]$.

Source: Elaborated by the author.

for the SIS model can be used as a lower bound for the SIR critical. Typically, the SIR epidemic threshold is slightly higher than the SIS threshold. With this in mind, we will compare the numerical approximations λ_c^{Δ} found for the SIR model with the $\lambda_) c^{NIMFA}$ and the upper bound λ_c^{UB} described in (4.1) and (5.7). This compression is shown in figure 26

For the ER and BA networks the numerical approximation critical rates found are in general slightly above the NIMFA critical rate, although is closer than expected for the ER networks. However, for the WS networks, the critical rate is considerably above the lower bound given by the NIMFA approximation. It is also worth noting that, for $\alpha = 2$ and $\alpha = 5$ the numerical approximation is above the upper bound, which was not expected. We saw in the previous section that for higher values of the shape parameter the numerical approximation gets further from the mean-field approach, so it is possible that in this case, the critical rate passing the upper bound is due to some additional care needed for these values of shape parameter.

Finally, let us compare the three models with each other. In figure 27 we plot the critical effective rates for the largest sizes for each network model as a function of the shape parameter.

Once again, this is very similar to the SIS process. All three networks have an increasing threshold as we take larger values of α. These results are coherent with the definition of the

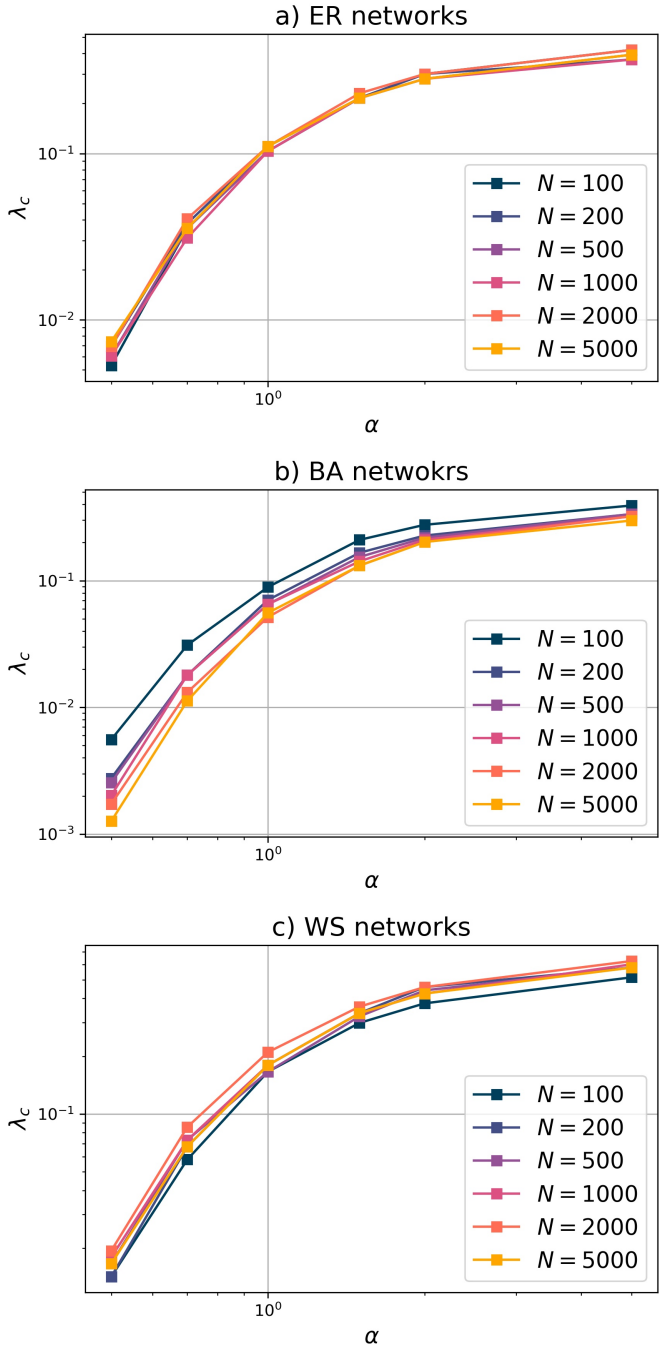

Figure 25 – Critical effective transition rate as a function of the shape parameter α for different network sizes and models for a SIR epidemic process. In a) critical rates for ER model, in b) for BA model and in c) a WS model.

Source: Elaborated by the author.

Weibull distribution and the role of the shape parameter. Smaller values of α generate bigger probabilities to smaller inter-event times, so we expect a faster epidemic where nodes infect their neighbors with a higher frequency per unit of time.

The BA network has in general lower threshold compared to the other two networks.

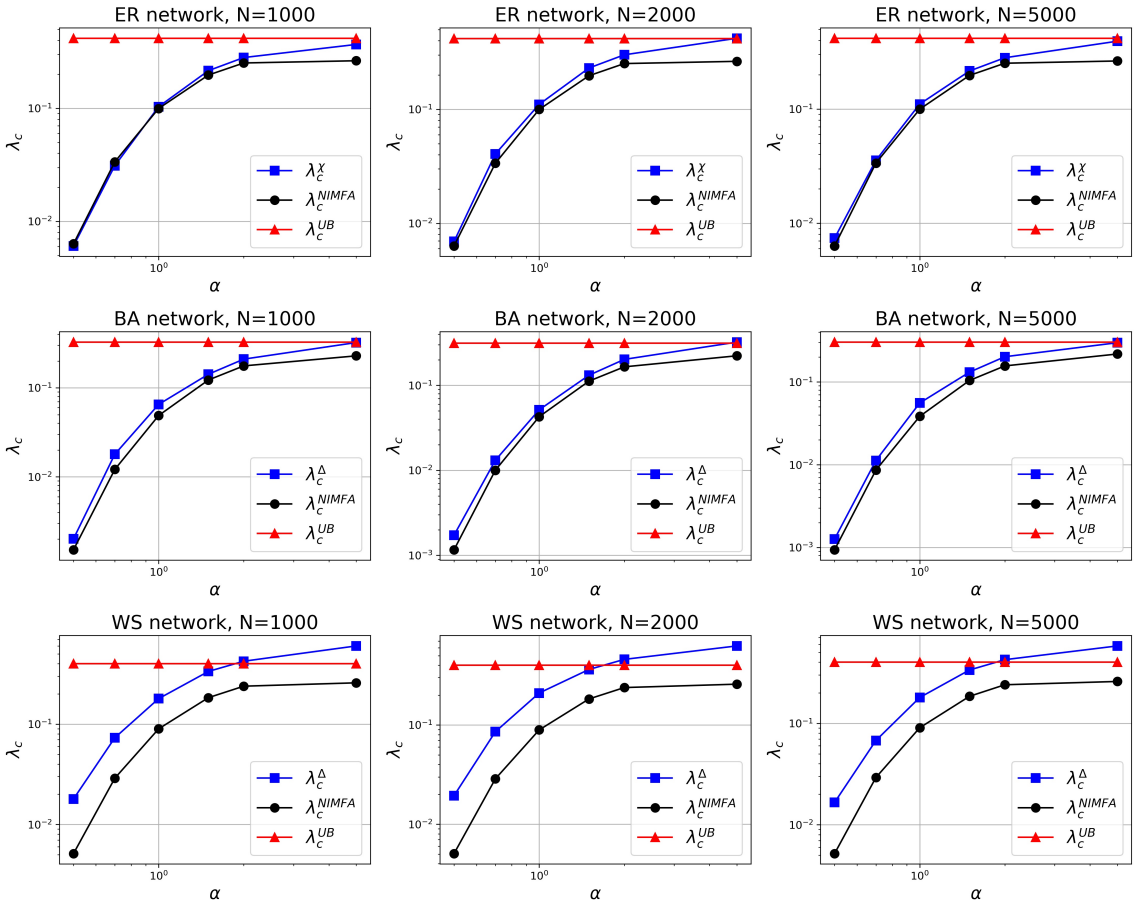

Figure 26 – Compression of the critical effective rate λ_c calculated with the vulnerability analysis in a SIR process with the NIMFA approximation λ_c^{NIMFA} and the upper bound λ_c^{UB} for the SIS epidemic model described in section 4.3.1. Each plot correspond to a different size and network model, stated in the title above of each image.

Source: Elaborated by the author.

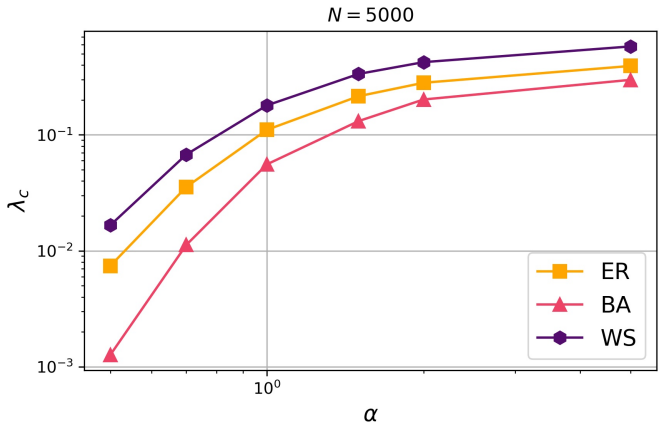

Figure 27 – Compression of the critical transition as a function of the shape parameter α in a SIR epidemic process between the three models: Erdős and Rényi, Barabási Albert, and Watts Strogatz.

Source: Elaborated by the author.

Regarding the WS network's critical rates being above the ER, the interpretation can be the same as the last subsection. The increase in clustering, due to the small-world characteristic, helps inhibit the epidemic and enlarges the critical rate of the homogeneous network.

5.5 Summary

In section 5.1, we described the implications of considering a Weibull infection process. The shape parameter of the Weibull distribution is associated with aging in the infection process. When $\alpha < 1$ there is a negative aging in which the hazard rate has a heavy-tailed behavior and indicates a high probability of early infections, and as time passes the infected node is less likely to transmit the disease. For $\alpha > 1$. there is a positive aging that causes a hazard rate that increases with time, which implicates a low probability of infection in early times and a high probability as the node stays infected.

We verified for both the SIS and SIR models that positive aging can inhibit the epidemic, generating smaller epidemic sizes and slower propagation. On the other hand, negative aging has the opposite effect, making it easier for the disease to spread and causing higher epidemic sizes. We also verified that the critical effective transmission rate λ_c can dramatically change for Weibull infection, specifically, increasing with α.

In this study, we investigated the behavior of SIS and SIR epidemics on heterogeneous networks generated by the Barabási-Albert (BA) model, as well as on homogeneous networks such as the Watts-Strogatz (WS) and Erdős-Rényi (ER) models. Our findings revealed that the critical rate in the BA networks decreased as the system size increased, implying a vanishing threshold. On the other hand, for the homogeneous WS and ER networks, we did not observe a clear change in the critical rate as the system size varied. The critical rate in these networks remained relatively stable with respect to the network size.

Additionally, we compared the numerically determined values of the critical rate, denoted as λ_c, for the SIS model with the NIMFA approximation reported in the literature [Cator, Bovenkamp and Mieghem 2013, Mieghem and Liu 2019]. Notably, our results demonstrated a good agreement between the numerical values and the NIMFA approximation, particularly for the ER model. For the SIR model, the critical rate generally fell between the upper bound described for the SIS model and the λ_c^{NIMFA}, except for higher values of the shape parameter ($\alpha = 5$), where the critical rate surpassed the upper bound. This intriguing finding indicates that

the critical rate for the SIR model may display unexpected behavior under certain conditions.

Finally, for both SIS and SIR model the WS networks presented higher values of the critical effective rate than the ER network, despite having the same average degree. The behavior probably occurs due to clustering effects. The BA network presented the lower values of the 3 models, and its expected that if larger networks are used then $\lambda_c \to 0$.

CHAPTER

6

SIR NON-MARKOVIAN EPIDEMIC PROCESS WITH WEIBULL INFECTION IN MODULAR NETWORKS

In the previous chapter, we talked about the SIS and SIR non-Markovian epidemic process with Weibull infection in different network models. These models had different characteristics like the small world property of the WS model or the presence of hubs in the BA model. However, there are still many unanswered questions related to the impact of network structure on non-Markovian epidemic processes. For instance, Modular structure has been demonstrated to be relevant for modeling both social and biological phenomena [Palla *et al.* 2005, Spirin and Mirny 2003, Ravasz *et al.* 2002, Watts, Dodds and Newman 2002] and spreading processes in networks [Ghalmane *et al.* 2019]. This modular organization generates groups of densely connected nodes [Fortunato and Hric 2016], making it a crucial feature that cannot be overlooked.

In this chapter, we explore the influence of community structure on a SIR epidemic process with a Weibull non-Markovian infection. Most of the results presented here are part of an article that is the final stages of preparation and will be submitted to Physical Review E.

In section 6.1 and 6.2, we present a brief review of epidemics in networks with community structure and explore the LFR benchmark algorithm used to create this kind of networks. After this, we see some specific cases of the non-Markovian SIR process with Weibull infection to have a general notion of the influence of aging in the system. In the proceeding section, we

explore how the shape parameter may change the prevalence and epidemic life, and how this non-Markovian process can alter the role of modular structure. Finally, in section 6.4 we show simulations in real networks that further support our previous results in synthetic networks.

6.1 Revision of epidemics on Modular networks

Topology plays a fundamental role in how diseases spread in networks. For instance, the presence of communities can drastically change the development of an epidemic outbreak. Some effort has been devoted to understanding the role of these communities in slowing down or speeding up contagious processes.

One of the most frequent results is that modularity can significantly slow SIR and SIS propagation if we straighten the modular structure [Salathé and Jones 2010, Nadini *et al.* 2018, Hui and Zi-You 2007]. Also, the presence of modules can reduce the final epidemic size [Huang and Li 2007, Huang, Park and Lai 2006]. This can be measured using the mixing parameter, which is a way to measure how strong the structure of a network is [Min *et al.* 2013, Pastor-Satorras *et al.* 2015].

Other studies focus on understanding how modularity affects epidemics in more networks with specific characteristics. For example, studies with modular time-varying networks [Nadini *et al.* 2018], or modular adaptive networks [Tunc and Shaw 2014]. Other researchers focus on developing immunization strategies using community structures [Gupta, Singh and Cherifi 2014, Ghalmane, Hassouni and Cherifi 2019].

6.2 LFR algorithm

To address the challenge of creating networks with community structure we used the Lancichinetti–Fortunato–Radicchi benchmark algorithm (LFR). This strategy originally was proposed to test community detection methods, but it also allows us to create a network with known communities [Lancichinetti, Fortunato and Radicchi 2008]. The main advantages of this algorithm is the explicit mixing parameter μ, which is the average fraction of inter-community links, this is, edges between nodes that don't belong to the same community. The algorithm consists in distributing the node degrees and the community sizes according to a power law distribution with different exponents Υ and Ω respectively. The general algorithm can be described:

- Generate a network with node degrees that follow a power law distribution with exponent Υ and an average degree $\langle k \rangle$.

