GUILHERME HENRIQUE DA SILVA COSTA

SPREADING PROCESSES ON COMPLEX NETWORKS: THEORY, SIMULATIONS AND APPLICATIONS

Thesis submitted to the Physics Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Silvio da Costa Ferreira Júnior

VIÇOSA - MINAS GERAIS
2022

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
C837s
2022

Costa, Guilherme Henrique da Silva, 1995-
Spreading processes on complex networks: theory, simulations and applications / Guilherme Henrique da Silva Costa. – Viçosa, MG, 2022.
1 tese eletrônica (186 f.): il. (algumas color.).

Texto em inglês.
Inclui apêndices.
Orientador: Silvio da Costa Ferreira Junior.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Física, 2022.
Referências bibliográficas: f. 168-186.
DOI: https://doi.org/10.47328/ufvbbt.2022.508
Modo de acesso: World Wide Web.

1. Física estatística. 2. Espalhamento (Física). 3. Epidemiologia - Métodos estatísticos. I. Ferreira Junior, Silvio da Costa, 1976-. II. Universidade Federal de Viçosa. Departamento de Física. Programa de Pós-Graduação em Física. III. Título.

CDD 22. ed. 530.1595

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

GUILHERME HENRIQUE DA SILVA COSTA

SPREADING PROCESSES ON COMPLEX NETWORKS: THEORY, SIMULATIONS AND APPLICATIONS

Thesis submitted to the Physics Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: April 25, 2022.

Assent:

Guilherme Henrique da Silva Costa
Author

Silvio da Costa Ferreira Junior
Adviser

Acknowledgments

Ao fim de mais um ciclo acadêmico, tentarei reunir todes que contribuíram para que eu chegasse aqui. Como um sistema caótico, sei que mínimas perturbações à trajetória durante esse caminhar poderiam levar a resultados complentamente diferentes. Não conseguirei cobrir toda essa trajetória mas, se você sente que fez parte disso de alguma forma e seu nome não se encontra abaixo, sinta-se agradecido.

Primeiramente agradeço aos meus familiares, em particular meus pais Elvio e Sheila e irmãos Elvio Jr. e Lucas pelo apoio e confiança durante esses longos anos. Em adendo, não poderia deixar de agradecer à Samantha também pela confiança, carinho e apoio.

Sou grato ao professor Silvio Ferreira, meu orientador desde 2015, pelos conselhos, ensinamentos e pela paciência durante esses anos. Foi sob sua orientação que cresci como pesquisador e pude chegar onde estou. Além de orientador, um amigo que já me ajudou em diversos momentos de dúvida. Que esta tese não seja o fim de nossa colaboração e amizade.

Agradeço aos diversos amigos feitos durante essa trajetória, em particular ao Túlio, Amanda, Hugo, Thainá, Milena, Rafael, Leonardo, Vitor, José Carlos, Juliane, Wesley e Diogo, pelas discussões e conversas mirabolantes, jogatinas de RPG e videogame, confraternizações e principalmente pelo apoio em momentos de fraqueza. Aos demais amigos que se afastaram por algum motivo e não foram citados aqui, vocês também fizeram parte disso. Além disso, agradeço ao Edson, Cleuza e demais funcionários da Livraria UFV, local em que frequentei religiosamente para tomar aquele cafezinho pós-almoço.

Agradeço também aos demais professores e funcionários da Universidade Federal de Viçosa, pela formação acadêmica e assistência durante todo esse período. Em particular aos professores Maximiliano Munford e Sidiney Alves, meus primeiros orientadores, por me inserirem no ambiente de pesquisa e pavimentarem minha chegada aos sistemas complexos. Além disso, agradeço aos meus colaboradores Wesley Cota, Arthur Schulenburg e Marcelo Martins, pelas discussões fomentadas e pesquisas desenvolvidas durante esse período.

Agradeço ao Departamento de Física (DPF/UFV), em particular ao Grupo de Investigação de Sistemas Complexos (GISC) e ao Núcleo de Simulação de Sistemas Complexos (NSSC), por proverem o ferramental computacional necessário para o desenvolvimento dos trabalhos dessa tese.

Por fim mas não menos importante, agradeço às agências de fomento Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001 pelo financiamento direto ou indireto de todos os trabalhos desenvolvidos nessa tese. Em particular, agradeço à CAPES pela bolsa de estudos.

"To absent friends, lost loves, old gods, and the season of mists; and may each and every one of us always give the devil his due"

— Neil Gaiman, 1990

Abstract

COSTA, Guilherme Henrique da Silva, D.Sc., Universidade Federal de Viçosa, April, 2022. **Spreading processes on complex networks: theory, simulations and applications**. Adviser: Silvio da Costa Ferreira Junior.

The use of complex networks as a substrate for the study of dynamic processes is a rich interdisciplinary area with several applications in real problems. In particular, the study of spreading processes in this framework are foundational for several models in biology, physics, and sociology. In this thesis, we develop studies on fundamental and theoretical aspects of dynamical processes, which have a connection with non-equilibrium statistical mechanics, as well as direct and practical applications for the study of real infectious diseases. We developed a quasi-stationary simulation method, one way to analyze systems with absorbing states, aiming at joining the best of existing methods: computational simplicity with precision in the characterization of localized transitions. After extensive simulations of the susceptible-infected-susceptible model in several types of complex networks, we were able to validate the method. Furthermore, we developed an heterogeneous mean-field theory for the symbiotic contact process with two species. Numerical integrations of the equations indicates the existence of a biestability region that depends on the heterogeneity of the complex network in addition to non-trivial finite size behaviors. Computer simulations in complex networks validated the theory and complemented the results. Finally, we developed works with applications for the spreading of COVID-19. In the beginning of the pandemic, we built a data-driven compartmental epidemic model with metapopulations to investigate how Sars-Cov-2 would spread in Brazil in a municipality level. We were able to compare several mitigation scenarios and verify the desynchronization of the epidemic outbreaks in Brazil, a result confirmed a posteriori. In addition, we developed a method to estimate the underreporting of documented COVID-19 cases in different locations in Brazil using a compartmental model. These contributions show the vast applications of dynamic process toolbox in complex networks, in addition to its importance for real situations.

Keywords: Dynamical processes. Theoretical epidemiology. Complex networks.

Resumo

COSTA, Guilherme Henrique da Silva, D.Sc., Universidade Federal de Viçosa, abril de 2022. **Processos de espalhamento em redes complexas: teoria, simulação e aplicações**. Orientador: Silvio da Costa Ferreira Junior.

A utilização de redes complexas como substrato para o estudo de processos dinâmicos é uma área interdiscplinar rica e com aplicações em diversos problemas reais. Em particular, o estudo de processos de espalhamento encontra nesse ferramental a base para diversos modelos em biologia, física e sociologia. Nesta tese, desenvolvemos tanto estudos em aspectos fundamentais e teóricos, os quais possuem conexões com a mecânica estatística de não equilíbrio, quanto aplicações mais diretas e práticas para o estudo de doenças infecciosas reais. Desenvolvemos um método de simulação quasi-estacionário, uma das formas de se analisar sistemas com estados absorventes, com o objetivo de agregar o melhor de métodos já existentes: simplicidade computacional com precisão na caracterização de transições localizadas. Após extensas simulações do modelo suscetível-infectado-suscetível, pudemos validar o método. Além disso, desenvolvemos uma teoria de campo médio heterogênea para o processo de contato simbiótico com duas espécies. Integrações numéricas sugerem a existência de uma região de biestabilidade que varia com a heterogeneidade da rede além de relações de tamanho finito não triviais. Simulações computacionais em redes complexas validaram a teoria e complementaram os resultados. Por fim, desenvolvemos trabalhos com aplicações para a progação da COVID-19. Próximo ao início da pandemia, construímos um modelo epidêmico compartimental com metapopulações e orientado por dados para investigar como se daria o espalhamento do Sars-Cov-2 no Brasil à nível de municípios. Pudemos comparar diversos cenários de mitigação e verificar a dessincronização dos surtos epidêmicos no Brasil, resultado confirmado a posteriori. Além disso, desenvolvemos um método para estimar a subnotificação de casos de COVID-19 documentados em diferentes localidades do Brasil a partir de um modelo compartimental. Essas contribuições mostram as vastas aplicações do ferramental de processos dinâmicos em redes complexas, além da importância para situações reais.

Palavras-chave: Processos dinâmicos. Epidemiologica teórica. Redes complexas.

List of Publications

i). Guilherme S. Costa and Silvio C. Ferreira. *Simple quasistationary method for simulations of epidemic processes with localized states*, Computer Physics Communications, 108046 (2021) [Chapter 4]

ii). Guilherme S. Costa, Marcelo M. de Oliveira and Silvio C. Ferreira. *Heterogeneous mean-field theory for two-species symbiotic processes on networks*, Physical Review E 106, 024302 (2022) [Chapter 5]

iii). Guilherme S. Costa, Wesley Cota and Silvio C. Ferreira. *Outbreak diversity in epidemic waves propagating through distinct geographical scales*, Physical Review Research 2 (4), 043306 (2020) [Chapter 6]

iv). Guilherme S. Costa, Wesley Cota and Silvio C. Ferreira. *Data-driven approach in a compartmental epidemic model to assess undocumented infections*, Chaos, Solitons & Fractals, 112520 (2022) [Chapter 7]

Other contributions

i). Guilherme S. Costa and Silvio C. Ferreira. *Nonmassive immunization to contain spreading on complex networks*, Physical Review E 101 (2), 022311 (2020)

ii). Arthur Schulenburg, Wesley Cota , Guilherme S. Costa and Silvio C. Ferreira, *Effects of infection fatality ratio and social contact matrices on vaccine prioritization strategies*, arXiv:2201.02869 (2022), Accepted for publication on **Chaos: An Interdisciplinary Journal of Nonlinear Science.**

List of acronyms and abbreviations

2SCP . . . . . Two-species symbiotic contact process

ABM . . . . . Agent based modeling

AS . . . . . Absorbing state

ASPT . . . . . Absorbing state phase transition

CFR . . . . . Case fatality ratio

CM . . . . . Configuration model

COMM . . . . . Commuting network

DP . . . . . Directed percolation

EPT . . . . . Equilibrium phase transition

EWF . . . . . External weak fied

FSS . . . . . Finite size scaling

GA . . . . . Gillespie algorithm

HMF . . . . . Heterogeneous mean-field

IBGE . . . . . *Instituto Brasileiro de Geografia e Estatística*

IBMF . . . . . Individual based mean-field

IFR . . . . . Infection fatality ratio

KPZ . . . . . Kardar-Parisi-Zhang

NEPT . . . . . Non-equilibrium phase transition

NPI . . . . . Non-pharmaceutical intervention

OGA . . . . . Optimized Gillespie algorithm

QMF . . . . . Quenched mean-field

r.h.s . . . . . right hand side

RAT . . . . . Reactivation per activity time

RBC . . . . . Reflecting boundary condition

RRN . . . . . Random regular network

SF . . . . . Scale-free

SIR . . . . . Susceptible-infected-removed

SIS . . . . . Susceptible-infected-susceptible

SQS . . . . . Standard quasistationary

UCM . . . . . Uncorrelated configuration model

UF . . . . . Federative unit, from Portuguese *Unidade Federativa*

WS . . . . . Watts-Strogatz

Contents

Part I

Preliminaries

Chapter 1

Introduction

Since the emergence of graph theory in the mid-18th century with the problem of Könisberg bridges [1], many scientists have applied this field of knowledge to describe systems in mathematics, physics, sociology, among other branches of science. With the explosion of network science at the end of the 90's [2] various processes such as social interactions [3, 4], epidemic outbreaks [5, 6, 7] and metabolic chains [8] were studied considering complex networks evolving on top of it. Furthermore, the theoretical and computational frameworks of classical areas of Physics, such as non-equilibrium statistical mechanics, have found several applications in these sub-areas of complexity science [9, 10, 11, 12]. In particular, the study of spreading processes, such as epidemic or information, evolving in complex networks has become one of the major areas of this interdisciplinary science, with several applications [7, 13, 14, 15, 16].

The different dynamic processes that occur in these complex networks are commonly treated by associating variables to the network elements that represent a certain characteristic of the dynamics and defining transition rates between these states [17]. In the context of epidemics spreading, this variable may be related to the stages of the disease and transitions refer to its evolution. By constructing relatively simple epidemic models, relevant information can be extracted to predict and understand the spread of pathogens and define public policies in order to reduce damage [18, 19, 20, 21]. The COVID-19 outbreak is an outstanding recent example of these applications. Starting in December 2019, within a few months the SARS-CoV-2 (pathogen that causes COVID-19) was already present in almost all countries and,

over more than 2 years, infected around 490 million people, leading almost 6 million of those to death [22]. In Brazil, in April of 2022, we approached 30 million cases with more than 660 thousand deaths [23]. Since the beginning of the outbreak, many scientists of different areas devoted efforts to understand and combat this novel treat [24, 25, 26, 27, 28, 29, 30]. In addition to these immediate applications, theoretical and fundamental aspects of some epidemic models are extremely important to understand absorbing state phase transitions and non equilibrium phenomena [31, 32], two classical areas of statistical mechanics [33].

In this thesis, we tackle several problems regarding spreading phenomena on complex networks, ranging from fundamental and theoretical studies on absorbing-state phase transitions to epidemic modeling focused on COVID-19. In Part I, we discuss the main concepts and theoretical foundations of this thesis. In Chapter 2, the main focus is the characterization and description of complex networks. After a brief historical overview, the most important concepts, structures and models are discussed. In Chapter 3, we discuss the main frameworks to study dynamical processes, general concepts and, most importantly, applications for spreading phenomena. Moreover, a brief introduction to mathematical epidemiology is made along with some important concepts, techniques and counterparts within absorbing state phase transitions.

In Parts II and III we tackle the problems themselves. On the former, we discuss more theoretical problems that have a direct connection with absorbing state phase transitions while for the latter we analyze spreading phenomena occurring on geographical networks, applied to COVID-19. In each chapter that makes up these parts, additional concepts (those which are specific enough) are discussed if necessary.

With the constant progress in computational power and the possibility of approaching more complex problems, the use of simulations to validate and complement theoretical analysis grows every day. Considering spreading processes, the use of computational simulations allow us to reach limits that would not be possible in other approaches. In particular, simulations of absorbing-state phase transitions on complex networks possess a major drawback: most theoretical analyses are initially valid for systems within thermodynamic limit (number of elements going to infinite) and need to be compared with finite size simulations. Thus, for many systems, one need to perform simulations on substrates as larger as possible in order to minimize

these finite size effects. Therefore, the algorithms for these simulations needs to be fast and efficient.

Another problem regarding computational simulations of models presenting absorbing transitions is to prevent the system to be trapped into the absorbing state since, strictly, they are the only steady-state for finite sizes [31]. A major approach to circumvent this hurdle is the quasistationary analysis (QS) [31], in which averages are performed only over the space of configurations that did not visit the absorbing state so far. Several methods were developed following this approach [34, 35, 36]. Each one of these methods has its pros and cons, some being statistically accurate but complicated and computationally demanding while others are simple but fail on more complex simulations. In order to improve this methodology, in Chapter 4 we develop a method that aims to be computationally simple and accurate at the same time. To validate the method, we perform simulations of susceptible-infected-susceptible [37] model, a widely used epidemic model that presents a complex behavior [38, 39, 40, 41], in terms of absorbing state phase transitions on heterogeneous complex networks, comparing it with two well established methods: the standard QS [34] (complex but statistically exact) and the reflecting boundary condition [36] (simple but limited). As a general result, we report that our method performs as well as the standard QS in the identification and characterization of the transitions, overcoming the limitations of simpler approaches such as the reflecting boundary condition, but with reduced algorithmic and computer complexity.