- Generate a community sizes that follow a power law distribution with exponent Ω. It is necessary to establish the s_{max}, the maximum community size, and s_{min} the minimum community size.

- We randomly distribute the N nodes to the M communities. If a node is put in a complete community, then we randomly choose another community.

- We rewire the links with to goal to approximate the average fraction of external connection to total connections to the pre-defined μ parameter.

An advantage of this algorithm is that allows us to construct modular networks with stronger or weaker community structures depending on the mixing parameter. If we set $\mu = 0$, we generate a graph with isolated communities without connections between modules. On the other hand, if we set $\mu = 1$ we have a fully mixed network where all connections are established between nodes from different communities, in this case, the resulting network is scale-free without a modular structure.

In figure 28 we can appreciate the effect of the mixing parameter in the LFR networks. With low values of μ we have sparse links between communities. As we increase the value of the mixing parameter we see an increase in inter-community links, and for the $\mu = 0.8$ we have an almost fully mixed structure.

6.3 Results on SIR non-Markovian process with Weibull infection in modular networks

First, just as in the last section, let us explore some specific cases to get a general notion of how this kind of epidemic behaves in modular networks. For this initial analysis, we will look at a small LFR network with $N = 300$, $\langle k \rangle = 7$, $\Upsilon = 2.5$, $\Omega = 1.5$, and $\mu = 0.05$. With these parameters, we generate a modular network with 6 communities. We will perform an epidemic simulation with $\lambda_{eff} = 0.75$, $\gamma = 1$. In figure 29 we first compare the Markov case with $\alpha = 1$ and with a negative aging $\alpha = 0.75$.

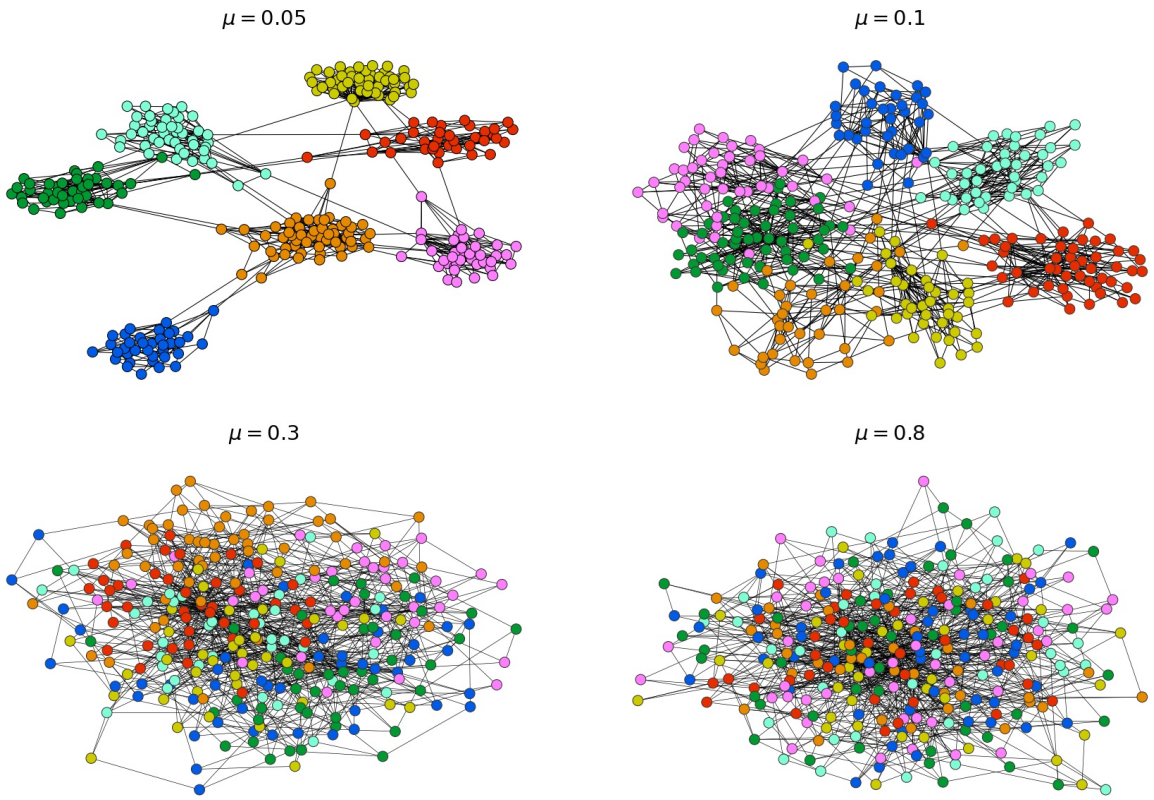

Figure 28 – Visual representation of synthetic modular networks generated by the LFR algorithm. Each network has a different mixing parameter.

Source: Elaborated by the author.

In figure 29, we see three-time instances $t = [0.5, 1, 2.5]$ for two identical networks with different epidemic processes, one with no aging (second column) and the other with a negative aging (first column). The first striking difference is that in negative aging, the disease has a faster dispersion from one community to another. We can see this comparing time $t = 1$, for which in case $\alpha = 0.75$ all communities have infected nodes, and three of them have more the half of the population infected, on the contrary for $\alpha = 1$ there are still 3 completely healthy communities. For time $t = 2.5$ almost all individuals have been infected and recovered for the negative aging, whereas for the traditional Markov case, the system is still at the peak number of infected nodes.

Now let us compare another case, but this time we will use an epidemic with no aging ($\alpha = 1$) and an epidemic with positive aging $\alpha = 2$. We see these results in figure 30

The positive aging process has the opposite effect of negative aging. In this case of positive aging, the probability of infection is low at early times and gets higher as time passes. In this specific case, this causes the disease takes a longer time to reach healthy communities. At time $t = 2.5$, for the Markov case with no aging the disease has reached all communities and

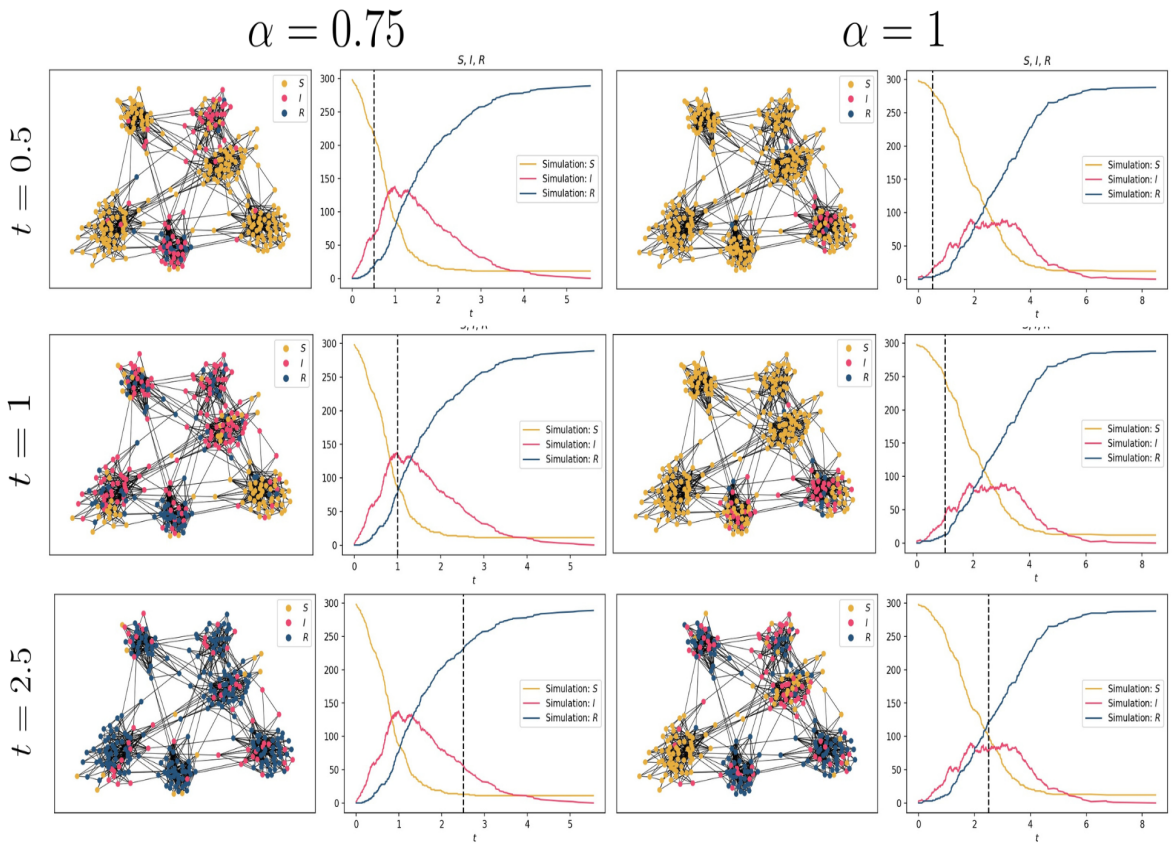

Figure 29 – Compression between two epidemic models in the same modular network, but with different shape parameters. The LFR network generated has 300 nodes and an average degree of 6. Each row represents a different instant in time, and each column has the network representation and dynamical curves for different shape parameters: a negative aging of $\alpha = 0.75$ in the first column, and the Markovian case with $\alpha = 1$ in the second column.

Source: Elaborated by the author.

the epidemic is near the maximum number of simultaneous infected individuals. However for the positive aging with $\alpha = 2$ at the same time, $t = 2.5$ half of the communities are completely healthy. For time $t = 5$ for $\alpha = 1$ we see almost all individuals recovered from the disease, whereas for $\alpha = 2$ many individuals are still susceptible, and infected individuals are starting to decrease, indicating that probably a big fraction of individuals will not be reached by the disease.

These examples are merely illustrative but can give us a quick perspective of the effect of the aging process in the LFR networks. It appears that the main difference between the compared cases is in how quickly the disease is able to move from one community to the other. In the next subsection, we will explore this idea in a more general way.

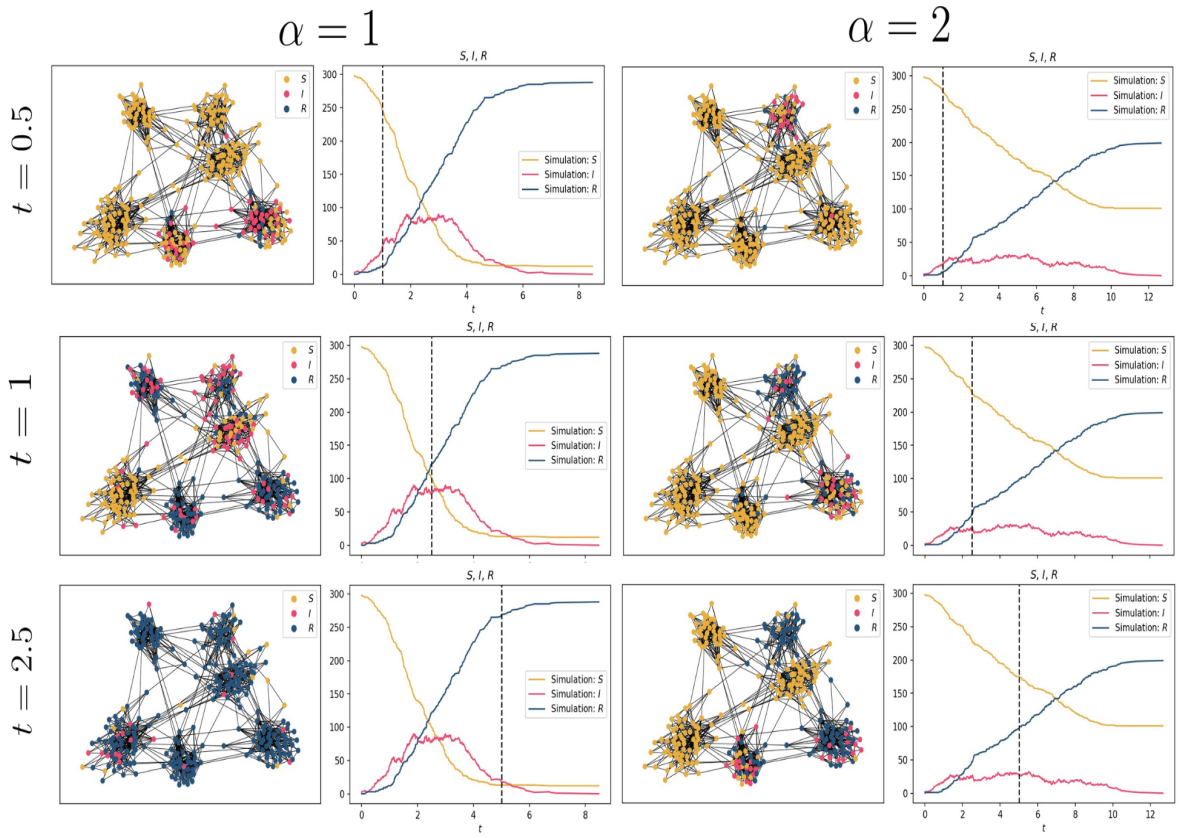

Figure 30 – Compression between two epidemic models in the same modular network, but with different shape parameters. The LFR network generated has 300 nodes and an average degree of 6. Each row represents a different instant in time, and each column has the network representation and dynamical curves for different shape parameters: the Markovian case with $\alpha = 1$ in the first column, and a positive aging with $\alpha = 2$ in the second column.

Source: Elaborated by the author.

6.3.1 Aging process delays the spread of the disease from communities

At the beginning of the epidemic, a small percentage of individuals are randomly selected to be initially infected at time zero. Since these initial infections occur randomly, the disease will begin spreading in a certain number of communities, denoted as M_0. Over time, the infection will gradually extend to the remaining communities. To assess the impact of aging on the rate of infection spread to healthy communities, we introduce the time T_M as the point in time when the disease has infected at least one individual in each community. By measuring T_M, we can have a notion of how the aging process affects the speed at which infections reach previously uninfected communities.

We consider an LFR network with $N = 7500$, $\Upsilon = 2.5$, $\Omega = 1.5$, and $\langle k \rangle = 6$. With these conditions, the resulting networks end with approximately 50 communities. Figure 31 illustrates the values of T_M as a function of the shape parameter for networks with different values of μ.

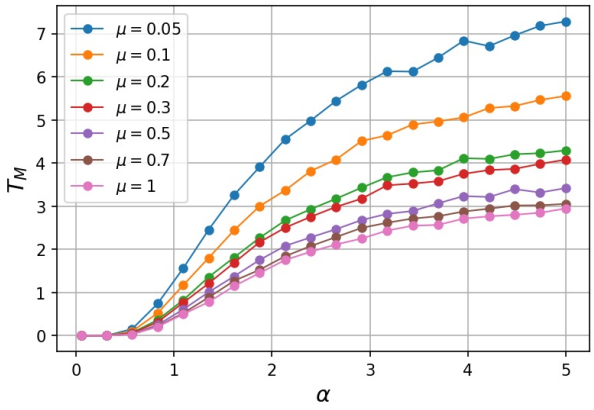

Figure 31 – Values of T_M, time until all communities have at least one infected individual, as a function of the shape parameter *alpha*. Each color represents a different mixing parameter from 0.05 to 1. Each point is calculated as the average time of 100 simulations.

It is observed that the values of T_M increase as the shape parameter increases for all curves. This behavior indicates that the disease takes longer time to reach healthy communities. When weaker community structures are present with higher values of μ, α still manages to introduce a delay in the disease's spread across communities, but the effect is relatively less pronounced. This phenomenon can be attributed to the aging process slowing down the spreading dynamics, thereby delaying the activation of inter-community links in networks with community structure.

It's essential to acknowledge that this analysis provides valuable insights into how aging can potentially influence the dynamics of an epidemic. However, it is not comprehensive enough to fully understand the overall impact of Weibull infection. To gain a more complete understanding, further investigation is required.

Now, we will turn our attention to the prevalence and epidemic life, as they are commonly used variables that help to measure the impact of an epidemic in a population.

6.3.2 Prevalence of the epidemic.

For the SIR epidemic process, we will define the prevalence, also called epidemic size, as the relative number of recovered individuals after the end of the simulation, which occurs when there are no more infected individuals in the population.

To understand how this variable is affected by the modular characteristic of the network and the aging associated with the non-Markvoian epidemic we perform extensive simulation varying the mixing parameter μ and the shape parameter α. The simulations were performed on a network with $N = 7500$, $\Upsilon = 2.5$, $\Omega = 1.5$, and $\langle k \rangle = 6$. With this condition, we obtain networks with an average of 50 communities. We perform 400 simulations for different values of α, μ, and λ_{eff} and calculate the average number of recovered people at the end of the simulation. The results are presented in figure 32 which has 6 heat maps, each one with different values of effective rate.

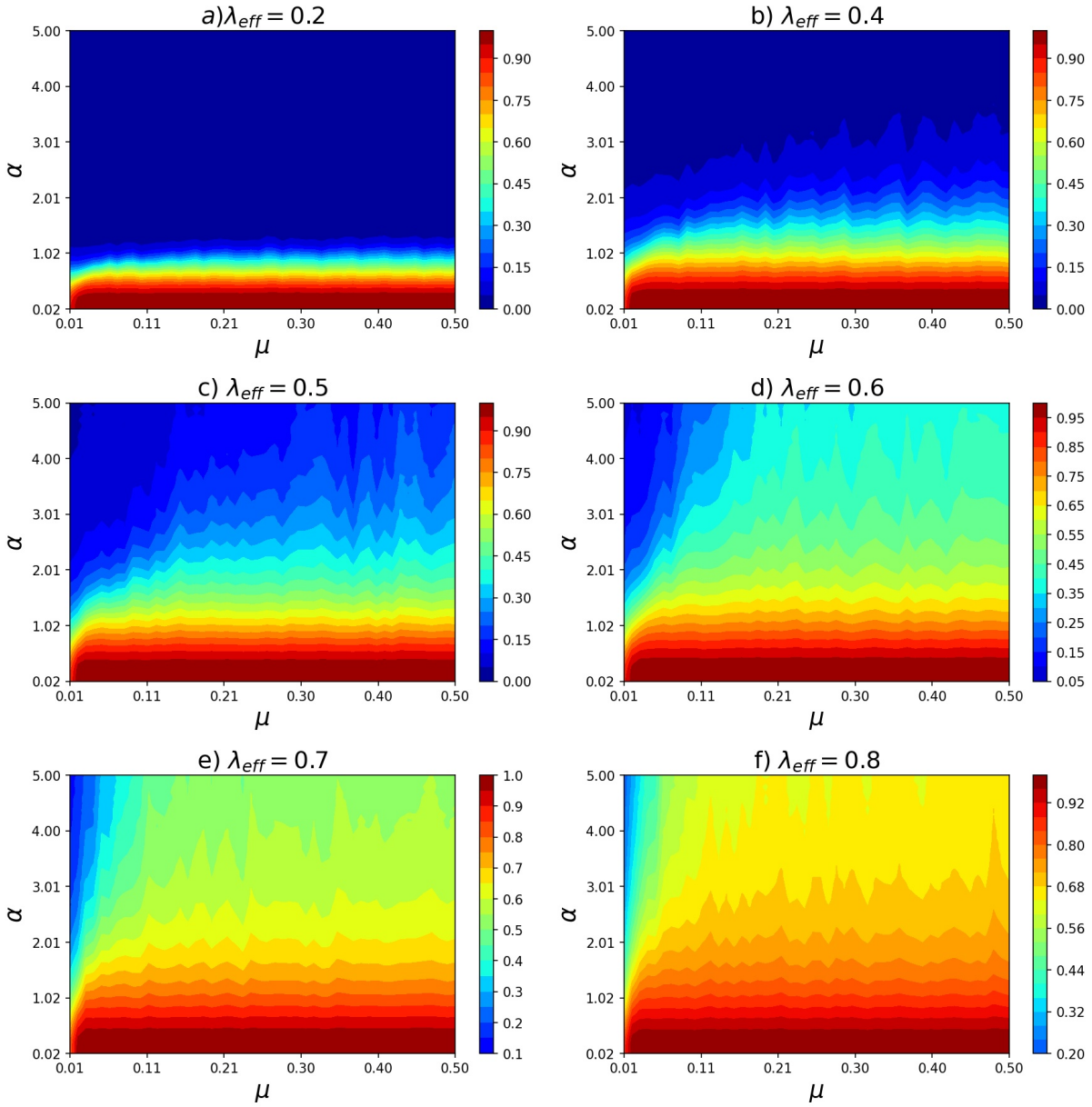

Figure 32 – Heat maps generated by a SIR process with Weibull infection on a modular network. The colors represent the prevalence found for each pair values of shape and mixing parameter. Each map corresponds to a shape parameter: a) $\lambda_{eff} = 0.2$, b) $\lambda_{eff} = 0.4$, c) $\lambda_{eff} = 0.5$, d)$\lambda_{eff} = 0.6$, e)$\lambda_{eff} = 0.7$, f)$\lambda_{eff} = 0.8$.

For the first two maps, 32a and 32b, we used smaller values of the effective transition rate. In these cases, most values of α above 1, in 32a, and values of α larger 2, in 32b, have prevalence values near zero. This indicates that the change in the critical effective ratio associated with the α parameter put the system in the absorbing state. For the values of α that result in an active state, the change in prevalence is primarily in the vertical direction, which indicates that the community structure does not play an important role in changing the final epidemic size.

As we increase the effective transition rate λ_{eff} the absorbing region shrinks and we see some more interesting behaviors. In case 32c (with $\lambda_{eff} = 0.6$), the shape parameter continues to dominate the changes in prevalence. As α increases, the fraction of the population affected by the epidemic process reduces. While we observe a subtle effect of the mixing parameter for higher values of α, it becomes more evident in cases 32d, 32e, and 32f, where we have values of λ_{eff} well above the critical effective rate. For these latter cases, when the mixing parameter is higher than 0.3, the influence of the shape parameter still remains the primary driver of changes in prevalence. However, the scenario changes when the mixing parameter decreases, especially for $\mu < 0.3$. In such situations, the effect of the community structure becomes evident and a notable reduction in prevalence values is observed. In other words, the μ parameter appears to be more relevant for higher shape parameters, indicating that the community structure's influence becomes more apparent.

In a summary, the prevalence of the epidemic consistently decreases with an increase in the shape parameter and a decrease in the mixing parameter. Although the non-Markov effect appears to exert a more substantial influence on the final epidemic size, strong community structures (indicated by small values of μ) still play a crucial when dealing with higher shape parameters. To gain a more comprehensive understanding of the epidemic dynamics, our focus now shifts to examining the epidemic life of the system.

6.3.3 Epidemic life

In the SIR epidemic model nodes that recovered from the disease can not get infected again, so the epidemic will always die after a certain amount of time, with the extreme case where all individuals of the network get infected and recovered. In this process, we define the epidemic life T_l as the time until there are no more infected individuals in the system, at which we say that the epidemic "dies".

We perform 400 simulations for different values of α, μ, and λ_{eff} and calculate the average number of epidemic life of the simulation. The results are presented in figure 33 which has 6 heat maps, each one with different values of effective rate.

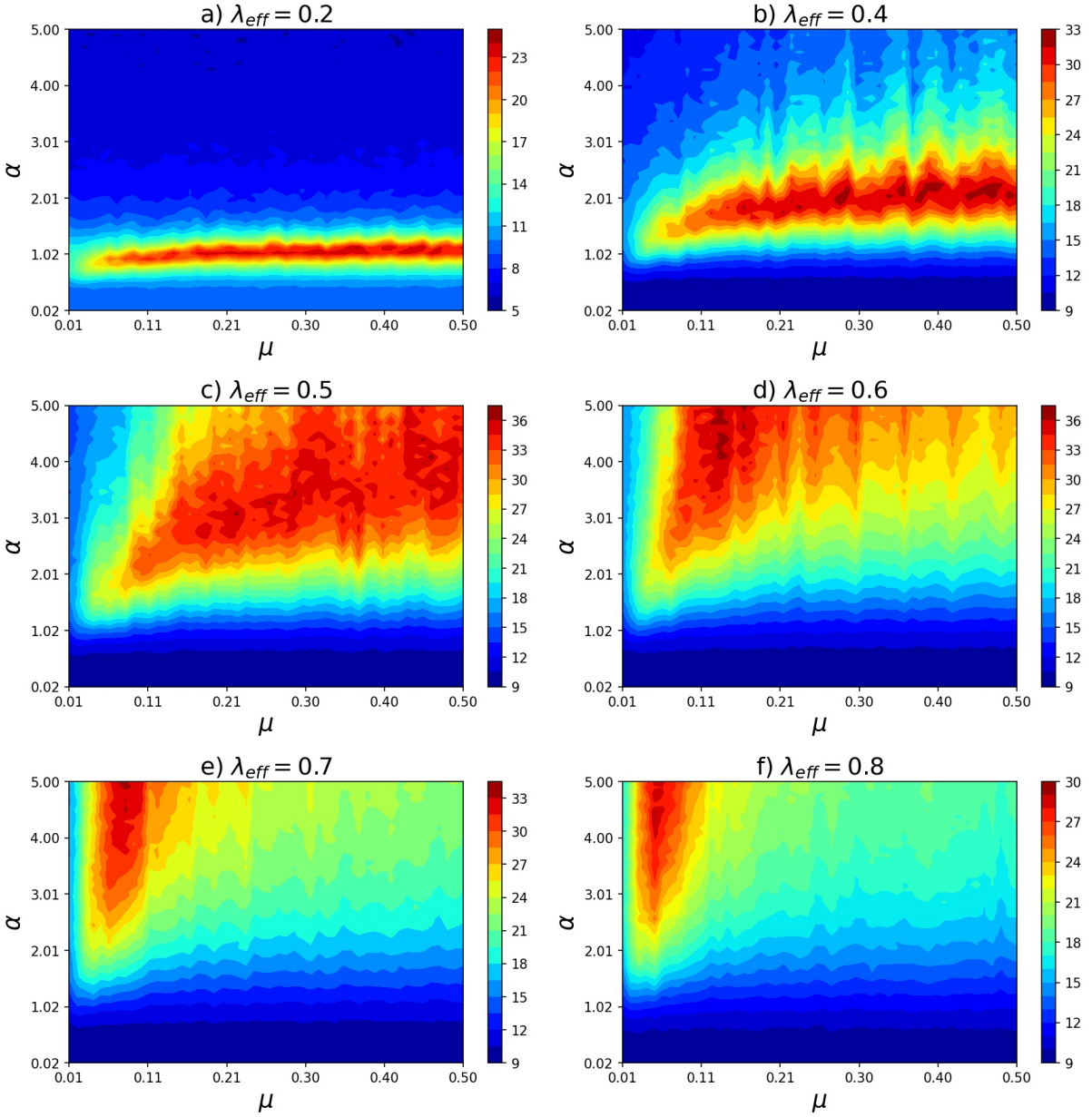

Figure 33 – Heat maps generated by a SIR process with Weibull infection on a modular network. The colors represent the life of the epidemic found for each pair values of shape and mixing parameter. Each map corresponds to a shape parameter: a) $\lambda_{eff} = 0.2$, b) $\lambda_{eff} = 0.4$, c) $\lambda_{eff} = 0.5$, d)$\lambda_{eff} = 0.6$, e)$\lambda_{eff} = 0.7$, f)$\lambda_{eff} = 0.8$

Again, we have to be careful with the case 33a and 33b as λ_{eff} is low and won't be above the critical transition rate for all values of α. Short epidemic lives could either be because there was a quick propagation of the epidemic or if the system is below the critical effective rate because the epidemic dies before spreading through the entire network. It is important to

compare this map to the maps in 32. In general, if there is a short epidemic life but the prevalence value is high, then we are dealing with a quick epidemic that spreads through the network. On the other hand, if the epidemic life is short and prevalence is low, then the system probably is below the critical effective rate and has an epidemic that dies out very quickly.

With this consideration, we can interpret results in figure 33. For the negative aging process $\alpha < 1$, almost all simulations have a relatively short epidemic life, and the μ parameter does not play a relevant role as there is basically no change as μ increases. Comparing with 32 we see that in this region there are high prevalence values, so we are dealing with a disease that spreads very quickly regardless of the community structure.

For positive aging processes $\alpha > 1$, we see a more interesting behavior. In the case of meat maps 33c and d33d the largest epidemic lives occur for higher values of α. In this case, we see a change with respect to μ. For very low values of the mixing parameter epidemic lives are short, and as μ increases we see that the epidemic lives get longer and longer until a maximum, then return to shorter lives. The interpretation of this behavior is as follows: for very strong community structures and positive aging the epidemic lives are shorter and prevalence values also are small, this probably occurs, as we saw in section 6.3.1, positive aging can cause slower propagation through each community, and for very strong community structure this can lead to a fast infection that quickly dies out before reaching healthy communities. As the communities start to mix with the increase in μ the positive aging is not enough to contain the disease inside the infected communities, we see this as the prevalence values increase with μ in this region. Despite this, longer epidemic lives are found, as it takes more for each individual to infect their neighbors.

The effect described above is more prevalent in the case e33 and f33 where the peak values of the epidemic life occur for very small mixing parameters. In this peak, the epidemic reaches a considerable prevalence value (near 0.4) but the epidemic has large values of T_l as the epidemic has a slow propagation through the network. After the peak in T_l as μ increases shorter epidemic lives appear again, which further indicates that communities are not just not able to contain the epidemic but they don't meaningful effect in the epidemic process.