In addition to the susceptible-infected-susceptible, another standard model for studying phase transitions for absorbing states is the contact process (CP) [42]. Considered a first approach to describe spreading or replication phenomena in lattices, the characterization of phase transitions of this model in several substrates, such as regular and disordered lattices [43, 44, 45] and complex networks [46, 47], is well known. Several more advanced models were proposed using original CP as a ground model [48, 49]. An interesting extension of CP occurs when we introduce cooperation dynamics between different species evolving in the same substrate. Called symbiotic processes, this type of dynamic is extremely rich and has applications in ecological models [50, 51]. In particular, the two species symbiotic contact process is a recently proposed model [52, 53] whose behavior on complex networks is not fully understood

and described theoretically. Although some recent works have clarified several aspects of the phase transition that occurs in this model in homogeneous substrates [54, 55], the same does not happen in highly heterogeneous substrates, such as complex networks with a power law degree distribution. Thus, we develop a heterogeneous mean field theory (see Chapter 5) for this model in an attempt to characterize its dynamics on complex networks. To validate this theory, we also performed quasistationary simulations on quenched and annealed complex networks.

Concomitantly with fundamental and theoretical works, we dedicate part of our efforts to develop models to describe and characterize the spreading of COVID-19 in Brazil. Since the outbreak's beginning, several modeling works have been developed [29, 56, 57, 58, 59] in order to understand and predict the spreading of SARS-CoV-2 for some countries. Inspired mainly by Reference [29], in Chapter 6 we develop a compartmental model [60] using metapopulations [15, 61] to study the spread of epidemics and apply it to the spread of COVID-19 in Brazil. The model is driven by demographic and mobility data obtained from public repositories [62, 63, 64]. We look at how diverse the epidemic curves are across the country rather than trying to quantitatively predict outbreaks in specific locations. Considering March 31, 2020 as the initial condition, in which there were approximately 6000 cases in 410 Brazilian municipalities [23], we define hypothetical mitigation scenarios and study the epidemic evolution in all municipalities in the country through integrated simulations. We qualitatively predict the variability in epidemic incidence curves across multiple geographic scales. Furthermore, we predict that the epidemic spread of some states is characterized by an outbreak focused initially in the capital, followed by an epidemic wave spreading through the countryside. Based on these results, we are able to consider strategies to contain the epidemic. In principle, the different behaviors in different regions call for decentralized strategies, depending on the region's needs. However, as municipalities are highly interconnected with each other, strategies must be integrated to account for these factors.

Having ventured into developing data-driven models, we saw that one of the biggest limitations that arise when working with them is the reliability and quality of data. In particular for COVID-19, it is known that there is a lot of under-reporting, due to infections by asymptomatic and presymptomatic individuals and testing limi-

tations [65, 66, 67, 68, 69]. This under-reporting may affect the time series of reported cases and, therefore, the quality of predictions for data-driven models. In order to circumvent this problem, in Chapter 7 we develop a methodology to estimate the quantity of individuals in *hidden* compartments, namely the asymptomatic, exposed and undiagnosed infectious individuals from epidemic surveillance data for cases and deaths. We estimate the levels of under-reporting in different localities and time windows. In order to validate the method, we apply it across different geographical scales on Paraná and Espírito Santo, two Brazilian states. Additionally, we estimate initial conditions for the hidden compartments and the effective infection rates, giving us the possibility to perform a short-time forecast for the series of confirmed cases.

To summarize, in Part I we discuss the foundations of this thesis: network science and dynamical processes. In Part II, we tackle fundamental and theoretical problems on absorbing state phase transitions such as quasistationary analysis and symbiotic processes. In Part III, we develop models and methods applied to COVID-19 both in an attempt to characterize the SARS-CoV-2 spreading and to estimate the under-reporting in Brazil. Finally, in Part IV we draw our final remarks and conclusions.

Chapter 2

Bridges, yeast and the internet: Network theory

The representation of interactions between elements of a system as complex networks motivated the study of a myriad of phenomena using this framework as a substrate, including modeling bridges and roads [70], metabolic chains on yeast [71] or routers connections in the Internet [72]. In this chapter, we will discuss the main concepts regarding networks used in this thesis. Starting with a brief historical review, we are going to pass through the main concepts and metrics, ending with some discussions and properties on some networks classes. For the interested reader, references [2] and [73] are two excellent textbooks on this subject.

2.1 Historical overview

Königsberg, Eastern Prussia. While walking throughout one of the many bridges of the city, residents of the Prussian capital asked themselves: Would it be possible to create a route through the city that crosses each of its bridges exactly once? It was while answering this question that Leonhard Euler, in 1735, built the foundation of graph theory, the mathematical framework behind network science.

The city of Königsberg (now Kaliningrad, Russia), had a river that traversed the entire city, dividing it into some landmasses. Across this river, there were seven bridges that connected the different masses of land, as illustrated in Figure 2.1-(a).

The puzzle, as previously mentioned, was to provide a route crossing all bridges exactly once. The answer, provided by the Swiss mathematician Leonhard Euler, was that this route does not exist i.e., it is not possible to traverse all bridges once and only once.

We will not delve into further details of the proof, just the general abstraction procedure: starting from the physical system in 2.1-(a), the mathematician got rid of every nonessential information such as shapes and lengths of the landmasses and bridges, resulting in the connection pattern saw in 2.1-(b). This representation of the system was called **graph**, a set of elements connected between them and following certain rules. In the Königsberg example, the landmasses are called nodes and the bridges are edges. By analyzing this graph it was relatively easy to prove that this path does not exist.

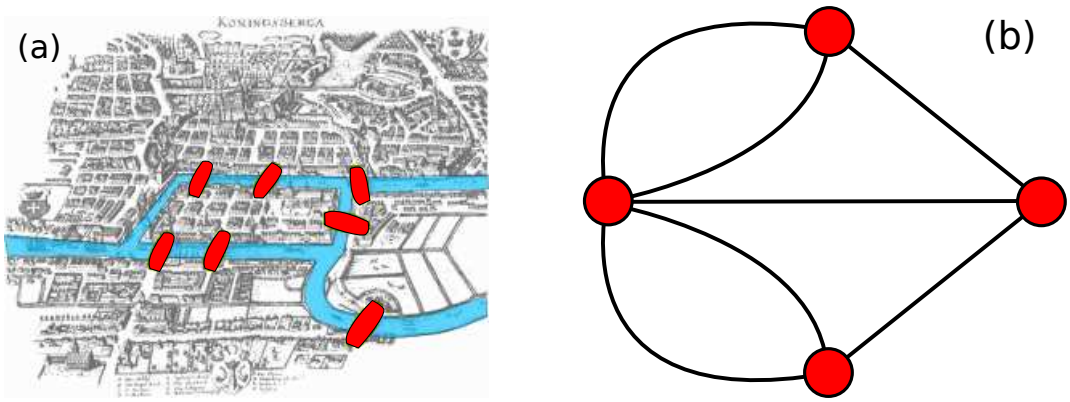


Figure 2.1: Abstraction steps followed by Leonhard Euler to solve the Seven Bridges of Königsberg's problem.

This reasoning was the major breakthrough that made possible the development of the area and is still used when extracting complex networks from real physical systems with the exception of a few cases, in which more details are needed. Figure 2.2 illustrates the extraction of a network from the Italian power grid. Starting from the identification of relevant components in (a) and going through a progressive abstraction in (b), (c) and (d). In the end, only the information regarding nodes and edges remains. It is worth discussing the connection between graphs and the structures that entitles this chapter, the networks. Usually, they are treated as synonyms in the literature, but there is a subtle distinction between them: a network is related to the physical system itself while a graph is the pure abstract representation of this system. Despite that, we will use both terms interchangeably.

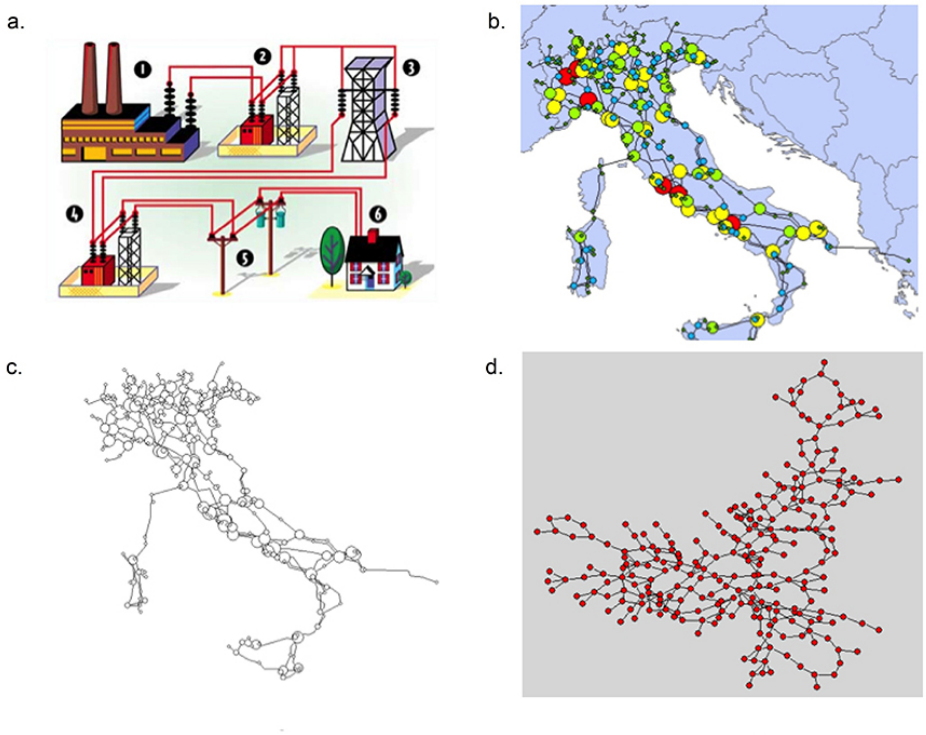


Figure 2.2: (a) Different elements of a power distribution grid. (b) Italian powergrid network, considering production and consumption in each node power station. (c) Spatial network extracted from (b). (d) Complex network in which all spatial information was disregarded, leaving only the vertices and edges. Adapted from [2].

Jumping in time, some of several important papers in the twentieth century regarding network science are due to Paul Erdõs and Alfred Rényi [74], that proposed a model to generate random networks, considered today as the simplest network to include stochasticity as an essential element [17], i.e., the connections are formed without any assumption on the nodes. Starting from a collection of N vertices, we connect each pair with probability p. For a sufficiently high p, a connected graph emerged from this rule, as can be seen in Figure 2.3. Those networks are known as Erdõs-Rényi networks (ER).

Some years later, Duncan Watts and Steven Strogatz [75] developed a model, hence called Watts-Strogatz model (WS), that was able to reproduce two important empirical properties of real networks: the high clustering coefficient and the small world effect, both to be discussed in Section 2.2. Starting from a periodic chain as in Figure 2.4-(a), in which each node is connected to $2m$ nearest neighbors, with probability p a link is rewired to a random vertice. By tuning the parameter p, the resulting object interpolates between a totally ordered system ($p = 0$) and a fully random network

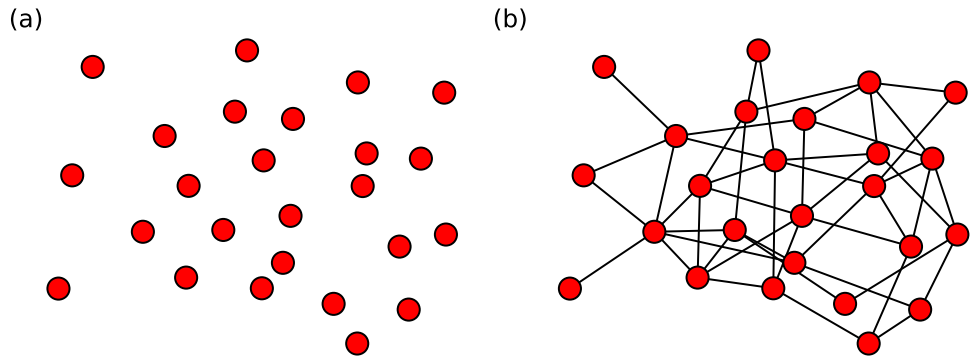


Figure 2.3: Example of Erdõs-Reyni network generated from a set of 25 nodes and $p = 0.1$

($p = 1$). Figures 2.4-(b) and (c) illustrate some cases.

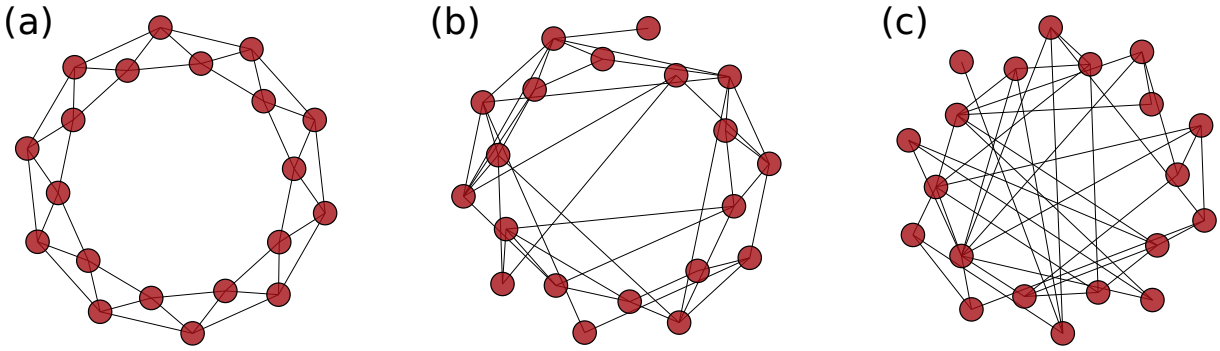


Figure 2.4: (a) Regular network with 20 nodes and 4 connections per vertice. Watts-Strogatz networks generated from (a) with (b) $p = 0.5$ and (c) $p = 1.0$.

Both of these models generate homogeneous networks, i.e., networks in which each node has roughly the same number of connections, pN for the ER case and $2m$ for WS networks. Comparing this property with real networks, we see that a lot of cases present a different reality: a highly heterogeneous connection pattern. To include this heterogeneity in the network models, Barabasi and Albert [76] developed an algorithm based on two mechanisms: growth of the network and preferential attachment. The first indicates that the network is dynamic, i.e, the number of nodes and connections changes as time evolves. The second mechanism acts as a connection rule: the newly added nodes have a higher chance to connect with highly connected ones. By putting these two mechanisms together, the authors were able to generate highly heterogeneous networks that described real datasets with much more precision than homogeneous models. We discuss this model further in the text.

With the popularization of network science and the advance of computational power, scientists were able to extract enormous datasets and study a variety of problems in physics, mathematics, social sciences and biology. In addition to the increase in applications of complex networks, the mathematics and new frameworks behind them evolved accordingly. Some examples include spatial networks [77], multiplex or multilayer networks [78] and higher order structures such as simplicial complexes [79]. Some of them will be further detailed in this text.