To this point, we have explored the general effect of the μ and α parameters in the epidemic process. But these have been just general observations of the effect of each variable in the process. Now, we have to further understand how the aging associated with the non-

Markvoainity can change the role of communities.

6.3.4 Difference in prevalence

In section 6.3.2 we presented the influence of the shape parameter and mixing parameter in the prevalence of the epidemic in LFR networks. The results suggested that the mixing parameter has a more significant role for higher values of α. To further explore this idea we will present another way to study non-Markovian epidemic in modular networks. We can make a plot of the epidemic size as a function of the shape parameter for different networks with different community structures. We set $\lambda_{eff} = 0.5$ and simulated SIR processes in two networks: one with a strong community structure ($\mu = 0.05$) and another with completely mixed communities ($\mu = 1$). The other parameters of the network are the same as the one described in previous section 6.3.2. The results of this simulations are shown in figure 34.

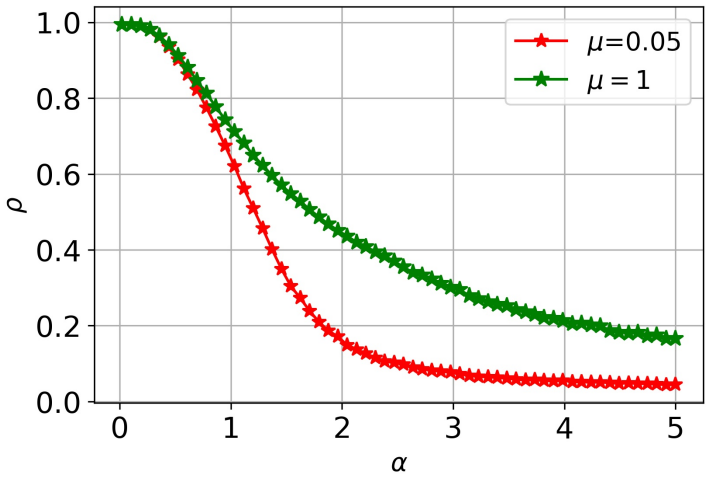

Figure 34 – Two curves of prevalence values as a function of the shape parameter of SIR Weibull process in LFR networks. The green curve are the ρ values of the simulation in an LFR network with mixing parameter $\mu = 1$, and the red curve on a network with $\mu = 0.05$.

For small values of α the two curves are nearly identical. However, as we increase the shape parameter, the gap between the curves gradually widens until it reaches a maximum, after which it begins to contract. This compression is valuable because we are comparing epidemic processes with the same characteristics, but one in a network with a strong community structure while the other one is in a network with very weak community structure. This analysis might help to identify when the community structure plays a pivotal role in the spreading process.

Inspired by this observation, we introduce an intuitive variable called the "difference in prevalence". This variable quantifies the disparity in prevalence values between two epidemic

processes with the same α parameter and λ_{eff} but distinct mixing parameters, μ_1 and μ_2, where the relationship holds that $\mu_1 > \mu_2$. The formal definition of this variable is presented in equation (6.1).

$$\Delta\rho = \rho(\alpha, \lambda_{eff}, \mu_1) - \rho(\alpha, \lambda_{eff}, \mu_2) \tag{6.1}$$

From this definition we can calculate the difference in prevalence for the curves in figure 34, which is shown in figure 35.

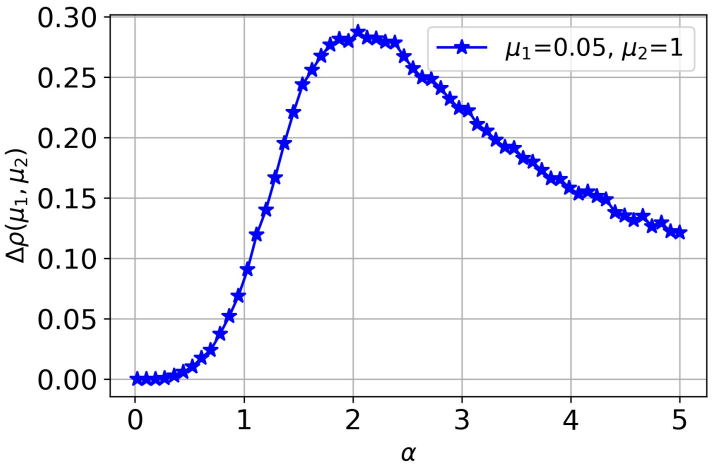

Figure 35 – Difference in prevalence of the curves of image 34.

It is essential to emphasize that the mixing parameter, μ, is a user-defined input for the LFR algorithm, allowing us to set it to any desired value. For the purpose of the following analyses in this subsection, we will keep μ_1 fixed at 1 while varying μ_2. This approach enables us to compare networks with different community structures against networks with mixed communities. With this consideration, we will call $\mu_2 = \mu$ and the difference in prevalence will be: $\Delta\rho = \rho(\alpha, \lambda_{eff}, 1) - \rho(\alpha, \lambda_{eff}, \mu)$.

With the introduction of this new variable, our first objective is to comprehend how the parameter α influences the "difference in prevalence." To achieve this, we conducted simulations using various values of the effective transition rate while varying the shape parameter. The difference was calculated different network structures. Figure 36 illustrates the results of these simulations, where each color on the plot represents a distinct value of the mixing parameter μ.

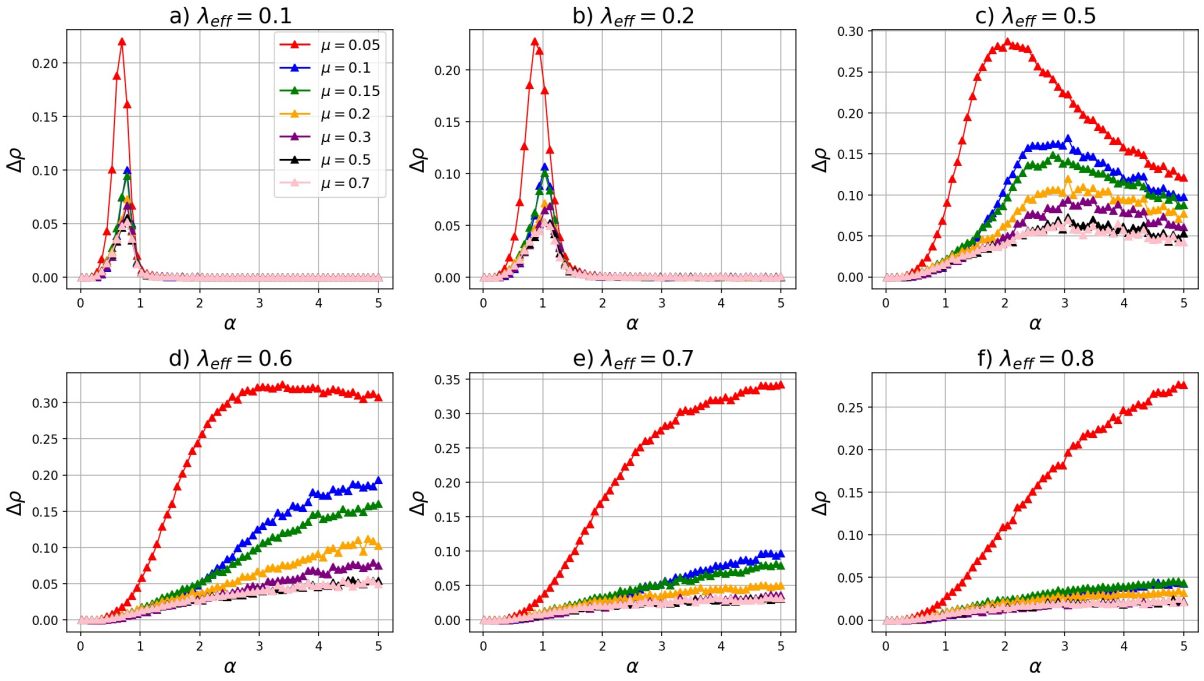

Figure 36 – Difference in prevalence as a function of the shape parameter α for different networks with community structures. Each plot corresponds to a different effective rate: a) $\lambda_{eff} = 0.1$, b) $\lambda_{eff} = 0.2$, c) $\lambda_{eff} = 0.5$, d) $\lambda_{eff} = 0.6$, e) $\lambda_{eff} = 0.7$, and f) $\lambda_{eff} = 0.8$. Each color correspond to a different modular structure: red $\mu = 0.1$, blue $\mu = 0.2$, green $\mu = 0.3$, orange $\mu = 0.5$, black $\mu = 0.7$.

Within the considered range of α, we observe two discernible types of behavior depending on the value of λ_{eff}. Figures 36a, 36b, 36c, and 36d, there exists a prominent peak in the difference in prevalence for all community structures at a specific value of the shape parameter. However, as the mixing parameter μ increases and we generate networks with weaker community structures, this peak gradually diminishes. This observation aligns with our expectations, as we approach μ closer to 1.

The presence of a peak in the difference in prevalence suggests that there is an optimal value of the shape parameter, denoted as α_{peak}, at which the impact of the community structure on prevalence differences is maximized.

We might explain the peak behavior as follows. For small shape parameter (indicating very negative aging) infected nodes have a high probability of quickly infecting their neighbors during the early stages of the infection. This rapid propagation reduces the impact of the community structure, as the disease spreads rapidly across the network. As the shape parameter α increases, the probability of infection at later times also increases. Consequently, the community structure begins to play a more significant role in the spread of the epidemic. As discussed in section 6.3.1, with higher shape parameters, it takes more time for the disease to escape a

community and infect healthy nodes. Thus, the community boundaries act as barriers to the epidemic's propagation. However, it's crucial to consider that as the shape parameter increases, the critical rate λ_c, which separates the absorbing from the active state of the epidemic, also rises [Mieghem and Bovenkamp 2013]. Larger shape parameters lead to a return to the absorbing phase as the effective rate falls below the threshold. Consequently, after the peak in the difference in prevalence, both networks, with and without community structure, revert to the absorbing phase, where there is no outbreak in the first place. As a result of these two effects, there exists a shape parameter that maximizes the difference in prevalence, indicating a stronger role for community structure in the propagation. However, it is important to note that the existence of a peak in the difference in prevalence does not solely arise from the relationship between the community structure and the aging process.

Another scenario arises for larger effective rates, as depicted in figures (36e and 36f), where the critical rates for all α values within this range fall below λ_{eff}. In such circumstances, increasing the shape parameter leads to a slower propagation of the epidemic due to the aging process. As a result, community structures play a more significant role in reducing the prevalence of the epidemic. The peak behavior might be observed in larger ranges of α, as at some point the critical rate will be above the effective rate chosen.

Another thing that can be highlighted from figure 36 is that the difference in prevalence of the higher mixing parameter ($\mu = 0.5$,$\mu = 0.3$ and $\mu = 0.7$) have very similar values. On the contrary the lower values of the mixing parameters are considerably larger. To have another perspective on this behavior made a similar analysis as the one described in figure 36, but using μ as the x-axes and choosing some values of α. The result of this analysis is presented in Figure 37, which shows curves of difference in prevalence as a function of μ with a logarithmic scale in the x-axes to facilitate interpretation.

In all six plots, we observe a rapid decrease in the difference in prevalence, $\Delta\rho$, for mixing parameters greater than 0.1. In most cases, if $\mu > 0.3$, the difference in prevalence becomes close to zero, indicating that running the epidemic process in networks with and without community structure yields similar final epidemic sizes. This effect is particularly evident for larger values of effective transition rates, where the decline after $\mu = 0.1$ is more pronounced. This phenomenon can be interpreted as the communities becoming sufficiently mixed beyond a certain value of μ, such that they no longer retain the epidemic. Beyond this point, further increases in the mixing parameter do not significantly alter the resulting epidemic process.

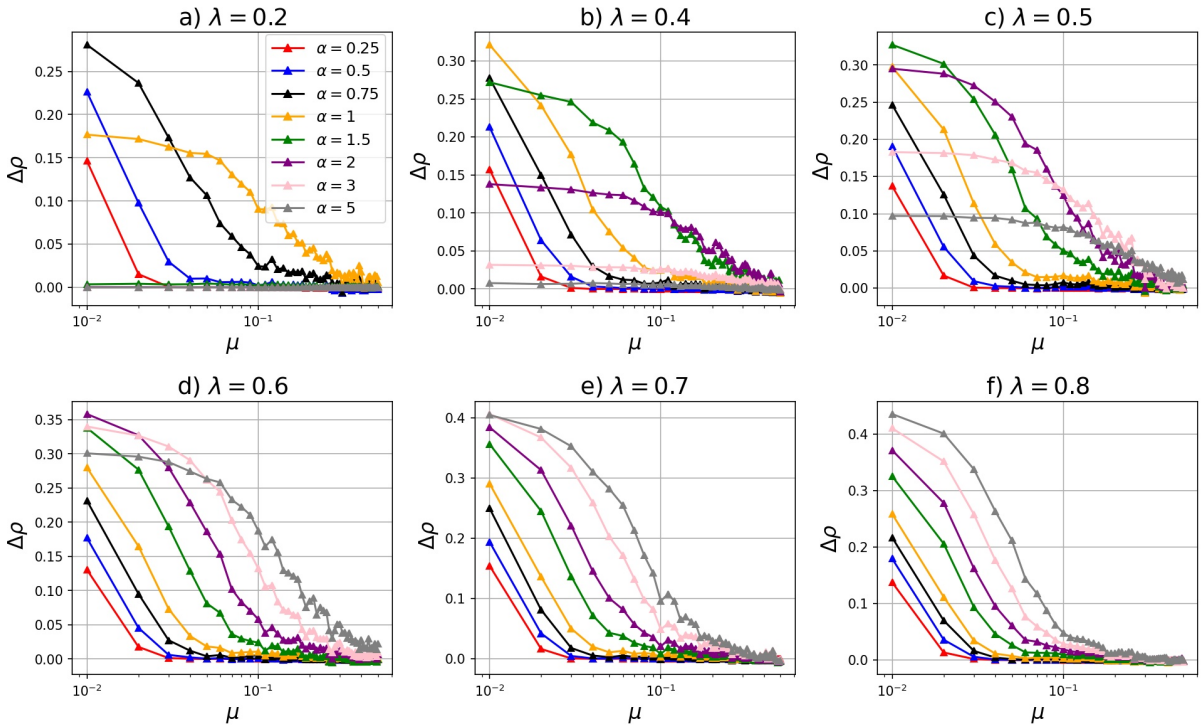

Figure 37 – Difference in prevalence as a function of the mixing parameter μ for different aging processes. Each plot corresponds to a different effective rate: a) $\lambda_{eff} = 0.2$, b) $\lambda_{eff} = 0.4$, c) $\lambda_{eff} = 0.5$, d) $\lambda_{eff} = 0.6$, e) $\lambda_{eff} = 0.7$, and f) $\lambda_{eff} = 0.8$

. Each color correspond to value of α associated with the aging process: red $\alpha = 0.25$, blue $\alpha = 0.5$, black $\alpha = 0.75$, orange $\alpha = 1$, green $\alpha = 1.5$, purple $\alpha = 2$, pink $\alpha = 3$, gray $\alpha = 5$.