2.2 Basic concepts and characterization of networks

The two main elements that constitute a network are nodes or vertices, that represent the element of the network itself (landmasses, proteins or internet routers, for example) and edges or links, that represent the interactions and connections among them as in a bridge connecting two landmasses or a biochemical reaction among proteins in a metabolic network. Figure 2.5 shows an illustration of a small network. The nomenclature node and edge is commonly used in graph theory while vertice and links are mainly used in network theory. In this thesis we use both interchangeably.

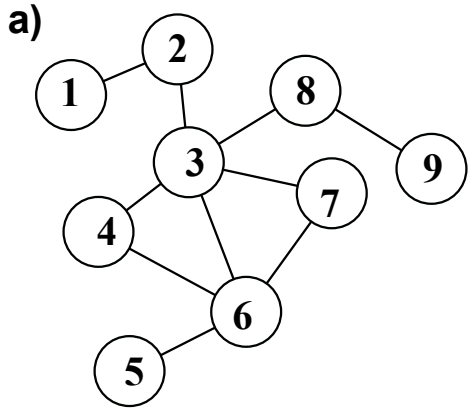
The usual mathematical and computational framework to study complex networks is using the adjacency matrix **A**, a N-squared matrix, N being the number of vertices, that encodes which vertices are connected to each other. The matrix A_{ij} follows this rules:

- $A_{ij} = 1$ if vertice i is connected to vertice j.
- $A_{ij} = 0$, otherwise.

The properties of **A** depend on the network structure itself. Regarding connections, if the edge $i \to j$ implies in an edge $j \to i$ for every pair of connected vertices ij, the network is said **undirected** and the adjacency matrix is symmetric, as in a sexual partnership network. Otherwise, as followers in Twitter, i following j does not imply that j follows i, the network is **directed**. In addition, complex networks are sometimes categorized according to density of connections: the graph is said **sparse** if the number of connections is way lower than the maximum allowed (a complete graph, consists of all nodes connected to each other, leading to $N(N-1)$ edges), otherwise, the

network is **dense**. Additional information concerning the intensity of connections can be stored on a weight matrix **W**, as for example, the number of people that travel daily between two cities. Matrices **A** and **W** provide all information about the network. Several measures and metrics can be extracted from them in order to describe and characterize the network.

a)



b)

	1	2	3	4	5	6	7	8	9
1	0	1	0	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0	0
3	0	1	0	1	0	1	1	1	0
4	0	0	1	0	0	1	0	0	0
5	0	0	0	0	0	1	0	0	0
6	0	0	1	1	1	0	1	0	0
7	0	0	1	0	0	1	0	0	0
8	0	0	1	0	0	0	0	0	1
9	0	0	0	0	0	0	0	1	0

Figure 2.5: (a) Representation of a small undirected network, showing the vertices and links and (b) the corresponding adjacency matrix A.

One of the most important metrics is the degree k of a vertice, which is the number of edges connected to a certain node i. In general, we can define two types of degree: k_{in}, that accounts for links pointing to a node ($j \to i$) and k_{out}, that accounts for edges exiting the vertice ($i \to j$). Both metrics can be calculated from the adjacency matrix:

$$k_{\text{in}}^{(i)} = \sum_j A_{ji} \qquad \text{and} \tag{2.1}$$

$$k_{\text{out}}^{(i)} = \sum_j A_{ij}. \tag{2.2}$$

For undirected networks, $k_{\text{in}} = k_{\text{out}}$ or simply k. From now on, all properties and metrics will be discussed considering undirected networks. Their generalization for the directed case is straightforward [73].

Since analyzing the degree of every node in large networks is little informative, a statistical description of the network can be constructed: considering a network with N vertices, we define the degree distribution of the network $P(k)$ as the probability that a randomly chosen vertice has degree k,

$$P(k) = \frac{N_k}{N}, \tag{2.3}$$

in which N_k is the number of vertices of degree k. We can calculate the n-th moment of this distribution as

$$\langle k^n \rangle = \sum_{k=1}^{k_{\max}} k^n P(k). \tag{2.4}$$

The first and second order moments are important to characterize the distribution giving the mean degree $\langle k \rangle$ and the variance ($\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$) of $P(k)$. Furthermore, the heterogeneity coefficient κ is defined as the moment ratio $\langle k^2 \rangle / \langle k \rangle$ with homogeneous networks having $\kappa \sim \langle k \rangle$ while highly heterogeneous ones are characterized by $\kappa \gg \langle k \rangle$.

To illustrate these and future concepts, we will consider a real network that will be studied further in this thesis: the commuting network for inter-municipality mobility in Brazil, constructed from official government data [62]. In this network, two cities are connected if the recurring mobility between them is non zero, i.e., if persons travel daily between the two cities to work or study. This network has $N = 5565$ nodes, $E = 99967$ edges and will be referred to as **COMM** from now on. Figure 2.6 shows $P(k)$ for this network in comparison with ER and WS networks of the same N and $\langle k \rangle$. Note that the COMM network has a prominent right tail while the others have more localized distributions. These differences will be further discussed in Section 2.3.

2.2.1 Degree correlations

Another important network metric is the nearest neighbor average degree K_{nn} of a vertice, defined as

$$K_{\text{nn}}^{(i)} = \frac{1}{k_i} \sum_{j=1}^{N} k_j A_{ij}, \tag{2.5}$$

which yields information about the connectivity tendencies of the vertices, if they are more likely to connect with their akins (in terms of degree) or their dissimilar. To further address this relationship, we can average $K_{\text{nn}}^{(i)}$ accordingly to the degree of

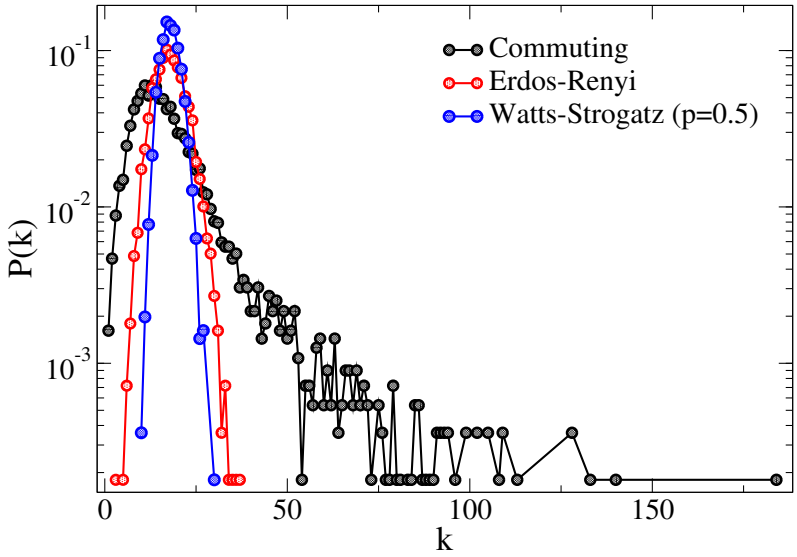


Figure 2.6: Degree distribution $P(k)$ for the commuting network for Brazil's inter-municipality mobility in comparison with equivalent (same N and $\langle k \rangle$) ErdõsRényi and Watts-Strogatz networks.

their node

$$K_{\mathrm{nn}}(k) = \frac{1}{N_k} \sum_{i=1}^{N} K_{\mathrm{nn}}^{(i)} \delta_{k,k_i}, \tag{2.6}$$

in which δ_{k,k_i} is the Kronecker's delta symbol. If the nodes are more likely to connect to their alike, $K_{\mathrm{nn}}(k)$ is an increasing function on k and the network has **assortative mixing** or positive degree correlation. Conversely, if higher degree nodes connect more likely to low degrees, $K_{\mathrm{nn}}(k)$ is a decreasing function on k and the network has **disassortative mixing** or negative degree correlation. Lastly, if there are no tendencies, $K_{\mathrm{nn}}(k)$ is approximately constant and the network is **uncorrelated**. In Figure 2.7, we show a schematic representation of $K_{\mathrm{nn}}(k)$ for these three correlation patterns.

Uncorrelated networks are widely applied to analytical calculations of epidemic models and other dynamical processes and are used as a first approach in a variety of problems. In this particular case, it is possible to calculate $K_{\mathrm{nn}}(k)$ analytically and in a closed form. First, we need to write $K_{\mathrm{nn}}(k)$ as

$$K_{\mathrm{nn}}(k) = \sum_{k'} k' P(k'|k), \tag{2.7}$$

in which $P(k'|k)$ is the conditional probability of a vertice of degree k being connected with a vertice of degree k'. To find a expression for this probability, let us define the edges matrix E, whose elements satisfy the following relations:

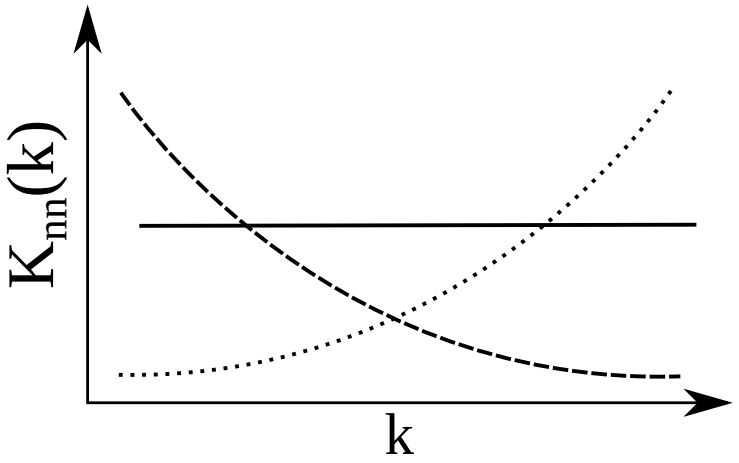


Figure 2.7: Schematic representation of nearest neighbor average degree K_{nn} for the three correlation patterns (full line - uncorrelated; dashed line - disassortative mixing; dotted line - assortative mixing).

- If $k \neq k' \rightarrow E_{kk'} =$ number of edges among vertices of degrees k and k'.
- If $k = k' \rightarrow E_{kk'} = 2\times$ (number of edges among vertices of same degree k)

Thus, $P(k'|k)$ takes the form

$$P(k'|k) = \frac{E_{kk'}}{kN_k}. \tag{2.8}$$

Next, we need to calculate the probability that a randomly chosen edge connects a vertice of degree k to a node with degree k'

$$P(k',k) = \frac{E_{kk'}}{\mathcal{E}} = \frac{E_{kk'}}{\langle k \rangle N}. \tag{2.9}$$

in which $\mathcal{E}$ is $2\times$ the number of edges of the network. By rearranging Equations (2.8) and (2.9), we can relate $P(k'|k)$ and $P(k',k)$

$$P(k',k) = \frac{kN_kP(k'|k)}{\langle k \rangle N}. \tag{2.10}$$

Since we have no hypothesis on k or k', we can interchange k and k' without penalty. Beyond that, we are considering undirected networks, so $P(k',k) = P(k,k')$. Thereby,

$$k'N_{k'}P(k|k') = kN_kP(k'|k). \tag{2.11}$$

Finally, summing in k on both sides of the equation above (by hypothesis, $P(k'|k)$ does

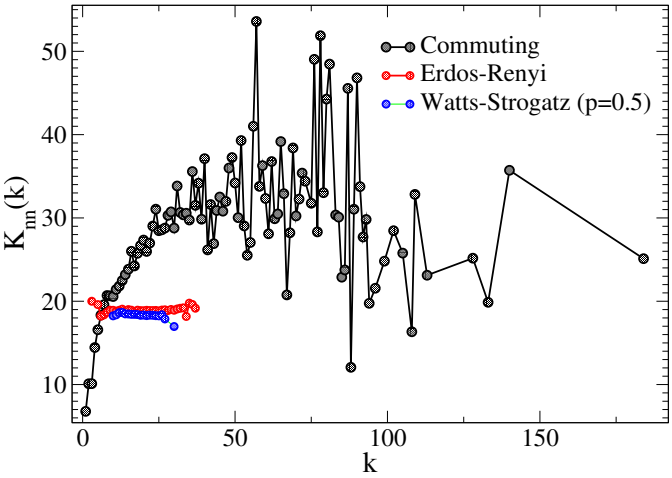


Figure 2.8: Nearest neighbor average degree $K_{\mathrm{nn}}(k)$ for the commuting network for Brazil's inter-municipality mobility in comparison with equivalent (same N and $\langle k \rangle$) Erdõs-Rényi and Watts-Strogatz networks.

not depend on k) results in

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle}, \tag{2.12}$$

which is introduced in Equation (2.7) to give

$$K_{\mathrm{nn}}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{2.13}$$

Figure 2.8 shows $K_{\mathrm{nn}}(k)$ for COMM network. It can be noticed that this graph is assortative for small k and approximately uncorrelated for large k, despite the fluctuations. In addition, we also have $K_{\mathrm{nn}}(k)$ for ER and WS networks of same N and $\langle k \rangle$. As expected, K_{nn} indicates that they are uncorrelated.

Another way to quantify the overall network topological correlations is the Pearson coefficient [73], defined as

$$P = \frac{\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{N\langle k \rangle}\right) k_i k_j}{\sum_{ij}\left(k_i \delta_{ij} - \frac{k_i k_j}{N\langle k \rangle}\right) k_i k_j}, \tag{2.14}$$

that lays within the interval $-1 \leq P \leq 1$, in which positive, null and negative values are related to assortative, neutral and disassortative mixing, respectively. The Pearson coefficient for COMM network is $P = 0.07$, implying in a slightly assortative mixing. For the sake of comparison, the coefficients for WS and ER networks are $P^{\mathrm{WS}} = -0.002$ and $P^{\mathrm{ER}} = 0.0001$.

2.2.2 Clustering

Another important concept is the idea of aggregation in networks. There are many different approaches to this concept, but the most used one is based on identifying triangles in the network: three vertices i, j and k that are connected between them. Figure 2.9-(a) illustrates this phenomenon. The **clustering coefficient** C is a metric that characterizes the neighborhood structure of a vertice i [17] by accounting to the number of triangles in it. One definition is the following:

$$C(i) = \frac{2e_i}{k_i(k_i - 1)}, \tag{2.15}$$

with e_i being the number of edges among neighbors of i and $k_i(k_i - 1)/2$ is the maximum possible value of e_i for a vertice with k_i links. Defined as it is, $C(i)$ lies in the interval $[0, 1]$; the higher $C(i)$, the higher the connectivity among its neighbors. The adjacency matrix yields a formal way to calculate e_i:

$$e_i = \frac{1}{2}\sum_{j,l} A_{ij}A_{jl}A_{li}. \tag{2.16}$$

Similarly to nearest neighbor degree K_{nn}, the clustering coefficient of vertices with same connectivity can be averaged as

$$C_k = \frac{1}{N_k}\sum_{i} C(i)\delta_{k,k_i}. \tag{2.17}$$

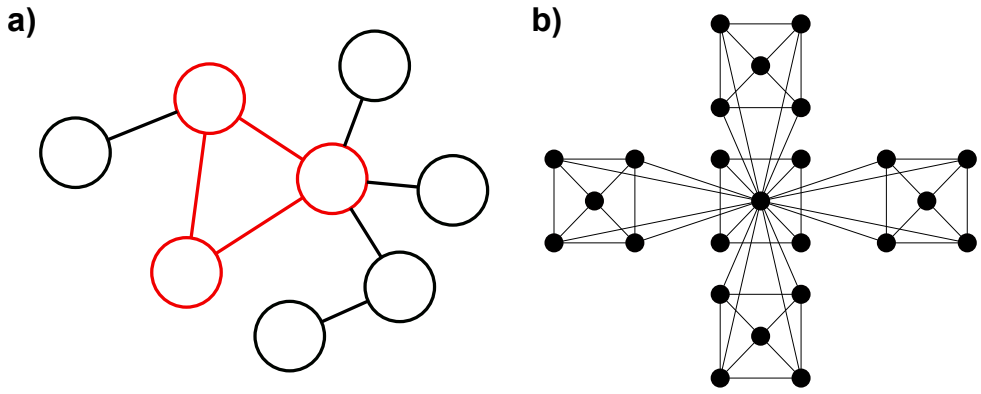


Figure 2.9: (a) Illustration of a triangle (in red) on a simple graph. (b) Example of a hierarchical network structure.

This dependency on k can reveal important aspects of the global structure of the network. Two particular cases are important: $C(k) \approx$ c, indicates that the network is

uncorrelated while $C(k) \sim k^{-\omega}$, $\omega > 0$ suggests a hierarchical and modular structure of the networks vertices [80]. Small portions of the network that are highly clustered and have low k are connected by nodes with high connectivity and low C. Figure 2.9-(b) illustrates a hierarchical network. Additionally, we can calculate the global clustering coefficient of the network as

$$\langle C \rangle = \frac{1}{N} \sum_i C(i), \tag{2.18}$$

that gives an average insight about aggregation on the graph. In Figure 2.10 we present $C(k)$ for COMM network in comparison with ER and WS networks with the same N and $\langle k \rangle$. ER networks have constant and small $C(k) \sim 1/N$ due to the completely random connections. The WS model presents a power law behavior due to vestiges of the initial ordering, shown in Figure 2.4-(a). Finally, the commuting network also has $C(k)$ as a power law, indicating some level of hierarchical structure.

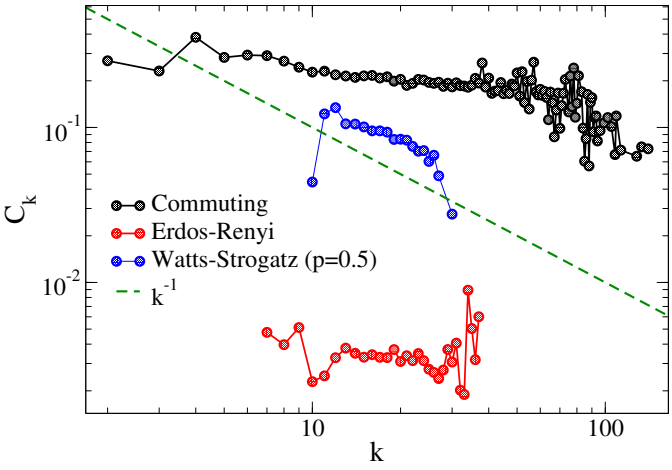


Figure 2.10: Clustering coefficient C_k for the commuting network for Brazil's inter-municipality mobility in comparison with equivalent (same N and $\langle k \rangle$) Erdõs-Rényi and Watts-Strogatz networks.

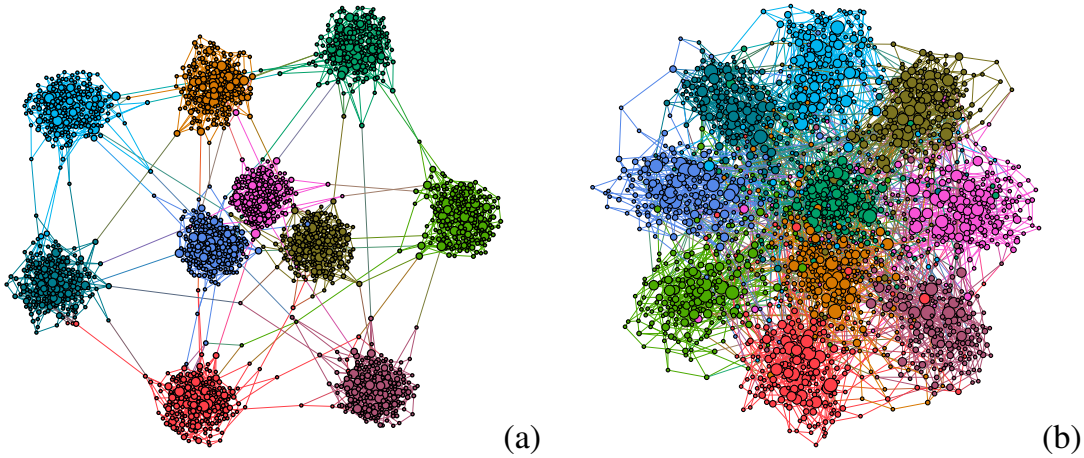
2.2.3 Modularity

In many real networks, regions appear that are locally denser, with vertices highly connected to each other while weakly connected outside, defining **communities** or **modules** within the graph. Formally, we can partition the N vertices of a network into M arbitrary communities, denoting the group to which a vertex i belongs as g_i. The main problem is to execute this partitioning in order to represent as faithfully

as possible the inherent community structure in the original graph. One of the main approaches is by using the modularity coefficient Q [2, 73], defined as

$$Q = \frac{1}{N\langle k\rangle}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{N\langle k\rangle}\right)\delta_{g_i g_j}, \tag{2.19}$$

in which $\delta_{g_i g_j}$ is the Kronecker delta symbol. The modularity lays in the range $[0,1]$, the larger, the better partitioned is the set $\mathcal{G} = \{g_1, g_2, ..., g_M\}$. Thus, by finding $\mathcal{G}$ that maximizes Q, one finds the optimal community structure of the network. Although theoretically simple, the evaluation of this optimal structure is computationally costly and difficult, with many algorithms developed to ease this task [2]. Figure 2.11 show some synthetic networks with high modularity.



(a) (b)

Figure 2.11: Examples of synthetic networks with high modularity. (a)Loosely and (b)Heavy connected modules. Different colors represent different communities. Courtesy of W. Cota [40].

2.2.4 Paths and distances

In network theory, it is highly common to neglect the space in which the network is embedded, abstracting the real physical system into something mathematical, in which the only information is who is connected to whom. In some cases, the physical structure does not exist such as in social or biochemical networks. So, the definition of **distances** and **paths** in complex networks is different from the usual euclidean metrics.

A path in a network is defined as an ordered sequence of successively connected vertices starting in a vertice i and ending in a node j. The length of this path ℓ is

the number of edges crossed to go from i to j through the path. Note that exists infinite paths among two nodes with a variety of lengths. Figure 2.12 illustrates some different paths in a sample network.

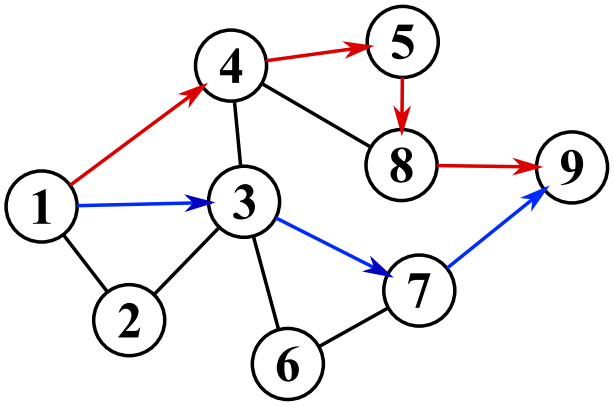


Figure 2.12: Scheme of a simple network highlighting two paths between vertices 1 and 9: blue path ($1 \to 3 \to 7 \to 9$) with length $\ell = 3$ and red path ($1 \to 4 \to 5 \to 8 \to 9$) with length $\ell = 4$.

An important concept regarding paths in a network is the **shortest path** or **distance** between two nodes ℓ_{ij}, which are the length of the shortest path(s) between two nodes i and j [1]. Starting from this definition, an important metric about the network is the **average shortest path** $\langle \ell \rangle$

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_i \sum_j \ell_{ij}, \tag{2.20}$$

which is an average between the distances of all pairs of nodes of the network. Also, the diameter of the network is defined as $D = \max_{ij}\{\ell_{ij}\}$.

Here, another big difference between networks and regular (Euclidean) lattices emerges: the dependency between $\langle \ell \rangle$ and N. For the latter, this relation is straightforward: $\langle \ell \rangle$ is proportional to the lattices size

$$\langle \ell \rangle \sim L \sim N^{1/d}, \tag{2.21}$$

in which L is the size of the lattice, N is the number of nodes and d is the euclidian dimension of the lattice. It is important to note that for complex networks, size and number of nodes are synonyms and are represented by N while for lattices there is a distinction between the two quantities. Taking a square lattice ($d = 2$) as example,

[1] $\ell_{ij} = \infty$ if there is no path between i and j

a size $L = 100$ implies in 100 nodes on each side of the square, giving a total of $N = 100^2$ nodes.

For complex networks, the reasoning is a bit more complicated, but Reference [2] presents an elucidating way to obtain this relation. Considering a random network with N vertices and mean degree $\langle k \rangle$, a node from this network has, in average, $\langle k \rangle$ vertices at distance $\ell = 1$, $\langle k \rangle^2$ vertices at distance $\ell = 2$, and so on. Thus, the number of vertices at a distance ℓ or less is given by

$$N(\ell) = 1 + \langle k \rangle + \langle k \rangle^2 + \langle k \rangle^3 + ... + \langle k \rangle^\ell = \frac{\langle k \rangle^{\ell+1} - 1}{\langle k \rangle - 1}. \tag{2.22}$$

If we assume that $\langle k \rangle \gg 1$, we can neglect the (-1) terms and find a much simpler relation: $N(\ell) \approx \langle k \rangle^\ell$. However, in a real network we are limited by the number of vertices N. Therefore,

$$N(D) = N \approx \langle k \rangle^D \rightarrow D \approx \frac{\ln N}{\ln \langle k \rangle}. \tag{2.23}$$

Since we assumed an average behavior during the calculations and D depends only on the largest paths, Equation (2.23) better describes the dependency between N and $\langle \ell \rangle$. Thus, $\langle \ell \rangle \sim \ln(N)$ in random networks. Comparing the behavior of $\langle \ell \rangle$ in networks and lattices in Figure 2.13, we see that distances are much smaller in the former. This is called **small world phenomenon**, proposed by Stanley Milgram in 1967 [81] and further studied in the following decades.

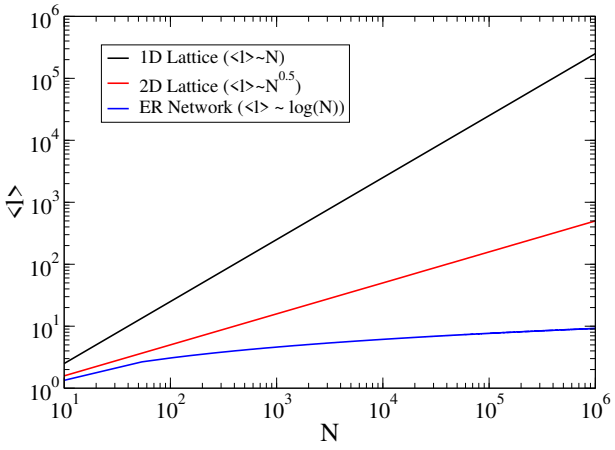


Figure 2.13: Relation between the average shortest path $\langle \ell \rangle$ and network size N for different topologies and geometries.

The experiment designed by Milgram had the purpose to measure distances in

contact networks. He wrote some letters addressed to residents in Massachusetts and asked citizens of Omaha and Wichita to try to deliver these letters by passing it to acquaintances. From the 296 written letters, 60 reached their destination with an average number of contacts of 5.2. Later, the term **theory of six degrees of separation** was wrought, which stated that every person is connected to any other by an average distance of six. Further studies involving social network as the Facebook claimed that actually, this number is even lower, about 3 or 4 [82, 83].

From these studies and experiments, emerged a lot of indexes in different media [84] to evaluate distances among people in some collaboration context:

- Erdos number: collaboration distance to the Hungarian mathematician Paul Erdos, famous for the high number of authored publications during his lifetime: ~ 1400.
- Bacon's number: distance to american actor Kevin Bacon in the co-stardom network, i.e., the collaboration graph of film actors.
- Sabbath's number: distance to any member of the British heavy metal band Black Sabbath in the musical collaboration network.

A few notable people have these three indexes: Carl Sagan, Richard Feynman, Brian May and Natalie Portman. The one who writes to you has only an Erdos number of 6. All this exotic examples allow to show that the small world effect permeates our reality in different environments and therefore, presents itself as a fascinating subject. More examples of statistics and metrics for some real networks are presented in Table 2.1.

2.2.5 Motifs

Besides the microscopic description of networks, focusing on properties of the vertices itself, the organization of nodes into larger structures, known as motifs, are also important for the throughout characterization of a network and, therefore, to enlighten the dynamical processes evolving on them [86, 87, 88].

One of the most basic structures that arise in a network is the **clique**, a densely connected fraction of the network that has an important role in some dynamical processes

Network	N	$\mathcal{E}$	$\langle k \rangle$	κ	P	C	$\langle \ell \rangle$	D
Reactome†	5873	145778	49.6	143	0.24	0.61	5.21	25
Air Traffic†	1226	2408	3.92	7.35	-0.015	0.07	6.92	18
Google†	15763	148585	18.9	902	-0.14	0.52	3.51	8
US PowerGrid†	4941	6594	2.67	3.87	0.003	0.08	19.9	47
Facebook (links)†	63392	816886	25.8	88.0	0.18	0.21	4.77	7
ErdõsReyni	5565	49788	17.9	18.9	10^{-4}	0.003	4.29	6
Watts-Strogatz	5565	50067	18.0	18.4	-0.002	0.009	4.42	6
Commuting	5565	49984	18.0	27.5	0.007	0.22	7.88	14

Table 2.1: Structural properties for some real and synthetic networks. From left to right: name of the network, number of vertices N , number of edges $\mathcal{E}$, average degree $\langle k \rangle$, heterogeneity coefficient κ, Pearson coefficient P, clustering C, average shortest path $\langle \ell \rangle$ and diameter D. †Networks extracted from [85].

(to be discussed further in the thesis). Formally, a clique is the maximal subgraph of fully connected vertices, i.e., every node in this set is linked to all others [73].

Another structure that is closely related to cliques are the ***k*-cores**, which are the maximal subgraph $\mathcal{G}_k$ such that $k_i \geq k, \forall i \in \mathcal{G}_k$ [73]. In particular, the **maximum *k*-core**, which is the largest k that results is a non-null $\mathcal{G}_k$ for a given network, is also densely connected and crucial to several dynamical processes on it [89]. Usually, k-cores are more studied and used than cliques because the former are relatively easier to find. The following algorithm can find all k-cores of a network:

i). Set $k_0 = 2$.

ii). Remove all vertices that have $k < k_0$ and all edges connected to them.

iii). If after step 2, new vertices with degree $k < k_0$ appears, remove them as well. If all remaining vertices have degree $k \geq k_0$, this set is the k_0-core of the network. All removed vertices in the steps (ii) and (iii) constitutes the $(k_0 - 1)$-shell.

iv). $k_0 \rightarrow k_0 + 1$

v). Iterate steps (ii) to (iv) until all vertices are removed.

To illustrate the different shells in a graph, we will consider a subgraph of COMM network regarding the Ceará (CE) state. This subgraph is a fraction of the original network that can be clearly visualized, as shown in Figure 2.14.