When $\mu < 0.1$, we observe significant changes in the difference in prevalence for different values of α. In case 37a, curves with higher shape parameters (1.5, 2, 3, and 5) show a difference in prevalence of zero for all values of μ, it is essential to note that this result occurs because the critical values associated with each shape parameter are higher than the chosen effective rate. As a result, both networks remain in the absorbing phase, where the epidemic does not affect a significant portion of individuals. The same phenomenon is observed in case 37b, but only for $\alpha = 5$. These cases are not informative as there is no epidemic occurrence in the first place. These cases are not interesting as there is not an epidemic in the first place.

Now, shifting our focus to Figure 37c, we find interesting observations. For smaller shape parameters (0.25, 0.5, 0.75, 1, and 1.5), increasing α enhances the role of the network structure, with curves having higher shape parameters positioned above those with smaller α. However, for shape parameters 2, 3, and 5, higher shape parameter curves are positioned below those with smaller α. This may initially appear confusing, but it reflects the same phenomenon as depicted in the previous section. For smaller effective rates, a peak behavior is observed concerning α, where curves with higher shape parameters are initially positioned above the other curves (α

from 0.25 to 1.5). However, after passing the peak, curves with higher α appear below the others (α from 2 to 5). This same behavior is also observed in curves a, b, and d in Figure 37, but the specific values of α change as the peak position shifts, similar to what was observed in Figure 36.

In figures 37e and 37f, we find ourselves sufficiently distant from the critical transition rate associated with each α value. As a result, increasing the shape parameter α significantly magnifies the effect of the mixing parameter μ. In these cases, the curves consistently remain positioned above the other curves with smaller shape parameter values.

These results, and those presented in the previous section were carried in networks with different sizes and higher average degrees, and the results obtained are qualitatively the same as in this case.

6.3.5 Effect of the Number o communities

So far we have just worked with networks with a number of communities close to $M = 50$. The LFR algorithm does not allow us to predetermine the number of communities, but this number can be altered by changing the s_{max} and s_{min} described in section 6.2.

The main goal of this project is not to study the effect of the number of communities. However, we can make a quick observation about the number of communities, especially to further support our choice of $M = 50$. In figure 38 we show some curves of prevalence as a function of the shape parameter α for different values of M, in the range

$$[10, 60]$$

in two structures: a weak one ($\mu = 0.7$) and a strong one ($\mu = 0.05$).

For figure 38 we see the prevalence curves for networks with weak community structure. We observe that the number of communities does not affect the prevalence curves as there is no particular tendency and curves appear to be very close to each other. However, for the strong community structure ($\mu = 0.05$), in figure 38b, as the number of communities M increases the prevalence curves have lower values. This means that as the number of communities increases, at least in this range of M, the community structure is more successful in containing the disease spread. Another important detail from figure 38b is that all curves of prevalence with strong community structure are below the prevalence of the weak community structure for $M = 60$ (black curve), independent on the number of communities.

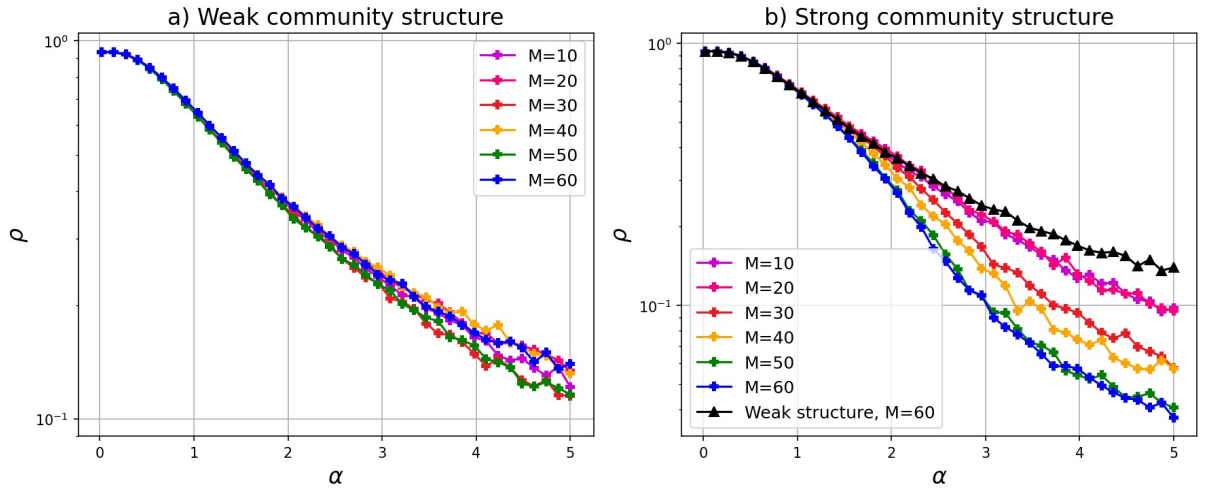

Figure 38 – Prevalence of the epidemic in log scale as a function of the shape parameters for LFR networks with a different number of communities in the range $[10, 60]$

An important consideration for this analysis is that it might change for a larger range of M or other specifications of the network. However, it is an interesting result as it helps us understand this specific case.

6.4 Simulations on real-world networks

To further support the behavior described in section 6.3.4, we conducted simulations of SIR non-Markovian epidemic processes on networks derived from real databases. The objective of this analysis was to verify if the observed properties can be seen in modular networks with different characteristics, specifically those not generated by the LFR algorithm.

This analysis focused on three real-world networks with known modular structures, obtained from diverse sources. The networks under consideration were:

- A Facebook network (F), represent social network interaction.
- A Contact network (C), derived from human interactions.
- An animal network (A) depicting voles' interactions.

Table 6.4 provides more detail descriptions of these networks and highlights other important properties associated with them.

Network	Description	N (Number of nodes)	V(Number of edged)	Average degree $\langle k \rangle$	Mixing parameter μ
Facebook network(F)	Composed of social circles of Facebook. Data was collected from a survey where Facebook users IDs where change to new valus to preser privacy. The data was extracted from the Stanford Network Analysis Platform (SNAP) [Leskovec and Krevl 2014] and has been utilized in previous works on epidemics in modular networks [Ghalmane *et al.* 2019].	4039	88234	43.69	0.04
Contact network (C)	This date describes face-to-face interactions attendees of exhibition called "INFECTIOUS: STAY AWAY" that took place in at the Science Gallery in Dublin in 2009. Nodes represent visitors, and a link is established if the visitor interacted for at least 20 seconds. The original data has weighted links depending how much time the interaction lasted. However, for this analysis we will leave this information out, as we intend just to look for the role of modular structure. The source of this network was the SocioPatterns data sets [Infectious network dataset – KONECT 2017].	411	2766	13.46	0.11
Animal (A)	In this case, the data is from vole interactions. Each node is a vole and an edge between two voles exists if they were captured in in the same trap at least ones during a trapping session. This data was extracted from the Network repository project [Rossi and Ahmed 2015].	1189	3576	6.02	0.08

Table 1 – Table with the three real network data sets with the number of nodes (N), number of edges (V), average degree ($\langle k \rangle$), and mixing parameter (μ, of each network.)

A parameter of particular interest in our analysis is the mixing parameter μ, as it serves as a crucial descriptor of the network's community structure. In this context, calculating the mixing parameter required several steps. First, we employed the Louvain algorithm for community detection to identify the distinct communities within the network. Once the communities were identified, we proceeded to compute the mixing parameter μ. This parameter was determined as the average was calculated as the average fraction of links between nodes of different communities and total number of links6.2.

Furthermore, the degree distributions of these networks are depicted in figure 39. It is

evident that both the Facebook network and the Animal network exhibit clear degree distributions that approximate a power law. On the other hand, the contact network has a slightly more complex degree distribution but still approximates better to a power-law distribution than the distributions observed in regular networks.

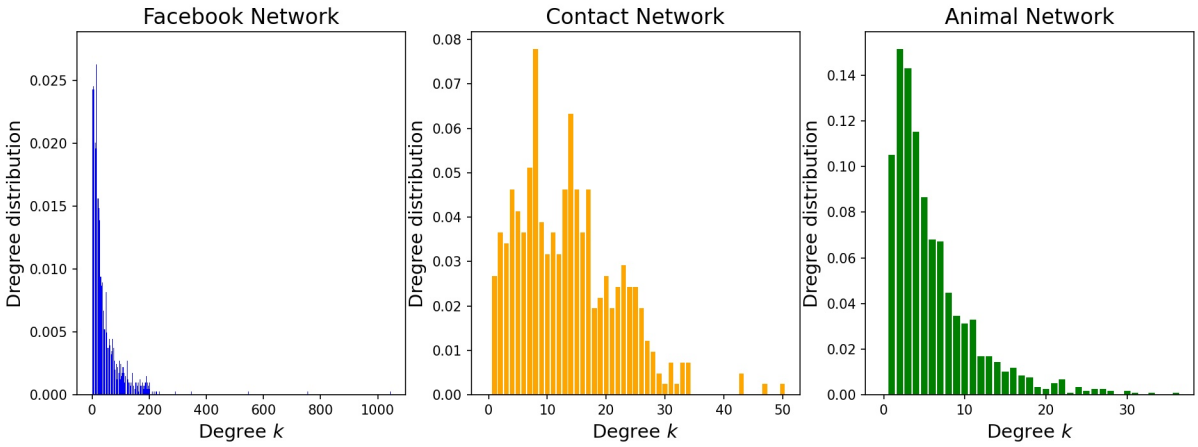

Figure 39 – Degree distribution of the real networks used.

Also, we show a visual representation of the networks in figure 40.

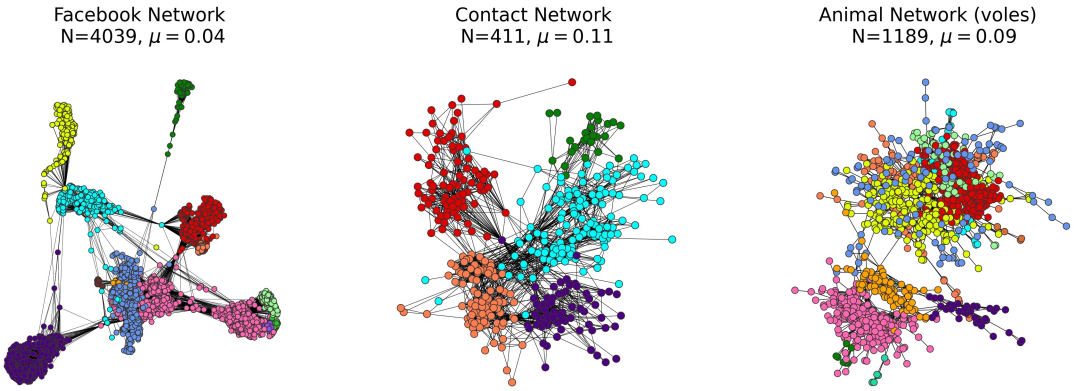

Figure 40 – Visual representation of the three networks created by real data: Facebook network, a Contact network and an Animal network.

Now if we would just compare the non-Markovian epidemic process in these three networks the results would be difficult to analyses as the three cases present very different properties. For instance the Facebook network has e much higher degree distribution compared to the animal network, so choosing the adequate effective transition rate would be difficult.

To address the challenge of obtaining data sets with networks that have similar topological properties but varying mixing parameters, performing a swapping process on existing networks

proves to be a viable solution. The swapping process involves randomly selecting pairs of links in the network and exchanging their connections. For example, if we have links (u_1, v_1) and (u_2, v_2), the swapping process would entail removing these links and establishing new connections (u_1, u_2) and (v_1, v_2). The swapping process effectively weakens the community structure gradually without altering other essential properties of the system, such as average degree and degree distribution. This is achieved by maintaining the same number of links and only changing the connections between nodes. As many links are swapped, the probability of choosing two links between the same communities becomes rare, leading to a gradual breakdown of the community structure. A representation of this process can be observed in figure 41 where a plot of the mixing parameter as a function of the number of swaps made in the Animal network is shown.

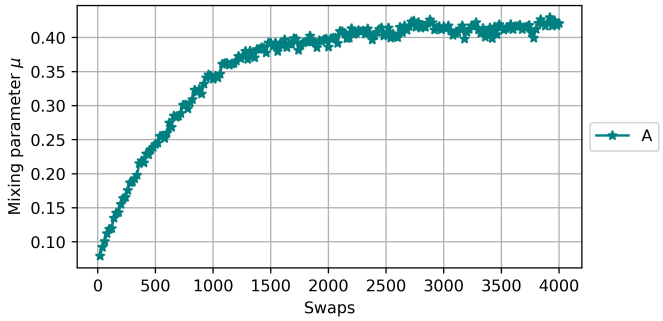

Figure 41 – Mixing parameter calculated of the A network as function of the number of swaps made in the network.Each color represents a different community.

Through the swapping process, it is observed that the mixing parameter increases as more swaps are performed, indicating a progressive mixing of the communities in the network. Additionally, it is worth noting that while the mixing parameter eventually reaches a plateau, it does not necessarily converge to $\mu = 1$.

For each network, using this swapping, we created four additional networks for each case. We will called these new networks as “swapped networks”. We use the same letters as the original networks with sub-index to identify each swapped networks, as the index increases more swaps were performed. For example A_1 and A_2 are swapped networks created from A, and A_2 suffered more swaps than A_1. Table 2 presents the names and some information of the swapped networks.

New network	Description	μ
F_1	F after 250 swaps	0.05
F_2	F after 500 swaps	0.05
F_3	F after 1259 swaps	0.06
F_4	F after 2500 swaps	0.1
C_1	C after 50 swaps	0.11
C_2	C after 100 swaps	0.21
C_3	C after 250 swaps	0.36
C_4	C after 500 swaps	0.51
A_1	A after 100 swaps	0.12
A_2	A after 200 swaps	0.16
A_3	A after 500 swaps	0.25
A_4	A after 1000 swaps	0.35

Table 2 – Information and names of the new networks generated after executing swaps in the original real networks F, C and A.

For each case different number of swaps was performed because the size of the network influence the number of swaps required to have a meaningful change in the mixing parameter. For example, the Facebook networks consist of 4039 nodes, necessitating a greater number of swaps to achieve a significant change in the mixing parameter compared to the smaller contact network.

Let X be a generic network that could be any of the original cases F, C, or A, and its respecting swapped networks X_1, X_2, X_3 and X_4.

X and X_4 exhibit highly similar characteristics as they share the same average degree and degree distribution. However, the community structure of X_4 is weaker due to the increased mixing parameter. In this context, defining the same "difference in prevalence" as observed in synthetic networks is not straightforward since the swapping process does not result in a network with a mixing parameter equal to 1.