To discuss the next structure, we need to define **hubs** and **outliers**. In a formal way, the definition of a hub in thermodynamic limit ($N \rightarrow \infty$) is any vertice with

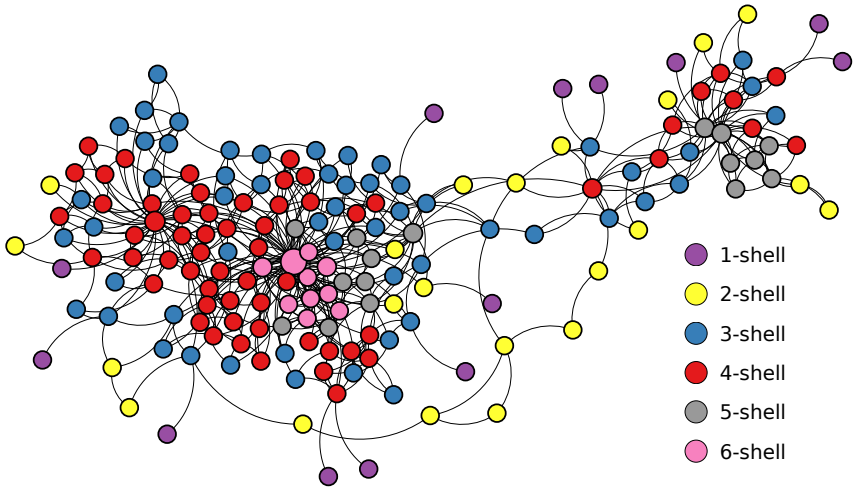


Figure 2.14: An example of k-shells of every vertice in the CE mobility network. Different colors indicate different values of $k-$shells. The visualization was generated using **graph-tool** package for Python 3.8 [90].

$k \sim N^{\alpha}$, $0 < \alpha < 1$. For finite networks, the definitions may change but it is widely considered that hubs are nodes that have $k \gg \langle k \rangle$. If the degree of a hub is larger than the rest of the network, except possibly a few other hubs, this vertice is an outlier. We can now discuss **stars** in networks. This structure consists of a hub and its neighbors, forming a star-shaped subgraph. The hub is called the center of the star while the neighborhood is called leaves. Hubs frequently are shortcuts between different parts of the network, implying that their presence lowers the average distance in the graph. Usually, a combination of those motifs appears in different parts of the network, as illustrated in Figure 2.15.

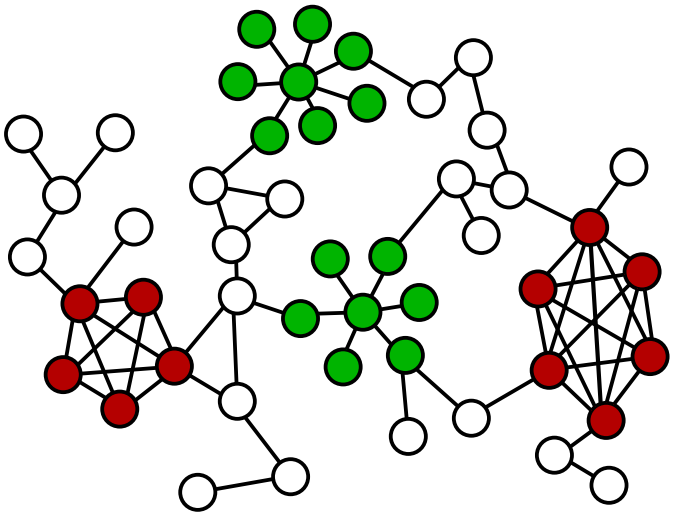


Figure 2.15: Schematic representation of a small graph, highlighting its stars (in green) and the cliques (in dark red).

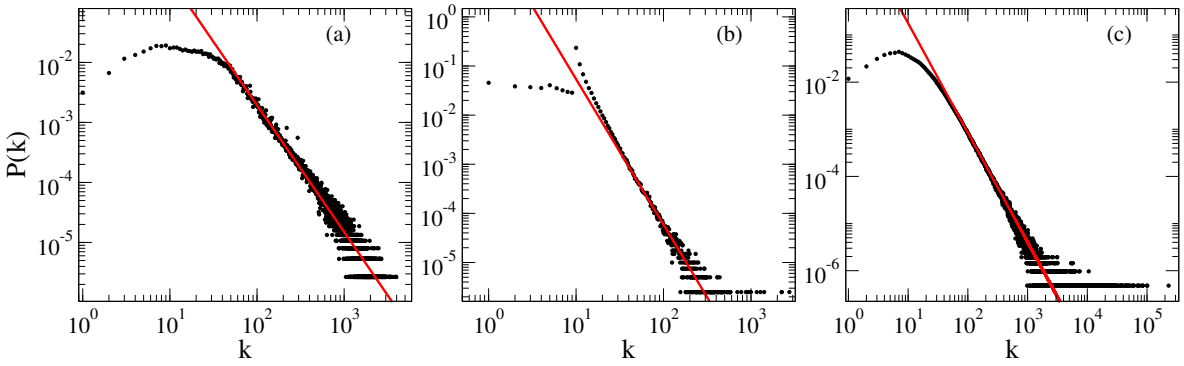


Figure 2.16: Degree distribution for some real networks. Full lines are asymptotic power laws. (a) Actor collaboration ; $\gamma = 2.11$. (b) Amazon ; $\gamma = 2.94$. (c) Pages in Wikipedia ; $\gamma = 2.24$. Descriptions of those networks are provided in [85].

2.3 Scale-free networks

Now we turn our attentions to the aforementioned scale-free (SF) networks, a class of graphs that appears in a wide range of contexts such as technological, biological, geographical and social networks [85]. There are various definitions of SF networks, some are strict while others are not. However, the most used definition is: a scale-free network is any graph that present large degree fluctuations, orders of magnitude larger than expected values [17]. Usually, they can be described by power law $P(k) \sim k^{-\gamma}$ degree distributions with $2 < \gamma < 3$. Figure 2.16 shows $P(k)$ for some real scale-free networks.

The mechanisms beyond this structures remained a mystery until Albert-László Barabási and Reka Albert , whom in 1999 [91] proposed a physical mechanism for the formation of these networks, based on two properties

- **Growth of the network:** starting from a small cluster of nodes, new vertices are added, one at a time.
- **Preferential attachment:** the new vertices have a higher probability to connect with more connected nodes.

By constructing a algorithm with these properties and connection probability

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}, \tag{2.24}$$

the authors were able to generate an synthetic network with asymptotic degree distri-

bution $P(k) \sim k^{-3}$, know as Barabasi-Albert network. Additional mechanisms were later proposed to generate networks with different degree exponents [92].

2.3.1 The scale-free property

In this section, we will assume a $P(k) = Bk^{-\gamma}$ limited between a minimum $k_{\min}$ and a maximum $k_{\max}$, $k_{\max} \gg k_{\min}$. Considering these bounds, we can obtain the normalization constant

$$\sum_{k=k_{\min}}^{k_{\max}} P(k) = 1 \rightarrow B \sum_{k_{\min}}^{k_{\max}} k^{-\gamma} = 1. \tag{2.25}$$

By approximating the sum as a integral, we can solve this equation for B

$$B \int_{k_{\min}}^{k_{\max}} k^{-\gamma} dk = 1 \rightarrow B \simeq (\gamma - 1) k_{\min}^{-\gamma+1}, \tag{2.26}$$

and calculate the moments of $P(k)$ as

$$\langle k^m \rangle = \sum_{k=k_{\min}}^{k_{\max}} k^m P(k) \simeq B \int_{k_{\min}}^{k_{\max}} k^{-\gamma+m} dk = B \frac{k_{\max}^{-\gamma+m+1} - k_{\min}^{-\gamma+m+1}}{-\gamma + m + 1}. \tag{2.27}$$

By considering $k_{\max} \gg k_{\min}$, we can write

$$\langle k^m \rangle \sim k_{\max}^{-\gamma+m+1}. \tag{2.28}$$

Another important concept is the idea of **cutoffs**, i.e., constraints on how arbitrarily large a certain degree can be in a given network model. For example, in a network with N vertices, the maximum allowed degree is $k_c = N - 1$, if multiple and auto connections are discarded. However, for many practical purposes, k_c is usually much smaller than N, either by imposition of the system or by probabilistic constraints i.e., the chance of a vertice with $k > k_c$ is really small.

To estimate the magnitude of k_c for degree distributions described by analytical functions, we define $\phi(k)$ as the number of vertices that have degrees higher or equal than k. Formally, it can be expressed as

$$\phi(k) = N \int_k^N P(k') dk'. \tag{2.29}$$

By setting $\phi(k_c) = 1$ and solving Equation (2.29) for k_c [93], we can estimate the **natural cutoff** of the network, i.e., a constraint for k purely associated with the degree distribution. For scale-free networks, this calculation yields the following scaling law

$$k_c \sim N^{\frac{1}{\gamma-1}}. \tag{2.30}$$

Thus, for synthetic networks with power law degree distribution, the maximum degree will have the same magnitude of k_c unless more rigid cutoffs are imposed. From Equation (2.30), we can analyze the behavior of the moments of $P(k)$ as we approach the **thermodynamic limit** $N \to \infty$. Considering $k_{max} = k_c$, the first two moments scale with N as follows

$$\langle k \rangle \sim N^{-\left(\frac{\gamma-2}{\gamma-1}\right)} \qquad \text{and} \qquad \langle k^2 \rangle \sim N^{-\left(\frac{\gamma-3}{\gamma-1}\right)}. \tag{2.31}$$

Thereby, $\langle k \rangle$ and $\langle k^2 \rangle$ diverges for $\gamma < 2$ and $\gamma < 3$ as $N \to \infty$, respectively. In the scale-free regime, we see that $\langle k^2 \rangle$ and therefore, the fluctuations of k, diverges. For the sake of comparison, we have $\sigma_{ER} \sim \langle k \rangle^{1/2}$ for random networks generated using the ER model and $\langle k \rangle$ is a good representative scale for the network degrees. Conversely, for $2 < \gamma < 3$ one expects vertices with arbitrarily large degrees with a non negligible probability due to the divergence of σ_{SF}. In this way, these networks do not have a typical representative for the degree of an arbitrarily chosen vertice, justifying the jargon scale-free.

2.3.2 Constructing synthetic scale-free networks

As previously mentioned, uncorrelated networks, despite being unrealistic in comparison with most of real systems, plays a fundamental role as a null model to confront analytical approximations of dynamic processes [94]. We have discussed several random network models in which no explicit assumptions on the network properties were adopted and the connections follow specific random rules. This lack of control over network properties can be a problem for the investigation of specific characteristics of a network.

To circunvent these limitations and have more control over the generated network,

we will discuss the Uncorrelated Configuration Model (UCM) [94], derived from the so called Configuration Model (CM) [95], both used to generate networks with predetermined $P(k)$.

In the CM, we assign k_i unconnected stubs according to a given degree distribution $P(k)$ ($k_{\min} < k < k_{\max}$) to each vertex of the network, constraining $\sum_i k_i$ to be even, so that no empty connections are left at the end of network's construction. Then, two stubs are randomly selected and its vertices are connected, avoiding multiple (i linked to j more than once) or self-connections (i linked to itself)) until all stubs are connected. For degree distributions whose $\langle k^2 \rangle$ is finite, the natural cutoff of the network is usually enough to suppress the chance of self and multiple connections. Therefore, for these distributions, the CM is able to generate a uncorrelated network without problems.

For the case of scale-free networks this is no longer true, since the amount of hubs and outliers implies in non negligible attempts of self and multiple connections. These attempts introduce disassortative correlations in the network since rejecting connections between hubs tends to connect them with lower degree vertices.

Catanzaro et al [94] showed that a sufficient condition to destroy the degree correlations on the usual CM for scale-free networks was to impose a more rigid cutoff $k_{\max} \sim (\langle k \rangle N)^{1/2}$ [93]. In this way, we limit exaggerated outliers and the correlations are again negligible. This model is called UCM model and is computationally implemented by following the steps

i). Draw the degrees of each vertice following the desired distribution $P(k)$ imposing the conditions: $k_{\min} < k_i < N^{1/2}$ and $\sum k_i$ even;

ii). Assign to each node a number of unconnected stubs corresponding to its degree;

iii). Randomly select two stubs, excluding self and multiple connections;

iv). Connect the selected vertices;

v). Repeat steps (iii) and (iv) until all network is connected. In case self or multiple connections are unavoidable, the sample is discarded and the process is restarted.

In Figure 2.17-(a) we plot degree distributions $P(k)$ for scale-free networks constructed by the UCM algorithm with different degree exponents, in comparison with the asymptotic power laws. In Figure 2.17-(b) we present $K_{\mathrm{nn}}(k)$ for the same networks as in (a). It can be noticed that K_{nn} is approximately constant (except for fluctuations), pointing that the networks are uncorrelated.

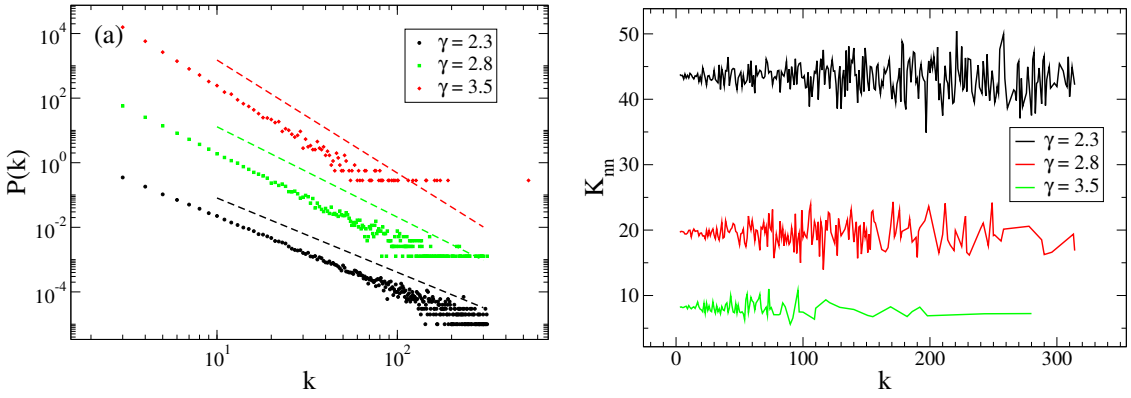


Figure 2.17: (a) Degree distributions for scale-free networks with $N = 10^5$ constructed with the UCM algorithm and different exponents ($P(k) \sim k^{-\gamma}$) ; the curves were vertically shifted to avoid overlapping ; dashed lines represents the asymptotic power laws. (b) $K_{\mathrm{nn}}(k)$ for the networks presented in (a).

2.4 Other network classes

In this section we discuss additional concepts in complex networks that will be eventually used throughout this thesis.

An important idea that so far has not been explored is the temporality in networks, i.e., graphs that change as time goes by, following some rule. A way to incorporate this aspect in network science are the concepts of **annealed** and **quenched** networks [73]. The former are probabilistic graphs, in which connections exists only instantaneously and in a probabilistic sense, preserving statistical properties such as $P(k)$ and $P(k'|k)$, defined previously. The latter refers to static networks, that do not change in time. All concepts introduced until here are for quenched networks and may need adaptations to be extended for annealed ones. The adjacency matrix A_{ij} becomes difficult to work with, due to connections that are made and unmade all the time. Thus, we can construct a probabilistic matrix A_{ij}^{ann} which indicates the probability that a vertex i is connected to j. For uncorrelated networks, the elements of A_{ij}^{ann} are given by

$$A_{ij}^{\text{ann}} = \frac{k_i k_j}{N\langle k \rangle}. \tag{2.32}$$

Annealed networks are extremely important for theoretical calculations on complex networks because the mean-field theories, discussed in Section 3.3.1, are commonly obtained considered this limit [87].

Another aspect of graph theory that was mentioned before is the degree heterogeneity of networks. Scale-free networks are the benchmarks for heterogeneous graphs, but even models with finite variance and exponential tails for the degree distribution such as WS and ER, commonly assumed as homogeneous networks, still present some level of degree heterogeneity. However, in this thesis we will consider the **random regular networks (RRNs)** as null model for homogeneity. All nodes of the network have same $k = m$ and the connections are randomly made according to CM model, described in Section 2.3.2

Finally, **multiplex graphs** [96, 97] are an extension of complex networks by constructing multiple layered substrates using the same set of nodes. In this systems, each vertice has a representation in each layer. However, the connections of each node may change within layers, i.e., if vertice i is connected to j and l in some layer, in another it may be connected to p and q instead. These systems are extremely useful to model ecosystems such as social relations or transportation networks [98]. Figure 2.18 illustrates some multiplex networks.

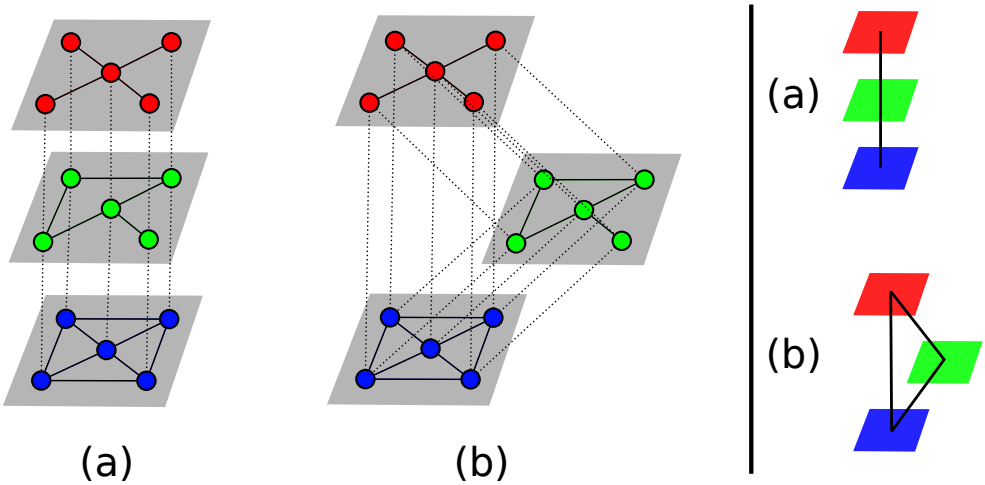


Figure 2.18: Schematic illustration of some three-layered networks. Right panels are microscopic representation of those networks while the left ones are a macroscopic relation between different layers. Adapted from [99].

Chapter 3

Dynamical processes: Applications to spreading phenomena

In this chapter, we discuss the main mathematical, physical and biological basis for the models studied and developed throughout the thesis. All this framework can be grouped in the dynamical processes area, a highly interdisciplinar subject [17, 100] that deals with phenomena that have many interacting elements. Initially, we cover some general concepts and classes of processes that will be important during this thesis, making connections to statistical physics. We then discuss the fundamentals of mathematical epidemiology, including theoretical analysis, computational simulations and applications.

3.1 General concepts

To start analyzing a dynamical process, we need to model the different states that each microscopic element of the system (molecules, individuals, magnets, airports, etc) can assume. This can be done by assigning a state variable σ_i to each one of the system's components. For the sake of mathematical and computational calculations, these states are usually mapped into integer numbers if they are discrete or to a certain range if continuous. Without loss of generality, let us assume that for a given process, there are Γ states, i.e., $\sigma_i = 1, 2, 3, ..., \Gamma$, and N elements. Thus, the macroscopic state σ of the system is given by the set $\sigma = \{\sigma_1, \sigma_2, ..., \sigma_N\}$. The dynamics of the

system is treated as transitions $\sigma \rightarrow \sigma', \sigma \neq \sigma'$, that alters one or more variables σ_i, following some rules that depend on the studied process. From now on, we will refer to the system's elements as nodes or vertices, since our goal is to study these dynamical processes on complex networks. However, this approach remains valid for other structures.

Since we are interested in systems with a large quantity of nodes, a probabilistic approach, highly used in statistical physics, is more suitable to describe the process. So, we introduce $P(\sigma, t)$ as the probability that the system is in the σ state at time t and $W_{\sigma \rightarrow \sigma'}$ as the transition rate between states σ and σ'.

An important class of dynamical processes are the **Markovian processes**, in which P at time t_n is fully determined by the probability distributions at t_{n-1}, i.e., the system's state at the next time step depends only on the immediately preceding state. Many real systems can be described by Markovian processes at some level, such as dissociation of gas molecules and Brownian motion.

Considering a Markovian dynamical process, we can write a differential equation to express the dynamic of $P(\sigma, t)$ in terms of the transition rates $W_{\sigma \rightarrow \sigma'}$

$$\frac{\partial P(\sigma, t)}{\partial t} = \sum_{\sigma'} \left[W_{\sigma' \rightarrow \sigma} P(\sigma', t) - W_{\sigma \rightarrow \sigma'} P(\sigma, t) \right], \tag{3.1}$$

that describes the probability flux incoming or outgoing of a determined microstate σ. This is the so-called **master equation** of the process and it allows us to calculate expected values of any relevant macroscopic quantity of the system. To exemplify, let us consider a variable B that is a function of the microstate σ. Then, we can track the evolution of the expected value of B as

$$\langle B(t) \rangle = \sum_{\sigma} B(\sigma) P(\sigma, t). \tag{3.2}$$

However, its not possible to calculate the formal solution of the master equation for most processes. Thus, alternative approaches are needed such as numerical calculations or computational simulations. The focus of most of the works in this thesis are simulations. Therefore, we detail further the computational aspects and difficulties of how to calculate P in those approaches. In some cases, P will converge to a stationary probability $P(\sigma, t \rightarrow \infty) = P_{\text{st}}(\sigma)$. Thus, the right hand side of Equation (3.1) is zero

and we have the following equilibrium condition

$$W_{\sigma\to\sigma'}P_{\text{st}}(\sigma') = W_{\sigma'\to\sigma}P_{\text{st}}(\sigma), \tag{3.3}$$

which is called **detailed balance**, a sufficient condition for equilibrium in those systems.

3.1.1 One-step and Poisson processes

An important class of Markov processes is the birth-and-death or **one step processes**, which are defined as a continuous time process whose state is described by integers n and transitions that occurs only on adjacent states [100]. One-step processes occurs whenever the stochastic system presents adsorption or emission of particles, birth and death of individuals, or simply the counting of events. Due to this definition, we may simplify the notation on probability distributions: $P(\sigma, t) \to P_n$. Then, for these processes, the master equation becomes

$$\dot{P}_n = \mu_{n+1}P_{n+1} + \lambda_{n-1}P_{n-1} - (\mu_n + \lambda_n)P_n \tag{3.4}$$

in which the coefficients μ_n and λ_n are transition rates that, being at n, the system jumps to $(n-1)$ or $(n+1)$, respectively. Figure 3.1 illustrates this transitions.

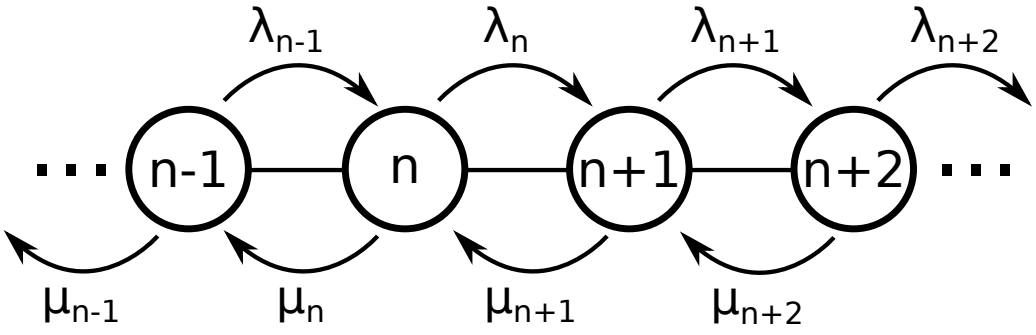


Figure 3.1: Transitions rates for one-step processes.

Additional descriptions of the system may be necessary if n is lower and/or upper bounded. In those cases, we call these limitations on n **boundaries** and we may need to introduce new equations to describe them. An important type is the **absorbing boundary**, that occurs when a certain state of the process traps or absorbs the dynamic, totally preventing the access to other states. In an epidemic spreading, for example, if we assume that no spontaneous infection occurs, a state with no infectious

individuals is an **absorbing state (AS)**, since the dynamics can no longer evolve. This kind of boundary also occurs on other problems such as gas dissociation or radioactive decay [100]. The existence of AS can be troublesome in some analysis and part of this thesis is dedicated to discuss and develop methods to circumvent the difficulties associated with absorbing states. They are discussed in Chapter 4.

Let us consider an one-step process with the following transition rates and initial condition:

$$\mu_n = 0 \qquad \lambda_n = q \qquad P_n(0) = \delta_{n,0}, \tag{3.5}$$

in which q is a constant parameter. Then, the master equation for this process become

$$\dot{P}_n = q(P_{n-1} - P_n), \tag{3.6}$$

which is a random walk over the integers with steps only to the right with probability q, i.e., the walker just moves away from the starting point at random times. The solution for this equation can be found using the generating function of the distributions P_n [101] and is

$$P_n(t) = \frac{(qt)^n}{n!}\,\mathrm{e}^{-qt}, \tag{3.7}$$

which is the probability distribution for the **Poisson process**, an important set of processes for modeling random events. In fact, Equation (3.7) may be interpreted as the probability of occurring n random events (each one with probability q) between some initial time $t_0 = 0$ and t.

Starting from this general proposition for the Poisson processes, let us study a simpler particular scenario: an event occurs in a time interval dt with rate q and probability qdt. We have that $p_0(t)$ is the probability of the event does not occur until time t and $p_1(t) = 1 - p_0(t)$ is the probability of occurrence. Thus, we can construct the master equations for this scenario

$$\dot{p}_0 = -qp_0,$$
$$\dot{p}_1 = qp_0.$$

By setting $p_0(t_0) = 1$ and $p_1(t_0) = 0$, the solution of these equations are $p_0(t) = \exp(-qt + t_0)$ and $p_1(t) = 1 - \exp(-qt + t_0)$. Without loss of generality, we can set

$t_0 = 0$ and write the particular solutions in a more convenient way

$$p_0(t) = \mathrm{e}^{-qt} \qquad \text{and} \qquad p_1(t) = 1 - \mathrm{e}^{-qt}. \tag{3.8}$$

Now, let us consider a set of Ω independent Poisson events, each one occurring with rate q_i $(i = 1, 2, 3, ..., \Omega)$. The probability that event i will occur in the time interval $[\tau, \tau + dt]$ is

$$G(\tau, i)dt = \left(\prod_{i=1}^{\Omega} \mathrm{e}^{-q_i \tau} \right) q_i dt, \tag{3.9}$$

in which the product is the probability that no event has occurred until τ and $q_i dt$ is the probability that i occurs in the given time interval. The product can be written in a more convenient way by defining $Q = \sum_i q_i$,

$$G(\tau, i)dt = \left(\frac{q_i}{Q} \right) (Q\mathrm{e}^{-Q\tau})dt. \tag{3.10}$$

Equation (3.10) can be interpreted as follows: the next event will be i with probability (q_i/Q) and it will occur after a time τ with distribution $P(\tau) = Q\mathrm{e}^{-Q\tau}$. Therefore, the Ω independent Poisson events might be described by a single exponential distribution with rate Q. This approach is the foundation of **Gillespie algorithm** [102], which is further described in Section 3.2 in the context of epidemic processes.

3.2 Mathematical epidemiology

Theoretical epidemiology is a highly interdisciplinar field with the involvement of areas such as biology, mathematics, physics and medicine. Through mathematical and computational frameworks, it was possible to study the spreading and dynamics of infectious diseases that have led to our current understanding of many pathogens. Furthermore, recent studies with the participation of physicists regarding $H1N1$ Influenza [18], Ebola [19] and COVID-19 [29] discuss mitigation strategies, designs scenarios and analyzes the effectiveness of public policies in order to prevent the spreading of these pathogens. Here in Brazil, the field has important studies on the spread of Zika [20], Dengue [21, 103] and, more recently, COVID-19 [104].

In addition to these immediately practical applications, spreading phenomena in general have strong connections with physical non-equilibrium phase transitions,

from which we can study traditional phenomena such as criticality and universality on these systems [31, 32].

3.2.1 Concepts and compartmental models

One approach to disease's mathematical modeling is the compartmental models, i.e., discretize the host-pathogen interaction dynamics into compartments, attributing to them the characteristics of different disease stages [60]. In Figure 3.2, we have a scheme of the compartments and their correspondence with immune responses and the pathogen concentration on a typical infectious disease. In this example, we divided the disease's dynamics into four compartments: susceptible, exposed, infectious and removed. The next step is to define characteristics for each of these stages to reproduce features of the biological dynamics. Among several characteristics that may be incorporated into a model, the main ones are: the transmission routes (by host, sexual partners, airborne, etc.) and the transitions rates, which are related to the disease time scale and probabilities of infection and death.

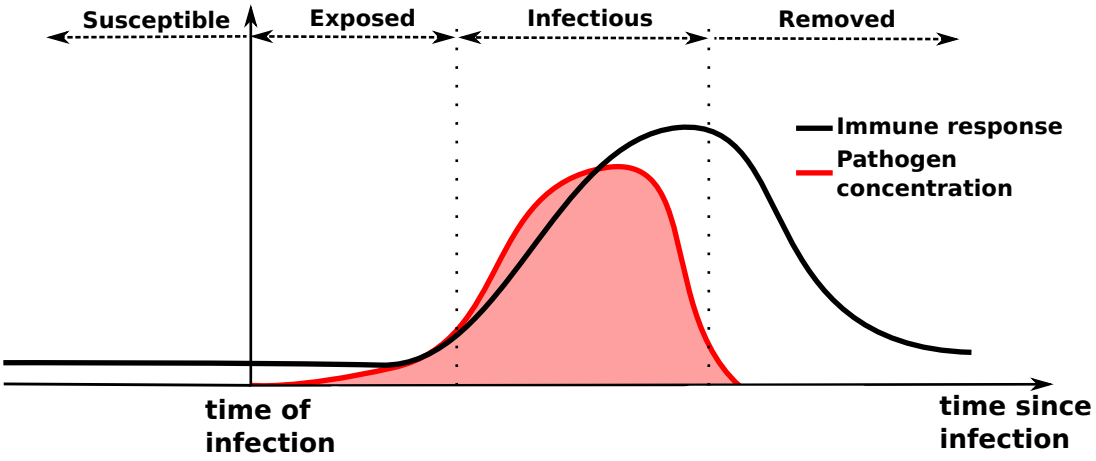


Figure 3.2: Scheme of a timeline of an infection, showing the pathogen concentration (shaded curve) and the immune response. The arrows above illustrate possible compartments to mimic this dynamics.