We will return to original definition of the difference in prevalence: $\Delta\rho = \rho(\alpha, \lambda_{eff}, \mu_1) - \rho(\alpha, \lambda_{eff}, \mu_2)$. In this definition, μ_1 and μ_2 can be chosen as any mixing parameter of two networks. Accordingly, we will utilize μ_1 from the equation as the mixing parameter of X_4, as it represents the network with the larger mixing parameter. Hence, the difference in prevalence for real networks will be defined as $\Delta\rho = \rho(\alpha, \lambda_{eff}, \mu_{X_4}) - \rho(\alpha, \lambda_{eff}, \mu)$. In this context, μ_{X_4} represents the mixing parameter of X_4, and μ will be the mixing parameter of the original network X or one of the three other swapped networks: X_1, X_2, or X_3.

6.4.0.1 Results of difference in prevalence for real networks

We conducted simulations the fifteen networks, consisting of the original network and four swapped networks for each of the three cases. For each network, we calculated the prevalence of X_4 while considering both the original network and the three other swapped networks. We performed these simulations across various values of the effective transition rate λ_{eff} and by varying the shape parameter of the Weibull infection. The results of these simulations are presented in figure 42.

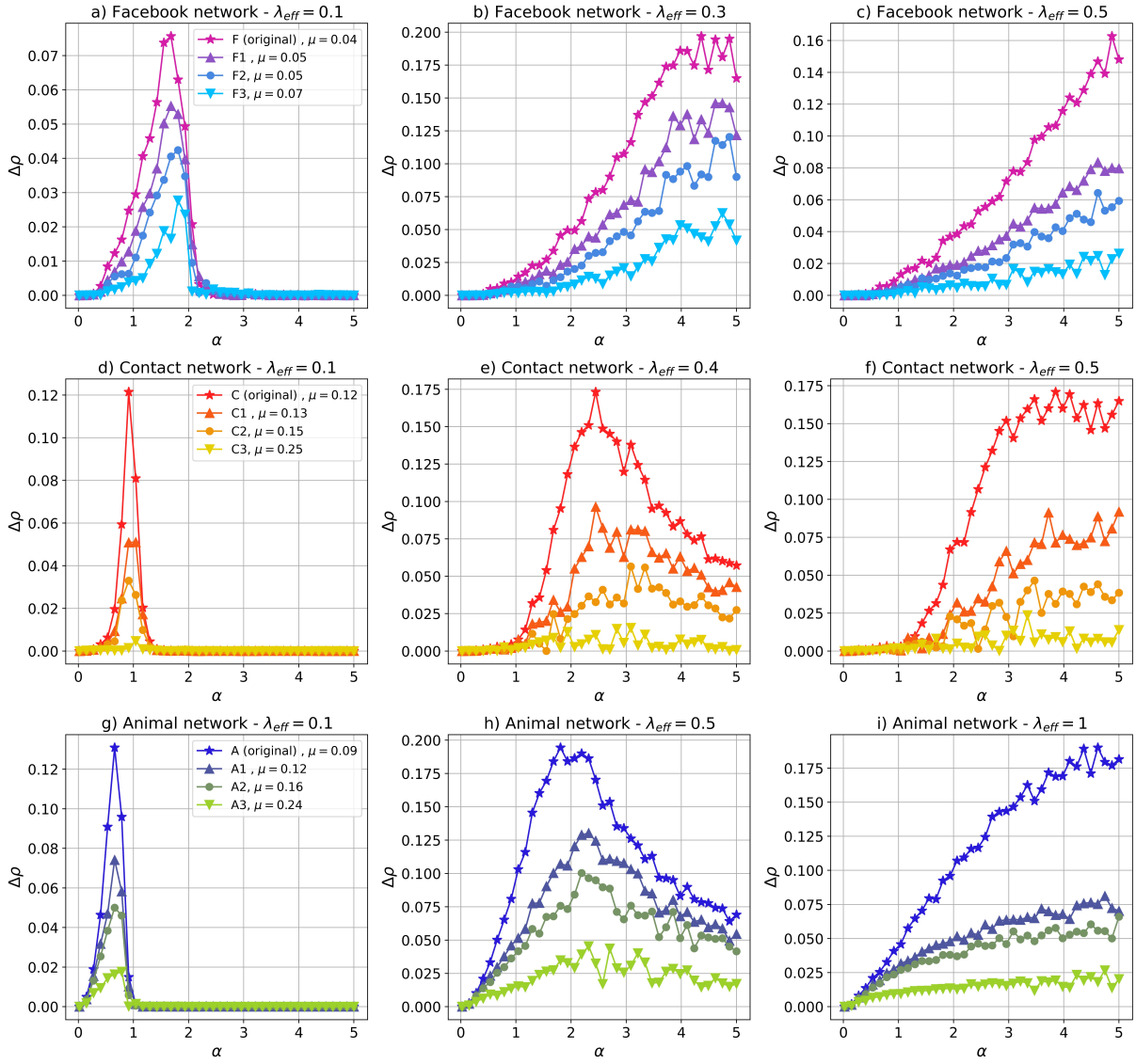

Figure 42 – Difference in prevalence as a function of the shape parameter α for the simulations performed in real network data sets. In the first row (a,b and c) corresponds to Facebook networks F, with each plot featuring a distinct effective rate. The second row (d,e and f) have simulations in the Contact networks C,while the last row (g, h, and i) illustrates plots based on the Animal network A. In all plots each curve, with a different color, represents a different community structures as described in the side boxes and following the notation in 2

An essential aspect to note is that we employed distinct values of λ_{eff} for each of the three cases. This decision stems from the fact that the average degree of the Animal network is significantly smaller than that of the other two networks. To compensate for this disparity and ensure accurate results, we opted for higher values of λ_{eff} for this case.

Across the three real networks we examined, we observed a behavior similar to that seen in the synthetic networks. For smaller values of the effective rate in 42 a, 42 d, 42g, 42e, and 42h there is a peak in the difference in prevalence. Lower shape parameters diminish the impact of the community structure by causing infections to occur primarily during the early stages of the outbreak. On the other hand, higher values of α lead to a change in the critical effective rate, resulting in the epidemic entering an absorbing phase.

Furthermore, as we increased the effective transition rate, the process of swapping links and consequently weakening community structures had a significant impact on reducing the epidemic's prevalence. This effect became more pronounced with an increase in the shape parameter.

However, it is important to highlight that the impact of the swapping process on the Facebook network was relatively lesser. This can be explained due to the Facebook network's significantly higher number of links compared to networks A and C. Despite performing a greater number of link swaps in these networks, the change in μ was not substantial enough to induce a significant shift in prevalence. Nevertheless, the prevalence still exhibited a maximum difference of up to 7%.

6.5 Summary

In this chapter, we explored the problem of an epidemic SIR process with Weibull infection in networks with community structure. The main goal was to understand how the aging process and the community structure change the resulting epidemic and how these two characteristics interact with each other. In section 6.3.2 we showed that larger values of the shape parameter, which cause positive aging, can significantly reduce epidemic sizes and alter the epidemic life. Also, strong community structures, for this case $\mu < 0.1$, can also contain the epidemic sizes. On the other hand for larger values of mixing parameter, in general, with $\mu > 0.3$, we generate networks with more connected communities, and a more limited impact of the community structure in the resulting epidemics is observed.

Furthermore, in section 6.3.4 we explored how the non-Markovian infection can alter the role of the modularity in the propagation. To accomplish this, we defined the variable called “difference in prevalence” which allows us to compare two epidemic processes in similar networks but with different community structures. For small values of λ_{eff}, near the critical rate, there is a shape parameter that maximizes the difference in prevalence. This indicates that for this value of α the community structure might play a central role in the epidemic. This is caused because for small α quick infections happen and the communities can’t contain the disease, and, for larger α the critical rate increases and the system may return to the absorbing phase. This analysis was also carried on in real networks constructed from databases, and the same behavior was observed.

We also explored the relevance of the number of communities and found that, for this mode, increasing the number of communities may help further contain the epidemic process 6.3.5.

CHAPTER

7

CONCLUSIONS

From diseases that threaten society to information dissemination, the challenge of understanding the dynamics of propagation phenomena is one most relevant topics of modern research. Nevertheless, the complexity of these systems is abysmal, so choosing which variables or characteristics will we take into account for making mathematical models is a challenge in itself. In this context, the use of network science has been a fundamental tool to conserve some complexity while still allowing relatively simple approaches to study epidemic processes. The use of Markovian assumption has been a pillar of some of the most used and studied models in epidemics on networks. This is a simplification that has been challenged, as shown in chapter 4.3, as real systems are rarely Moravian, especially those associated with human interaction and biological processes.

It is well known that considering non-Markovian characteristics can drastically change the outcome of an epidemic. For example, the works of Piet Van Mieghem and colleagues [Cator, Bovenkamp and Mieghem 2013, Mieghem and Liu 2019, Mieghem and Bovenkamp 2013] have shown that considering a SIS non-Markovian infection characterized by a Weibull distribution can change the critical transition rate λ_c and epidemic size, which are fundamental metrics used in the epidemic process in complex networks. However, despite the increase in research that focuses on non-Markovian epidemics, this is still a relatively unexplored topic that has many answered questions. This dissertation had the main goal of contributing to some of these unexplored challenges, focusing on two problems related to SIS and SIR non-Markovian epidemics with the Weibull infection process and a Markovian recovery.

The first problem developed in this research, described in chapter 5, was characterizing and comparing SIS and SIR non-Markovian epidemics with Weibull infection in three different network models: a random regular networks described by the Erdős and Rényi model (ER), a heterogeneous scale-free networks described by the Barabási Albert model (BA), and a network with small world characteristic described by the Watts Strogatz model (WS). The SIS non-Markovian epidemic with Weibull infection is one of the most studied models in the literature of non-Markovian epidemics. Some of our funding further support the behavior described in the literature. The SIR model has been less explored and some of our findings for this model are new.

From a general point of view, for both the SIS and SIR non-Markovian epidemic process with Weibull infection and for all three network models, the shape parameter of the Weibull distribution can dramatically change the total size of the epidemic and the critical effective transition rate (λ_c). When we consider a negative aging ($\alpha < 1$), as the disease ages in the infected node the probability of having smaller inter-event infection times increases, and as a consequence, there is a quick spreading increasing the size of the epidemic and decreasing the λ_c, almost vanishing for small values of α. On the other hand, when $\alpha > 1$, there is a positive aging, and as time goes by there is a higher chance of larger inter-event times, with a localized maximum, and so epidemics with smaller sizes are observed with larger values of λ_c.

Besides that, for the SIS model, a susceptibility analysis was made to find the critical rate for different network sizes for the three models. As N increases for the WS and ER models, the critical rate does not suffer major changes, which was the expected behavior as they are homogeneous networks. On the contrary, for the BA model which generates heterogeneous networks, the critical rates decrease with the system size, indicating a vanishing threshold in the thermodynamic limit. With these results, we conclude that the vanishing threshold associated with heterogeneity also occurs for this kind of non-Markovian epidemic.

We compared the critical rate found through the susceptibility analysis with the mean-field approximations described in [Cator, Bovenkamp and Mieghem 2013, Mieghem and Liu 2019]. For the ER and BA models, there was a good agreement between the NIMFA approximation and the numerical values, especially for larger network sizes. In the case of the WS models the numerical values of λ_c were slightly above the mean-field approximation, which is probably attributed to the clustering effect of these networks, which are not taken into account in the NIMFA approximation. The upper bound $\lambda_c^{UB} = 1/ln(\lambda + 1)$ described in [Barabási 2005],

also works well for the three models. For all cases, higher values of α appeared to divert more from the mean-field approximations.

For the SIR epidemic processes, the critical effective rate was numerically calculated using the vulnerability analysis [Shu *et al.* 2015]. We found that this method, originally used for the Markovian case, also works well for the non-Markovian dynamics. The same general results of the SIS model were found for the SIR model. The scaling of the system indicates a vanishing threshold for the heterogeneous BA networks, as the critical value consistently shrank as N increases. As expected, for the homogeneous models WS and ER this behavior was not observed, from which we can conclude that the vanishing threshold associated with heterogeneity also applies to this non-Markovian SIR process. Despite not having a mean-field approach for the SIR model, the NIMFA critical rate for the SIS model serves well as a lower bound, as stated in [Cator, Bovenkamp and Mieghem 2013], and the upper bound λ_c^{UB} can be used for smaller values of α.

Finally, for both the SIS and SIR models, we compared the critical rates for the three network models. We found that the BA network, with finite size, has lower values of critical rate. The WS model has the highest critical rates, even compared with ER models with the same largest eigenvalue and average degree. This higher value of λ_c can be attributed to the clustering effect that has been observed to inhibit the spread of epidemics in homogeneous networks.

The second main problem discussed in this dissertation, developed in chapter 6, was the study of a non-Markovian SIR process with Weibull infection in networks with modular structure. To accomplish this, we studied the effect of the shape parameter in LFR benchmark networks, which present community structure, and have the mixing parameter μ, which is the average fraction of inter-community links, and is a way to measure how strong is the community structure. The general behavior observed was that the prevalence, or size of the epidemic has larger values when the shape parameter decreases. This is consistent with previous works in other types of network structures [Mieghem and Bovenkamp 2013]. Strong community structures can also reduce epidemic sizes. For larger values of mixing parameters, in general, with $\mu > 0.3$, we generate networks with more connected communities. In this case, small changes in prevalence are observed when increasing μ, which indicates a limited impact of the modular structure on the total number of infected individuals.

Additionally, we investigated how non-Markovian infection can affect the role of modu-

larity in disease propagation. First, in section 6.3.1 we made an analysis of the time T_M, defined as the time for which the disease has reached at least one individual of each community. This quantity consistently increases with α indicating that the aging process delays the passage of the disease between communities.

The effect of this phenomenon was further supported by the study of "difference in prevalence" defined as the difference between the epidemic sizes of two epidemic processes with the same infection characteristics (λ_{eff} and α) but different community structure. For small values of λ_{eff}, near the critical rate, there is a shape parameter that maximizes the difference in prevalence. This indicates that for this value of α the community structure plays a more important role. This is caused because for small α quick infections are common, and the communities can't contain the infection. On the contrary, for larger α there is an increase in the critical rate, and if this increase is enough λ_{eff} falls below λ_c, causing the system to return to the absorbing state. For larger values of λ_{eff} this is less evident in a small range of α, however, if a large enough range is chosen the maximum behavior will probably persist.

The more important conclusion from this part of the project is that community structure can significantly reduce the fraction of individuals infected, but this effect is reduced when dealing with negative aging processes ($\alpha < 1$), which prioritize infections in early age times. However, positive aging ($\alpha > 1$), as long as the critical transition rate is below the effective rate chosen, can boost the effect of the community structure delaying the spread of the disease to healthy communities. In other words, if the infection process favors late infection times, the presence of communities can help contain the disease in the network.