To illustrate some technical concepts, let us consider a simple and widely used model in literature: the **susceptible-infected-removed (SIR)** [105]. In this model, the disease dynamics is divided into three compartments:

- Susceptible: healthy individual who has not yet come into contact with the disease. Upon contact with an infectious, it becomes infected with rate λ per

contact.

- Infected: individual that can transmit the pathogen. Upon contact with a susceptible individual, it infects them with a rate λ per contact and becomes spontaneously removed with rate μ.
- Removed: individual who does not participate in the disease dynamics, either by being cured or having died. In both cases, the individual does not transmit and cannot get infected again.

With the model defined, there are several ways to analyze it. You can perform computer simulations or build differential equations in an attempt to extract information and infer behaviors. In both cases, one needs to define the substrate, i.e., the environment in which the spreading process occur. If the model represents a computer virus, does it spread on the physical network of routers? If so, how to model this network? It could be an ER graph, a scale-free network or any given model, but the possibilities should be suited to the specific type of problem.

For infectious diseases, a common first approach is to consider the so-called **homogeneous mixing hypothesis**, in which any pair of individuals of a certain population N can interact with equal chance and all individuals have the same number of daily contacts k. Although it seems like an unrealistic assumption, some important results can be extracted from the differential equations for the SIR under this hypothesis, that are given by

$$\frac{dI}{dt} = \lambda k s(t) I(t) - \mu I(t), \tag{3.11}$$

$$\frac{dR}{dt} = \mu I(t), \tag{3.12}$$

and the closure relation $S + I + R = N$. The terms $\lambda k s(t) I(t)$ and $\mu I(t)$ are the total infection and healing rates, respectively, and $s(t)$ is the fraction of the susceptible population ($S(t)/N$). An integration of these equations can be seen in Figure 3.3-(a) for $\lambda k = 0.2$, $\mu = 0.1$, $N = 1000$ and initial condition $I(0) = 1, R(0) = 0$ and $S(0) = 999$, showing a typical behavior of epidemic curves.

The general behavior of the spreading can be easily predicted knowing the epidemic parameters and analyzing the right hand side of Equation (3.11): since $I(t)$ is strictly positive, we can write the following conditions:

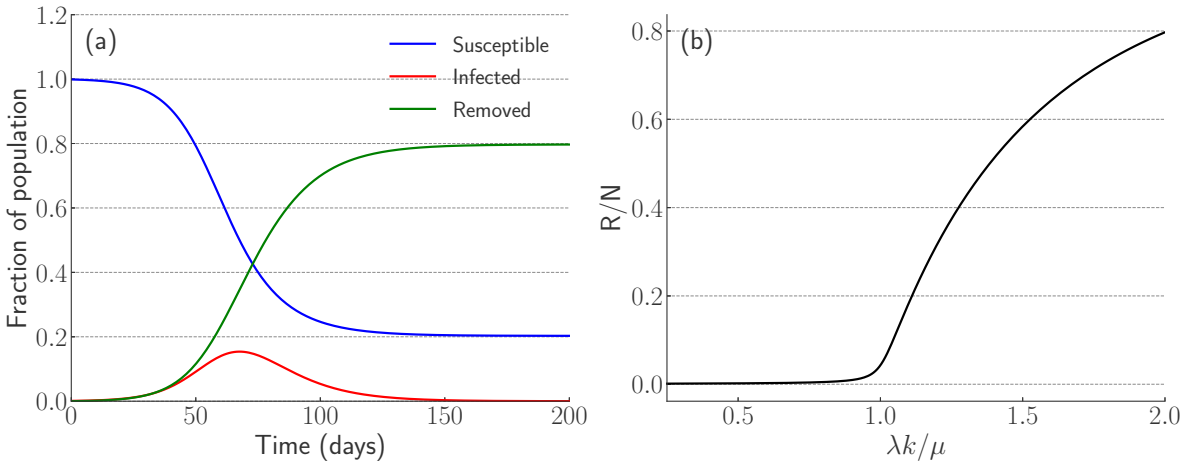


Figure 3.3: (a) Integration of SIR equations with $\lambda k = 0.2$, $\mu = 0.1$ and $N = 1000$. (b) Recovered individuals R at the end of a simulation as a function of the ratio $\lambda k/\mu$.

- $\frac{\lambda k s(t)}{\mu} > 1 \Longrightarrow \frac{dI}{dt} > 0 \Longrightarrow$ the outbreak is increasing

- $\frac{\lambda k s(t)}{\mu} < 1 \Longrightarrow \frac{dI}{dt} < 0 \Longrightarrow$ the outbreak is decreasing

Figure 3.3-(b) illustrates the different outcomes of the spreading process by plotting the fraction of recovered individuals at the end of epidemic in function of $\lambda k/\mu$ for the same initial condition as in Figure 3.3-(a). It can be seen that for $\lambda k/\mu < 1$, the disease hardly spreads while for $\lambda k/\mu > 1$ an appreciable fraction of the population is affected.

This analysis can be settled within an wider concept in epidemiology, the **basic reproduction number** R_0. In a simple way, R_0 is the average number of individuals infected by a single infectious person introduced in a population composed only of susceptible individuals [60]. If we consider an arbitrary homogeneous mixed population, we can write the **effective reproductive number** as $R_t = R_0 \times S/N$. Both concepts are important because they condense information about the epidemic into just one dimensionless number. If R_t is greater than one, the infection is spreading while if R_t is smaller than one, the outbreak is decreasing. In Table 3.1 we present estimated R_0 values for some infectious diseases.

For more complex models with mobility, multiple infectious compartments or age stratification, the analysis done for SIR and consequently the calculation of R_0 is not straightforward. In general, it can be calculated from the summation over all infec-

Infectious Disease	Host	Estimated R_0
Feline Immunodeficiency Virus	Domestic Cats	$1.1 - 1.5$
Rabies	Dogs (Kenya)	2.44
Phocine Distemper	Seals	$2 - 3$
Tuberculosis	Cattle	2.6
Influenza	Humans	$3 - 4$
Foot-and-Mouth Disease	Livestock farms (UK)	$3.5 - 4.5$
Smallpox	Humans	$3.5 - 6$
Rubella	Humans (UK)	$6 - 7$
Chickenpox	Humans (UK)	$10 - 12$
Measles	Humans (UK)	$16 - 18$
Whooping Cough	Humans (UK)	$16 - 18$
COVID-19 [25]	Humans	$2.3 - 4.1$

Table 3.1: Some estimated basic reproductive ratios. Adapted from [60].

tious compartments $\mathcal{C}$ of the infection rate $\lambda_{\mathcal{C}}$ multiplied by the average time that an individual remains on $\mathcal{C}$ (expressed as the inverse of healing rate $\mu_{\mathcal{C}}$), multiplied by the probability $P_{S\to\mathcal{C}}$ that $\mathcal{C}$ belongs to the epidemic chain, in which $S \to \mathcal{C}$ represents the transition starting in a susceptible individual S and reaching the infectious compartment $\mathcal{C}$. The equation bellow condenses this calculation

$$R_0 = \sum_{\mathcal{C}} P_{S\to\mathcal{C}} \frac{\lambda_{\mathcal{C}}}{\mu_{\mathcal{C}}}. \tag{3.13}$$

For the SIR model, we have just one infectious compartment with infection rate λk and healing rate μ. Since we have just one possible epidemic chain ($S \to I \to R$), the probability $P_{S\to I}$ is 1. Therefore, $R_0 = \lambda k/\mu$. To illustrate a model with multiple epidemic chains, let us consider a slightly modified SIR model, with two types of infectious compartment: I_1, with mild symptoms and contact infection rate λ_1 and I_2, with severe symptoms and infection rate λ_2. Individuals in I_1 can evolve spontaneously to I_2 or R with rates α_I and α_R, respectively. Individuals in I_2 spontaneously become R with rate β. This model is illustrated in Figure 3.4.

This model contains two epidemic chains ($S \to I_1 \to I_2 \to R$ and $S \to I_1 \to R$) and two infectious compartments $\mathcal{C} = \{I_1, I_2\}$. Both contribute to R_0 but possesses different transition rates and $P_{S\to\mathcal{C}}$. The compartment I_1 is present in both epidemic chains, therefore, $P_{S\to I_1} = 1$. However, the compartment I_2 appears only on one chain. The probability $P_{S\to I_2}$ is directly proportional to the transition rate between compartments I_1 and I_2 and inversely proportional to the total transition rate of I_1.

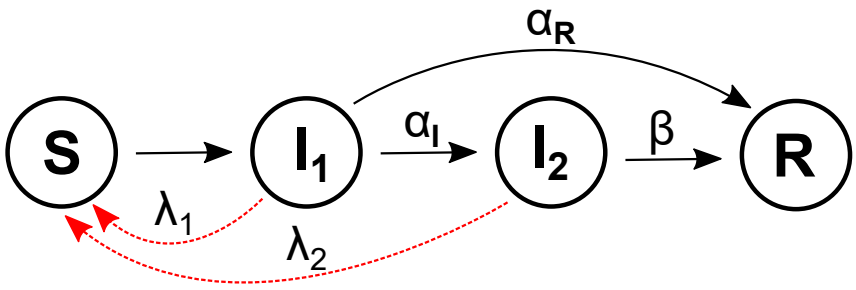


Figure 3.4: Schematic illustration of compartments and transitions of the SIR model with two infectious states.

Therefore, $P_{S\to I_2} = \dfrac{\alpha_I}{\alpha_I + \alpha_R}$. The complete expression for R_0 become

$$R_0 = \frac{\lambda_1 k}{\alpha_I + \alpha_R} + \left(\frac{\alpha_I}{\alpha_I + \alpha_R}\right)\frac{\lambda_2 k}{\beta}. \tag{3.14}$$

3.3 Epidemics as phase transitions

One of the main issues of equilibrium statistical mechanics is the study of phase transitions, phenomenon characterized by changes in the macroscopic behavior of the system when an external parameter, such as the temperature, is varied [17]. For example, liquids can be transformed into gas or solid upon changes in temperature or pressure, ferromagnets can change their magnetization upon changes in a external magnetic field or temperature, among others. These two examples are part of a group called **equilibrium phase transitions (EPT)**, in which a general statistical description can be made in terms of the micro-states probabilities, given by the Boltzmann weight [33]. The same cannot be said of the so-called **non-equilibrium phase transitions (NEPT)**, characterized by breaking the detailed balance, which implies in a non-zero flow of probabilities between pairs of micro-states [32].

Although we are dealing purely with non-equilibrium systems in this thesis, many equivalences can be made with EPTs. So, we discuss the most paradigmatic EPT theoretical system: the **Ising model**, in which a collection of spins are arranged on a crystalline lattice, usually a regular one (squared, cubic, hexagonal, etc.) and each spin can interact with nearest neighbors. Each spin can assume the values $\sigma_i = -1$ or 1 and, at external magnetic field $\mathbf{B} = 0$, the system's Hamiltonian is defined as

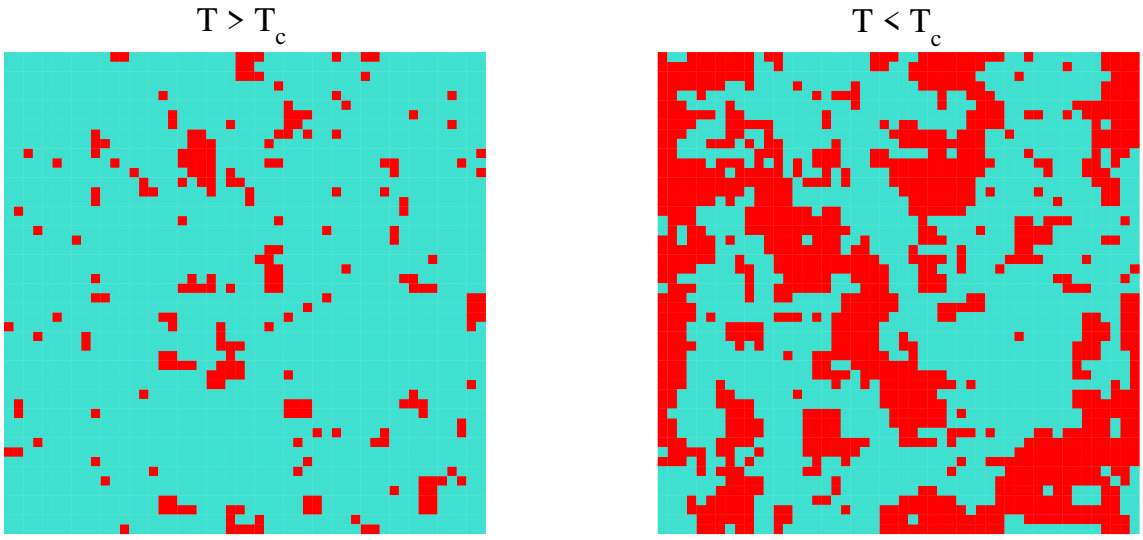


Figure 3.5: Typical snapshots of a two-dimensional Ising system below and above the critical temperature. Red represents $\sigma_i = -1$ and cyan $\sigma_i = 1$.

$$\mathcal{H} = -\sum_{\{i,j\}} J_{ij}\sigma_i\sigma_j, \tag{3.15}$$

in which $\{i, j\}$ denotes all pairs of neighbors and J_{ij} is a coupling parameter that represents the interaction between spins i and j. Experimentally, it is observed that at low temperatures, the spin ordering of ferromagnetic materials ($J_{ij} > 0$) generates an appreciable residual magnetization M. As the temperature increases above a certain threshold T_c, the thermal fluctuations destroys this magnetic ordering and M drops to zero. This transition between an ordered (ferromagnetic) to a disordered (paramagnetic) behavior characterizes the phase transition. In Figure 3.5, we show simulations for typical configuration of Ising model bellow and above the critical temperature T_c, illustrating this order-disorder transition.

Figure 3.6 shows the average magnetization $\langle M \rangle = \sum_i \sigma_i / N$ as we vary the temperature T for the Ising model on a 50×50 square lattice. It can be seen a clear distinction between the ferromagnetic and paramagnetic phases as $\langle M \rangle$ decreases to zero. In this case, T is called **control parameter** and $\langle M \rangle$ is an **order parameter**, as it quantifies the ordering degree of the system. It is important to notice that the smooth behavior near the transition is due to the system's finite size. On thermodynamic limit, one expects the magnetization to abruptly but continuously decreases to zero. In addition, note the similarities between this curve and Figure 3.3-(b).