7.1 Limitations and future woks

Despite having many interesting results concerning the effect of the non-Markovian epidemic process with Weibull infection in different network typologies, there are still some limitations of the results studied. In this subsection, we will highlight some of these limitations and propose possible solutions that could be implemented in the future. These limitations are enumerated and described as follows:

1. **Implement a better algorithm to calculate the susceptibility and vulnerability:** We mentioned in section 3.2.2.2 that the standard approach for calculating the susceptibility and vulnerability of a Markovian epidemic process is the Quasi-stationary state analysis [Shu *et al.* 2015]. This technique is very efficient and results in a better approximation for the critical values than other "brute force" approaches. Although there are some more refined methods than the one used in this project [Cator, Bovenkamp and Mieghem 2013], there still needs more research to optimize the calculation of the susceptibility/vulnerability. A possible solution could be to make an adaptation of the Quasi-stationary state analysis used in the Markovian case, or if this ends up unfeasible just applying a more refined approach to obtain better approximations for the critical rates would be a big improvement for the results presented in 5.3.1 and 5.4.1.

2. **Make a critical exponent analysis:** The susceptibility/variability analysis to find a critical effective transition rate is just a part a more complete analysis of phase transitions. Critical exponents are calculated to completely characterize the phase transition of the system. This kind of analysis has been made for the SIS non-Markovian model [Starnini, Gleeson and ná 2017], and it would be interesting to extend these results to the SIR process.

3. **NIMFA Mean field approach for the SIR non-Markovian model:** Although there is pairwise mien field approximation for the SIR non-Markovian process, there has not been developed an N-intertwined approximation for this model. This could be interesting to compare an eventual critical rate with the one found for the SIS model and with the values found through the vulnerability analysis.

4. **Explore different modular network structures:** Although the LFR algorithm is very versatile to generate networks with different topological characteristics, and in particular different community structures. It would be interesting to analyze the behavior of non-

Markovian epidemics in LFR networks with different characteristics. For instance, the influence of the power law exponents for generating the community sizes and degrees, which in this work were held constant.

5. **Calculate the critical transition rate for the SIR non-Markovian epidemic model:** We mentioned trough out chapter 6 that many times the systems return to the absorbing phase for some values of α. These observations where made by the fact that the average prevalence values for these cases where near zero. However, it would be interesting to calculate the critical rate and its dependence on α and possibly μ, to further support the behavior described.

Besides these limitations, there are also some interesting topics not covered in this text that would be valuable for future works. We can highlight the followings:

1. **Explore other compartmental model:** It would be valuable to study compartmental models with more compartments like the SIRS model or SEIR model.

2. **Explore other distributions for the inter-event times:** The focus of this project, and the majority of the literature is on the non-Markovian epidemic with Weibull infection. Despite exiting some results on general distributions [Mieghem and Liu 2019], there has been little exploration for other distributions that might have a more biological-inspired characteristics.

3. **Dynamic-sensitive methods to detect vital nodes on networks:** This is a technique where epidemic processes are performed to find which nodes have a bigger influence in the propagation processes, this has been studied for Markovian processes [Lü *et al.* 2016], but as we saw in the preliminary results considering a non-Markovian transmission process can drastically change the outcome of an epidemic, so a natural follow up would be to verify if dynamic-sensitive methods with non-Markovian processes have similarities or differences with the traditional method.

4. **Explore rumor spreading models with non-Markovian characteristic:** The Maki Thompson model, which uses a variation of the SIR model to describe rumor-spreading phenomena, is a very popular model for rumor-spreading. For instance, there have been works with multilayer networks where a rumor dynamic spread in one layer and an epidemic

process in another one, with the goal to model real epidemics that can be altered by public knowledge and information dissemination.

5. **Use the renewal approach for non-Markovian dynamics in other dynamical problems:** Applying this approach based on renewal processes to model non-Markovian processes in problems from other research field like: ecological problems, as in the predator pray models, or social behavior problems, as in the voting models.

Finally, we hope this work inspires new debates and works on the topic of memory effects in spreading phenomena.

BIBLIOGRAPHY

ALBERT, R.; JEONG, H.; BARABASI, A.-L. Diameter of the world-wide web. **Nature**, v. 401, p. 130–131, 09 1999. Citation on page 45.

ANDERSON, R. M.; MAY, R. M. **Infectious diseases of humans: dynamics and control**. [S.l.]: Oxford University Press, 1992. Citation on page 31.

BARABÁSI, A.; PóSFAI, M. **Network Science**. Cambridge University Press, 2016. ISBN 9781107076266. Available: <https://books.google.com.br/books?id=iLtGDQAAQBAJ>. Citations on pages 32, 40, 41, 43, 45, and 49.

BARABÁSI, A.-L. The origin of bursts and heavy tails in human dynamics. **Nature**, Springer Science and Business Media LLC, v. 435, n. 7039, p. 207–211, may 2005. Available: <https://doi.org/10.1038%2Fnature03459>. Citations on pages 33, 68, and 130.

BARABASI, A.-L.; ALBERT, R. Albert, r.: Emergence of scaling in random networks. science 286, 509-512. **Science (New York, N.Y.)**, v. 286, p. 509–12, 11 1999. Citation on page 45.

BARRAT, A.; BARTHéLEMY, M.; VESPIGNANI, A. **Dynamical Processes on Complex Networks**. [S.l.]: Cambridge University Press, 2008. Citation on page 61.

BIKHCHANDANI, S.; HIRSHLEIFER, D.; WELCH, I. A theory of fads, fashion, custom, and cultural change as informational cascades. **Journal of Political Economy**, v. 100, p. 992–1026, 10 1992. Citation on page 32.

BLYTHE, S.; ANDERSON, R. Variable infectiousness in hfv transmission models. **IMA journal of mathematics applied in medicine and biology**, v. 5, p. 181–200, 02 1988. Citation on page 68.

BOGUñá, M.; LAFUERZA, L. F.; TORAL, R.; SERRANO, M. Simulating non-markovian stochastic processes. **Phys. Rev. E**, American Physical Society, v. 90, p. 042108, Oct 2014. Available: <https://link.aps.org/doi/10.1103/PhysRevE.90.042108>. Citations on pages 32, 33, 71, 73, 80, and 82.

BROCKMANN, D.; HUFNAGEL, L.; GEISEL, T. The scaling laws of human travel. **Nature**, v. 439, p. 462–5, 02 2006. Citations on pages 33 and 68.

BUTTS, C. Revisiting the foundations of network analysis. **Science (New York, N.Y.)**, v. 325, p. 414–6, 08 2009. Citation on page 32.

CAI, X.; FRY, C.; WAGNER, C. International collaboration during the covid-19 crisis: Autumn 2020 developments. 11 2020. Citation on page 31.

CATOR, E.; BOVENKAMP, R. van de; MIEGHEM, P. V. Susceptible-infected-susceptible epidemics on networks with general infection and cure times. **Phys. Rev. E**, American Physical Society, v. 87, p. 062816, Jun 2013. Available: <https://link.aps.org/doi/10.1103/PhysRevE.87.062816>. Citations on pages 33, 70, 72, 86, 88, 90, 100, 129, 130, 131, and 133.

CHAKRABARTI, D.; WANG, Y.; WANG, C.; LESKOVEC, J.; FALOUTSOS, C. Epidemic thresholds in real networks. **ACM Trans. Inf. Syst. Secur.**, Association for Computing Machinery, New York, NY, USA, v. 10, n. 4, jan 2008. ISSN 1094-9224. Available: <https://doi.org/10.1145/1284680.1284681>. Citation on page 62.

CHEN, H.; WANG, S.; SHEN, C.-s.; ZHANG, H.-F.; BIANCONI, G. Non-markovian majority-vote model. **Physical Review E**, v. 102, 12 2020. Citation on page 68.

CUYPERE, E. D.; TURCK, K. D.; WITTEVRONGEL, S.; FIEMS, D. Markovian sir model for opinion propagation. In: **Proceedings of the 2013 25th International Teletraffic Congress (ITC)**. [S.l.: s.n.], 2013. p. 1–7. Citation on page 73.

DUCH, J.; ARENAS, A. Community detection in complex networks using extremal optimization. **Phys. Rev. E**, American Physical Society, v. 72, p. 027104, Aug 2005. Available: <https://link.aps.org/doi/10.1103/PhysRevE.72.027104>. Citation on page 50.

EBADI, H.; SAEEDIAN, M.; AUSLOOS, M.; JAFARI, G. Effect of memory in non-markovian boolean networks illustrated with a case study: A cell cycling process. **EPL (Europhysics Letters)**, v. 116, p. 30004, 11 2016. Citation on page 68.

ERDÖS, P.; RÉNYI, A. On random graphs i. **Publicationes Mathematicae Debrecen**, v. 6, p. 290–297, 1959. Citation on page 42.

ERDÖS, P. L.; RENYI, A. On the evolution of random graphs. **Transactions of the American Mathematical Society**, v. 286, p. 257–257, 1984. Citation on page 42.

FENG, M.; CAI, S.-M.; TANG, M.; LAI, Y.-C. Equivalence and its invalidation between non-markovian and markovian spreading dynamics on complex networks. **Nature Communications**, v. 10, p. 1–10, 08 2019. Citations on pages 33, 72, and 73.

FERDOUSI, T.; COHNSTAEDT, L.; MCVEY, D.; SCOGLIO, C. Understanding the survival of zika virus in a vector interconnected sexual contact network. **Scientific Reports**, v. 9, 05 2019. Citation on page 73.

FERREIRA, S. C.; CASTELLANO, C.; PASTOR-SATORRAS, R. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. **Phys. Rev. E**, American Physical Society, v. 86, p. 041125, Oct 2012. Available: <https://link.aps.org/doi/10.1103/PhysRevE.86.041125>. Citation on page 65.

FIENBERG, S. E. A brief history of statistical models for network analysis and open challenges. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 21, n. 4, p. 825–839, 2012. Available: <https://doi.org/10.1080/10618600.2012.738106>. Citations on pages 42, 43, and 47.

FORTUNATO, S.; HRIC, D. Community detection in networks: A user guide. **Physics Reports**, Elsevier, v. 659, p. 1–44, 2016. Citation on page 103.

GHALMANE, Z.; HASSOUNI, M. E.; CHERIFI, C.; CHERIFI, H. Centrality in modular networks. **EPJ Data Science**, v. 8, 12 2019. Citations on pages 103 and 121.

GHALMANE, Z.; HASSOUNI, M. E.; CHERIFI, H. Immunization of networks with non-overlapping community structure. **Social Network Analysis and Mining**, v. 9, 08 2019. Citation on page 104.

GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. **Journal of Computational Physics**, Academic Press, v. 22, n. 4, p. 403–434, 1976. Citations on pages 64, 72, and 82.

GOFFMAN, W.; NEWILL, V. Generalization of epidemic theory: An application to the transmission of ideas. **Nature**, v. 204, p. 225–228, 1964. Citation on page 32.

GONZALEZ, M. C.; HIDALGO, C.; BARABASI, A.-L. Understanding individual human mobility patterns. **Nature**, v. 453, p. 779–82, 07 2008. Citations on pages 33 and 68.

GROSSMANN, G.; BORTOLUSSI, L.; WOLF, V. Efficient simulation of non-markovian dynamics on complex networks. **PLoS ONE**, v. 15, 10 2020. Citation on page 73.

GUPTA, N.; SINGH, A.; CHERIFI, H. Community-based immunization strategies for epidemic control. **2015 7th International Conference on Communication Systems and Networks, COMSNETS 2015 - Proceedings**, 11 2014. Citation on page 104.

HAN, L.; LIN, Z.; YIN, Q.; TANG, M.; GUAN, S.; BOGUNA, M. **Non-Markovian epidemic spreading on temporal networks**. 2023. Citation on page 74.

HU, Y.; LI, M.; ZHANG, P.; FAN, Y.; DI, Z. Community detection by signaling on complex networks. **Phys. Rev. E**, American Physical Society, v. 78, p. 016115, Jul 2008. Available: <https://link.aps.org/doi/10.1103/PhysRevE.78.016115>. Citation on page 50.

HUANG, L.; PARK, K.; LAI, Y.-C. Information propagation on modular networks. **Phys. Rev. E**, American Physical Society, v. 73, p. 035103, Mar 2006. Available: <https://link.aps.org/doi/10.1103/PhysRevE.73.035103>. Citation on page 104.

HUANG, W.; LI, C. Epidemic spreading in scale-free networks with community structure. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2007, n. 01, p. P01014–P01014, jan 2007. Available: <https://doi.org/10.1088/1742-5468/2007/01/p01014>. Citation on page 104.

HUI, Z.; ZI-YOU, G. Modular epidemic spreading in small-world networks. **Chinese Physics Letters**, IOP Publishing, v. 24, n. 4, p. 1114–1117, apr 2007. Available: <https://doi.org/10.1088/0256-307x/24/4/073>. Citation on page 104.

INFECTIOUS network dataset – KONECT. 2017. Available: <http://konect.cc/networks/sociopatterns-infectious>. Citation on page 121.

JACKSON, M. O. **Social and Economic Networks**. USA: Princeton University Press, 2008. ISBN 0691134405. Citation on page 32.

JO, H.-H.; PEROTTI, J. I.; KASKI, K.; KERTÉSZ, J. Analytically solvable model of spreading dynamics with non-poissonian processes. **Physical Review X**, American Physical Society (APS), v. 4, n. 1, mar 2014. Available: <https://doi.org/10.1103%2Fphysrevx.4.011041>. Citation on page 71.

KARLIN, S.; TAYLOR, H. **A First Course in Stochastic Processes**. Elsevier Science, 1975. (First course in stochastic processes / Samuel Karlin; Howard M. Taylor, v. 1). ISBN 9780123985521. Available: <https://books.google.com.br/books?id=Vq69ZDOyLIUC>. Citation on page 52.

KEELING, M.; DANON, L.; FORD, A.; HOUSE, T.; JEWELL, C.; ROBERTS, G.; ROSS, J.; VERNON, M. Networks and the epidemiology of infectious disease. **Interdisciplinary Perspectives on Infectious Diseases**, Hindawi Publishing Corporation, v. 2011, 2011. ISSN 1687-708X. Citation on page 32.

KEELING, M.; EAMES, K. Networks and epidemic models. **Journal of the Royal Society, Interface / the Royal Society**, v. 2, p. 295–307, 10 2005. Citations on pages 32 and 36.

KELLA, O.; STADJE, W. Superposition of renewal processes and an application to multi-server queues. **Statistics & Probability Letters**, v. 76, p. 1914–1924, 11 2006. Citation on page 68.

KENAH, E. Contact intervals, survival analysis of epidemic data, and estimation of r0. **Biostatistics (Oxford, England)**, v. 12, p. 548–66, 11 2010. Citation on page 68.

KERMACK, W. O.; MCKENDRICK, A. G.; WALKER, G. T. A contribution to the mathematical theory of epidemics. **Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character**, v. 115, n. 772, p. 700–721, 1927. Citations on pages 31 and 35.

KISS, I.; MILLER, J.; SIMON, P. **Mathematics of Epidemics on Networks**. [S.l.: s.n.], 2017. ISBN 978-3-319-50804-7. Citations on pages 35, 36, 37, 38, 39, 73, and 82.

KISS, I. Z.; RÖST, G.; VIZI, Z. Generalization of pairwise models to non-markovian epidemics on networks. **Phys. Rev. Lett.**, American Physical Society, v. 115, p. 078701, Aug 2015. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.115.078701>. Citation on page 71.

KUGA, K.; TANIMOTO, J. Effects of void nodes on epidemic spreads in networks. **Scientific Reports**, v. 12, 03 2022. Citation on page 73.

LANCICHINETTI, A.; FORTUNATO, S.; RADICCHI, F. Benchmark graphs for testing community detection algorithms. **Phys. Rev. E**, American Physical Society, v. 78, p. 046110, Oct 2008. Available: <https://link.aps.org/doi/10.1103/PhysRevE.78.046110>. Citations on pages 50 and 104.

LAURO, F. D.; KHUDABUKHSH, W. R.; KISS, I. Z.; KENAH, E.; JENSEN, M.; REMPALA, G. A. **Dynamic Survival Analysis for non-Markovian Epidemic Models**. arXiv, 2022. Available: <https://arxiv.org/abs/2202.09948>. Citation on page 73.

LESKOVEC, J.; KREVL, A. **SNAP Datasets: Stanford Large Network Dataset Collection**. 2014. <http://snap.stanford.edu/data>. Citation on page 121.

LIU, Q.; MIEGHEM, P. V. Burst of virus infection and a possibly largest epidemic threshold of non-markovian susceptible-infected-susceptible processes on networks. **Phys. Rev. E**, American Physical Society, v. 97, p. 022309, Feb 2018. Available: <https://link.aps.org/doi/10.1103/PhysRevE.97.022309>. Citations on pages 72 and 88.

LU, Z.; YU, Y.; CHEN, Y.; REN, G.; XU, C.; WANG, S. Stability analysis of a nonlocal sihrdp epidemic model with memory effects. **Nonlinear Dynamics**, v. 109, 07 2022. Citation on page 73.

Lü, L.; CHEN, D.; REN, X.-L.; ZHANG, Q.-M.; ZHANG, Y.-C.; ZHOU, T. Vital nodes identification in complex networks. **Physics Reports**, v. 650, 07 2016. Citation on page 134.

MASUDA, N.; ROCHA, L. E. C. A gillespie algorithm for non-markovian stochastic processes. **SIAM Review**, Society for Industrial & Applied Mathematics (SIAM), v. 60, n. 1, p. 95–115, jan 2018. Available: <https://doi.org/10.1137%2F16m1055876>. Citations on pages 32, 64, 73, and 82.

MIEGHEM, P. V. **Performance Analysis of Communications Networks and Systems**. [S.l.]: Cambridge University Press, 2006. Citations on pages 56, 58, and 69.

MIEGHEM, P. V.; BOVENKAMP, R. van de. Non-markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. **Phys. Rev. Lett.**, American Physical Society, v. 110, p. 108701, Mar 2013. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.110.108701>. Citations on pages 33, 70, 117, 129, and 131.

MIEGHEM, P. V.; LIU, Q. Explicit non-markovian susceptible-infected-susceptible mean-field epidemic threshold for weibull and gamma infections but poisson curings. **Phys. Rev. E**, American Physical Society, v. 100, p. 022317, Aug 2019. Available: <https://link.aps.org/doi/10.1103/PhysRevE.100.022317>. Citations on pages 33, 69, 72, 88, 95, 100, 129, 130, and 134.

MIN, Y.; JIN, X.; GE, Y.; CHANG, J. The role of community mixing styles in shaping epidemic behaviors in weighted networks. **PloS one**, v. 8, p. e57100, 02 2013. Citation on page 104.

MORENO, Y.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic outbreaks in complex heterogeneous networks. **Physics of Condensed Matter**, v. 26, 07 2001. Citation on page 62.

MORITA, S. Six susceptible-infected-susceptible models on scale-free networks. **Scientific Reports**, Springer Science and Business Media LLC, v. 6, n. 1, mar 2016. Available: <https://doi.org/10.1038%2Fsrep22506>. Citation on page 60.

MUSCOLONI, A.; CANNISTRACI, C. A nonuniform popularity-similarity optimization (npso) model to efficiently generate realistic complex networks with communities. **New Journal of Physics**, v. 20, 05 2018. Citation on page 50.

NADINI, M.; SUN, K.; UBALDI, E.; STARNINI, M.; RIZZO, A.; PERRA, N. Epidemic spreading in modular time-varying networks. **Scientific Reports**, v. 8, 02 2018. Citation on page 104.

NEWMAN, M. **Networks: an introduction**. [S.l.]: Oxford University Press, 2018. Citations on pages 40, 41, and 43.

PALLA, G.; DERéNYI, I.; FARKAS, I.; VICSEK, T. Uncovering the overlapping community structure of complex networks in nature and society. **Nature**, v. 435, p. 814–818, 07 2005. Citations on pages 49 and 103.

PANG, G.; PARDOUX, E. **Functional Limit Theorems for Non-Markovian Epidemic Models**. arXiv, 2020. Available: <https://arxiv.org/abs/2003.03249>. Citation on page 71.

PANISCHEV, O.; YULMETYEV, R.; DEMIN, S. A. Age-related alterations of non-markovian effects in stochastic dynamics of heart rate variability. In: **2005 11th International Scientific and Practical Conference of Students, Post-graduates and Young Scientists - Modern Technique and Technologies**. [S.l.: s.n.], 2005. p. 101–103. Citation on page 68.

PASTA, M. Q.; ZAIDI, F. **Leveraging Evolution Dynamics to Generate Benchmark Complex Networks with Community Structures**. arXiv, 2016. Available: <https://arxiv.org/abs/1606.01169>. Citation on page 50.

PASTOR-SATORRAS, R.; CASTELLANO, C.; MIEGHEM, P. V.; VESPIGNANI, A. Epidemic processes in complex networks. **Reviews of modern physics**, APS, v. 87, n. 3, p. 925, 2015. Citations on pages 31, 32, 35, 38, 51, 58, 59, 64, 65, and 104.

PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic dynamics and endemic states in complex networks. **Phys. Rev. E**, American Physical Society, v. 63, p. 066117, May 2001. Available: <https://link.aps.org/doi/10.1103/PhysRevE.63.066117>. Citations on pages 32 and 60.

____. Epidemic spreading in scale-free networks. **Phys. Rev. Lett.**, American Physical Society, v. 86, p. 3200–3203, Apr 2001. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.86.3200>. Citation on page 60.

PERALTA, A. F.; KHALIL, N.; TORAL, R. Reduction from non-markovian to markovian dynamics: the case of aging in the noisy-voter model. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2020, n. 2, p. 024004, feb 2020. Available: <https://doi.org/10.1088%2F1742-5468%2Fab6847>. Citation on page 68.

RAVASZ, E.; SOMERA, A. L.; MONGRU, D. A.; OLTVAI, Z. N.; BARABA

SI, A.-L. Hierarchical organization of modularity in metabolic networks. **Science**, American Association for the Advancement of Science (AAAS), v. 297, n. 5586, p. 1551–1555, aug 2002. Available: <https://doi.org/10.11262Fscience.1073374>. Citations on pages 49 and 103.

REDNER, S. How popular is your paper? an empirical study of the citation distribution. **The European Physical Journal B: Condensed Matter and Complex Systems**, v. 4, n. 2, p. 131–134, 1998. Available: <https://EconPapers.repec.org/RePEc:spr:eurphb:v:4:y:1998:i:2:p:131-134>. Citation on page 45.

RINNE, H. **The Weibull Distribution: A Handbook**. CRC Press, 2008. ISBN 9781420087444. Available: <https://books.google.com.br/books?id=6wdcTfiLNS4C>. Citations on pages 52, 55, 79, and 81.

ROSS, S. **A First Course in Probability**. Pearson Prentice Hall, 2010. ISBN 9780136033134. Available: <https://books.google.com.br/books?id=Bc1FAQAAIAAJ>. Citations on pages 53 and 54.

ROSSI, R. A.; AHMED, N. K. The network data repository with interactive graph analytics and visualization. In: **AAAI**. [s.n.], 2015. Available: <https://networkrepository.com>. Citation on page 121.

SALATHé, M.; JONES, J. Jones, j.h.: Dynamics and control of diseases in networks with community structure. plos comput. biol. 6(8), e1000736. **PLoS computational biology**, v. 6, p. e1000736, 04 2010. Citation on page 104.

SCHAUB, M. T.; DELVENNE, J.-C.; ROSVALL, M.; LAMBIOTTE, R. The many facets of community detection in complex networks. **Applied Network Science**, Springer Science and Business Media LLC, v. 2, n. 1, feb 2017. Available: <https://doi.org/10.10072Fs41109-017-0023-6>. Citation on page 50.

SHERBORNE, N.; MILLER, J. C.; BLYUSS, K. B.; KISS, I. Z. **Mean-field models for non-Markovian epidemics on networks: from edge-based compartmental to pairwise models**. arXiv, 2016. Available: <https://arxiv.org/abs/1611.04030>. Citation on page 72.

SHU, P.; WANG, W.; TANG, M.; DO, Y. Numerical identification of epidemic thresholds for susceptible-infected-recovered model on finite-size networks. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, v. 25, p. 063104, 06 2015. Citations on pages 65, 66, 131, and 133.

SONG, B.; SONG, Y.; JIANG, G.-P. How clustering affects epidemics in complex networks. In: . [S.l.: s.n.], 2017. p. 178–183. Citation on page 91.

SPIRIN, V.; MIRNY, L. Spirin v, mirny la.. protein complexes and functional modules in molecular networks. proc natl acad sci usa 100: 12123-12128. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, p. 12123–8, 10 2003. Citations on pages 49 and 103.

STARNINI, M.; GLEESON, J. P.; NÁ, M. B. Equivalence between non-markovian and markovian dynamics in epidemic spreading processes. **Phys. Rev. Lett.**, American Physical Society, v. 118, p. 128301, Mar 2017. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.118.128301>. Citations on pages 33, 73, 81, and 133.

STOUFFER, D.; MALMGREN, R.; AMARAL, L. Comment on barabasi, nature 435, 207 (2005). 11 2005. Citation on page 33.

TASGIN, M.; HERDAGDELEN, A.; BINGOL, H. **Community Detection in Complex Networks Using Genetic Algorithms**. arXiv, 2007. Available: <https://arxiv.org/abs/0711.0491>. Citation on page 50.

TOMOVSKI, I.; BASNARKOV, L.; ABAZI, A. **Endemic state equivalence between non-Markovian SEIS and Markov SIS model in complex networks**. arXiv, 2021. Available: <https://arxiv.org/abs/2111.04110>. Citation on page 73.

TRAVIS, J.; MILGRAM, S. An experimental study of the small world problem**the study was carried out while both authors were at harvard university, and was financed by grants from the milton fund and from the harvard laboratory of social relations. mr. joseph gerver provided invaluable assistance in summarizing and criticizing the mathematical work discussed in this paper. In: LEINHARDT, S. (Ed.). **Social Networks**. Academic Press, 1977. p. 179–197. ISBN 978-0-12-442450-0. Available: <https://www.sciencedirect.com/science/article/pii/B9780124424500500183>. Citation on page 47.

TUNC, I.; SHAW, L. B. Effects of community structure on epidemic spread in an adaptive network. **Phys. Rev. E**, American Physical Society, v. 90, p. 022801, Aug 2014. Available: <https://link.aps.org/doi/10.1103/PhysRevE.90.022801>. Citation on page 104.

VAZQUEZ, A.; RÁCZ, B.; LUKÁCS, A.; BARABÁSI, A.-L. Impact of non-poissonian activity patterns on spreading processes. **Physical Review Letters**, American Physical Society (APS), v. 98, n. 15, apr 2007. Available: <https://doi.org/10.1103%2Fphysrevlett.98.158702>. Citations on pages 33 and 70.

VESPIGNANI, A. Predicting the behavior of techno-social systems. **Science (New York, N.Y.)**, v. 325, p. 425–8, 08 2009. Citation on page 32.

WANG, W.; TANG, M.; STANLEY, H.; BRAUNSTEIN, L. Unification of theoretical approaches for epidemic spreading on complex networks. **Reports on Progress in Physics**, v. 80, 12 2016. Citations on pages 61, 62, and 63.

WATTS, D. J.; DODDS, P. S.; NEWMAN, M. E. J. Identity and search in social networks. **Science**, American Association for the Advancement of Science (AAAS), v. 296, n. 5571, p. 1302–1305, may 2002. Available: <https://doi.org/10.11262Fscience.1070120>. Citations on pages 49 and 103.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. **Nature**, v. 393, n. 6684, p. 440–442, 1998. Citations on pages 45 and 47.

YANG, G. Empirical study of a non-markovian epidemic model. **Mathematical Biosciences**, v. 14, n. 1, p. 65–84, 1972. ISSN 0025-5564. Available: <https://www.sciencedirect.com/science/article/pii/0025556472900090>. Citation on page 70.

YIN, H.; LIU, S.; WEN, X. Optimal transition paths of phenotypic switching in a non-markovian self-regulation gene expression. **Phys. Rev. E**, American Physical Society, v. 103, p. 022409, Feb 2021. Available: <https://link.aps.org/doi/10.1103/PhysRevE.103.022409>. Citation on page 68.

YOUSSEF, M.; SCOGLIO, C. An individual-based approach to sir epidemics in contact networks. **Journal of theoretical biology**, v. 283, p. 136–44, 06 2011. Citation on page 63.

YULMETYEV, R. M.; DEMIN, S. A.; PANISCHEV, O. Y.; HäNGGI, P. Age-related alterations of relaxation processes and non-markov effects in stochastic dynamics of r–r intervals variability from human ECGs. **Physica A: Statistical Mechanics and its Applications**, Elsevier BV, v. 353, p. 336–352, aug 2005. Available: <https://doi.org/10.1016%2Fj.physa.2005.01.042>. Citation on page 68.

YULMETYEV, R. M.; EMELYANOVA, N. A.; DEMIN, S. A.; GAFAROV, F. M.; HäNGGI, P.; YULMETYEVA, D. G. Non-markov stochastic dynamics of real epidemic process of respiratory infections. **Physica A: Statistical Mechanics and its Applications**, v. 331, n. 1, p. 300–318, 2004. ISSN 0378-4371. Available: <https://www.sciencedirect.com/science/article/pii/S037843710300863X>. Citation on page 70.


ICMC
USP
SÃO CARLOS