When studying and characterizing phase transitions, an important concept are the

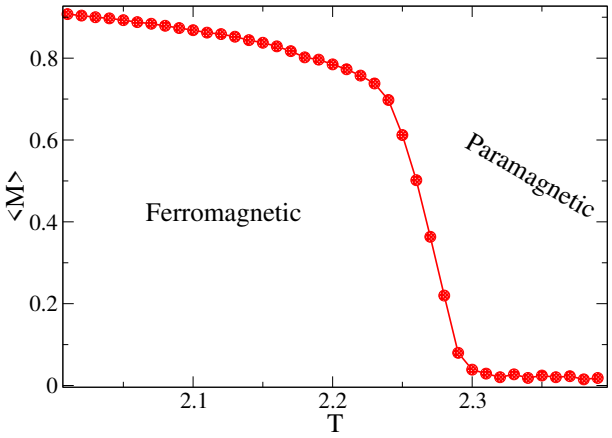


Figure 3.6: Average magnetization $\langle M \rangle$ as a function of temperature T for the simulations of Ising model on a 50×50 square lattice.

correlations functions C, that captures how two system variables influence each other (spatially or temporally). In the Ising model, for example, we can define the spatial correlation function using the spins

$$C(r) = \langle \sigma_i \sigma_j \rangle_r - \langle \sigma_i \rangle \langle \sigma_j \rangle, \tag{3.16}$$

in which $\langle ... \rangle_r$ is averaged over all spin pairs i and j that are at a distance r from each other. For NEPTs, we can also define a temporal auto-correlation function $\mathcal{T}(\tau)$, that correlates the state of an element with itself on different times. Generally speaking, these correlation functions have characteristic lengths ξ, namely **correlation lenghts**, that indicates the range of this influence. A major feature of ξ is the divergence at the transition point following a scaling law

$$\xi \sim |T - T_c|^{-\nu}, \tag{3.17}$$

where T is the control parameter, T_c is the transition point and ν is a critical exponent. When the transition point has diverging correlations length it is called a **critical point** and involves a critical transition. Many other quantities have this power law behavior near the transition with different exponents. These scaling laws are benchmarks of universality classes and will be explored further in this thesis.

From now on, we turn our attentions to non-equilibrium phase transitions, in particular those of epidemic models. As aforementioned, Figures 3.3-(b) and 3.6 shows the similarities between these two phenomena if we consider $\lambda k/\mu$ as control and

R/N as an order parameter. However, to introduce the subject and discuss some techniques and theories, we will let the SIR aside for a moment and focus on the **Susceptible-infected-susceptible** (SIS) model, which forms the basis of the models studied in Chapters 4 and 5.

The SIS model [37] is one of the most fundamental models to study epidemics or information spreading, besides the aforementioned SIR. A fundamental aspect of the SIS model is the instantaneous waning of immunity, i.e., it is possible for the same individual to be reinfected by the pathogen immediately after healing. Although this is not true for most pathogens, certain viruses such as Influenza produce new variants that evade the immune system quickly [106] and therefore, their dynamics may be approximated by a SIS-like model. Figure 3.7 illustrates the transitions in the SIS model.

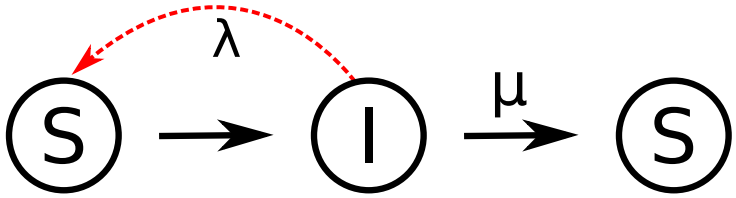


Figure 3.7: Schematic illustration of compartments and transitions of the SIS model.

Just as we did for the SIR model, let us start our analysis with the differential equations under the homogeneous mixing hypothesis. For practicality, we will write the equations in terms of densities: $\rho = I/N$ and $s = S/N$. The density of infected nodes ρ is sometimes called **epidemic prevalence**. In addition, the closure relation become $\rho + s = 1$ and the dynamical equation for SIS can be written as

$$\frac{d\rho}{dt} = (\lambda k - \mu)\rho - \lambda k \rho^2, \tag{3.18}$$

which has the following fixed points ($\dot{\rho} \approx 0$): $\rho_1 = 0$ and $\rho_2 = \dfrac{\lambda k - \mu}{\lambda k} = \alpha$. By performing a stability analysis, we notice that the system undergoes a transcritical bifurcation [107] as λ varies, i.e., the fixed points switch their stability. The transition point can be calculated when ρ_1 looses stability, which gives $\lambda_c = \mu/k$. Figure 3.8 illustrates this transition in terms of ρ by varying λ. We can see two well defined behaviors, an **inactive phase**, in which the epidemic cannot sustain itself ($\lambda < \lambda_c$) and an **active phase**, in which the epidemic spreads endlessly ($\lambda > \lambda_c$). The exact point

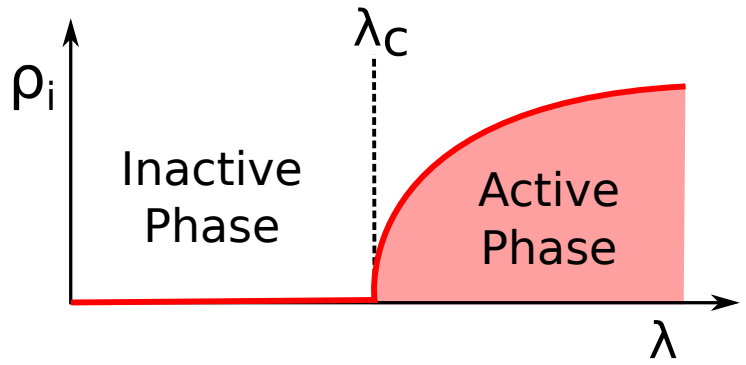


Figure 3.8: Scheme of the SIS model phase transition, highlighting the existence of a critical point λ_c that separates the two phases.

where this transition occurs ($\lambda = \lambda_c$) is a critical point or **epidemic threshold** of the model. This example represents an **absorbing state phase transition** (ASPT), a class of NEPT in which absorbing states are involved directly in the transition.

In principle, these considerations for the SIS model are valid under the simplified homogeneous mixture hypothesis, which may not be realistic in some scenarios. For the study of other more complicated substrates, such as contact networks, we need to rely on other types of frameworks, such as mean field theories and computer simulations. Some of these methods, applications and discussions will be reported in the sequence of this section. Except where mentioned, in all cases we are considering the SIS model evolving in a complex network with degree distribution $P(k)$. In these cases, the networks' vertices are individuals who can become infected and the links represent contacts amongst them.

3.3.1 Mean-field theories

The main assumption of a mean-field theory is to consider that the local field is replaced by the averaged one over the whole system. For example, in the Ising model the interactions among nearest-neighbors are replaced by a field equivalent to the average magnetization of the whole sample. In the context of network theory, it consists of partially neglecting the actual structure of the network (by assuming average interactions) or the dynamical correlations between the vertices (influence of a node state on its neighbors).

The simplest mean-field theory is an extension of the homogeneous mixture hypothesis to heterogeneous systems, hence called **Homogeneous Mean Field Theory**,

in which one assumes that all vertices are statistically equivalent and interact with $\langle k \rangle$ neighbors. Thus, the dynamical equation for this theory is dramatically similar to Equation (3.18),

$$\frac{d\rho}{dt} = -\mu\rho + \lambda\langle k \rangle \rho(1-\rho), \tag{3.19}$$

in which we replaced $k \to \langle k \rangle$ and assumed $\mu = 1$ without loss of generality since this choice only fixes the time scale. Thus, the epidemic threshold for this case is $\lambda_c = 1/\langle k \rangle$.

This approach can be improved if, instead of considering that all vertices are statistically identical, we consider that certain groups of vertices (divided according to some feature) are equal. One of the first theories proposed following this direction was the **degree-based mean field** or **heterogeneous mean field** (HMF) theory [108], in which one assumes that all vertices of degree k are statistically equivalent, irrespective of its actual neighborhood. This assumption also means that k-degree vertices are connected with probability $P(k'|k)$ to vertices with degree k', which is a improvement on the homogeneous theory regarding the network structure. Thus, one can construct the dynamical equations for SIS under this hypothesis to obtain

$$\frac{d\rho_k}{dt} = -\rho_k + \lambda k(1-\rho_k) \sum_{k'} P(k'|k)\rho_{k'}, \tag{3.20}$$

in which the first term of the r.h.s. of the equation represents spontaneous healing of vertices while the second one is the infection of k-degree susceptible vertices by all k'-degree infectious neighbors. Looking at the stationary solutions of this equation, we note that $\rho_k = 0, \forall\ k$, is a fixed point. By performing a linear stability analysis around this point, we have

$$\frac{d\rho_k}{dt} = \sum_{k'} L_{kk'}\rho_{k'}, \tag{3.21}$$

in which $L_{kk'}$ is the system's Jacobian matrix

$$L_{kk'} = -\delta_{kk'} + \lambda k P(k'|k). \tag{3.22}$$

From linear equations theory, the fixed points of the system are stable if the largest eigenvalue of $L_{kk'}$ is negative and unstable otherwise. Thus, the epidemic threshold

can be written as

$$\lambda_c = \frac{1}{\Lambda_{\max}}, \tag{3.23}$$

in which $\Lambda_{\max}$ is the largest eigenvalue of the correlation matrix $C_{kk'} = kP(k'|k)$. In order to calculate the epidemic threshold explicitly, we need to assume a form for $P(k'|k)$. One of the simplest and most useful cases are the uncorrelated networks, presented in Chapter 2. In this case, $P(k'|k)$ is given by Equation (2.12) and the correlation matrix assume the form

$$C_{kk'} = \frac{kk'P(k)}{\langle k \rangle}. \tag{3.24}$$

It is direct to show that $C_{kk'}$ has $v_k = k$ as one of its eigenvectors with $\Lambda = \langle k^2 \rangle / \langle k \rangle$ being the corresponding eigenvalue. In order to show that Λ is the largest eigenvalue of $C_{kk'}$, we will make use of the **Perron-Frobenius theorem**, whereby a non-negative irreducible square matrix always has a positive real eigenvalue $\Lambda_{\max}$ and the magnitude of no other eigenvalue exceeds $\Lambda_{\max}$. Furthermore, the theorem also attests that the eigenvector corresponding to $\Lambda_{\max}$ has all positive inputs and it is not degenerate. Thereby, $C_{kk'}$ satisfies the conditions for using the theorem and we can guarantee that $\Lambda = \langle k^2 \rangle / \langle k \rangle$ is the largest eigenvalue of this matrix. Therefore, the epidemic threshold for SIS on HMF theory is

$$\lambda_c^{\mathrm{HMF}} = \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{3.25}$$

In Section 2.3, we calculated $\langle k \rangle$ and $\langle k^2 \rangle$, concluding that for $2 < \gamma < 3$, $\langle k^2 \rangle$ diverges in the thermodynamic limit while $\langle k \rangle$ remains finite. Therefore, we have that $\lambda_c^{\mathrm{HMF}} \to 0$ as $N \to \infty$ in this γ range. For $\gamma > 3$, both $\langle k \rangle$ and $\langle k^2 \rangle$ are finite, leading to a non-zero threshold as $N \to \infty$.

Other common mean-field approach on theoretical epidemiology is the use of **individual-based mean-field (IBMF) theories** , in which we aim to write down the probability ρ_i^{X} that node i belongs to compartment X [109]. Dynamical correlations between nodes may or not be taken into account by different approaches [110]. The simplest IBMF theory, called **quenched mean field (QMF)** considers that neighboring nodes are statistically independent, i.e., the state X of a node i does not influence the state Y of its neighbor j. Because it is a theory at the individual level, we keep infor-

mation of the full structure of the network by imputing the adjacency matrix A_{ij} in the theory to obtain

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda(1-\rho_i)\sum_j A_{ij}\rho_j. \tag{3.26}$$

Similarly to Equation (3.20), we perform a linear stability analysis around the stationary solution $\rho_i = 0$,

$$\frac{d\rho_i}{dt} = \sum_j J_{ij}\rho_j, \tag{3.27}$$

in which the Jacobian matrix J_{ij} is

$$J_{ij} = -\delta_{ij} + \lambda A_{ij}. \tag{3.28}$$

The epidemic threshold λ_c will be given by the inverse of largest eigenvalue of A, denoted $\Lambda_{\max}$. These eigenvalues were calculated in [111] for random uncorrelated networks with $P(k) \sim k^{-\gamma}$, leading to the following estimates for λ_c [112]:

$$\lambda_c^{QMF} = \frac{1}{\Lambda_{\max}} \simeq \begin{cases} \langle k\rangle / \langle k^2\rangle & \text{if } 2 < \gamma < 5/2 \\ 1/\sqrt{k_c} & \text{if } 5/2 < \gamma \end{cases}$$

The first important result for QMF is that for $2 < \gamma < 5/2$, the calculated threshold is identical to HMF while for $\gamma > 5/2$ the two theories predict different values of λ_c. In Section 2.3, we also saw that $k_c \sim N^{\frac{1}{\gamma-1}}$ for uncorrelated networks. Thus,

$$\lambda_c^{QMF} \sim N^{-\frac{1}{2(\gamma-1)}}, \tag{3.29}$$

leading to a vanishing threshold for $\gamma > 5/2$ as $N \to \infty$. So, for $\gamma > 3$, there is a remarkable difference between HMF and QMF theories. While the former predicts a non-zero threshold, the latter predicts a vanishing one, as the system approaches the thermodynamic limit. Even for $5/2 < \gamma < 3$, the theories differ regarding the dependency of λ_c and N, despite both theories predicting vanishing thresholds.

Figure 3.9 show a comparison between epidemic thresholds for the SIS model on UCM networks with $\gamma = 2.3$ (a), 2.8 (b) and different sizes, calculated by homogeneous mean-field, QMF and HMF theories. Stochastic simulations (to be discussed in

Section 3.3.2) are also presented. It can be noted that homogeneous theory performs worst in predicting the threshold, since we are studying highly heterogeneous systems. QMF and HMF performs really well for $\gamma = 2.3$, following the scaling law and converging to a common value as N increases. For $\gamma = 2.8$, there are differences in the predictions, with QMF closer to the simulations than HMF.

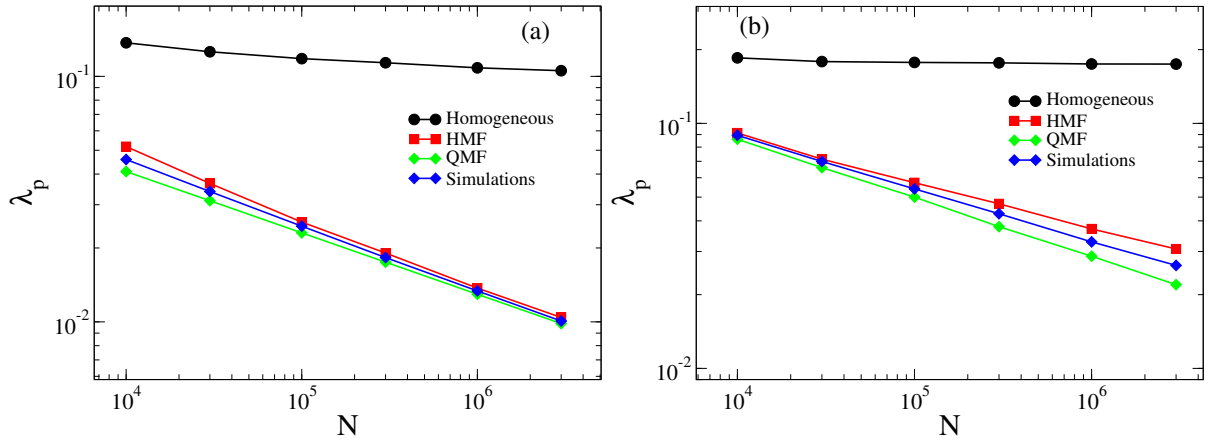


Figure 3.9: Epidemic thresholds calculated by different mean field theories in comparison with simulations on UCM networks. (a) $\gamma = 2.3$; (b) $\gamma = 2.8$.

3.3.2 Computational simulations and scaling theory

The use of computational simulations to study and analyze complex systems has become increasingly recurrent as advances in computer processing and data collection makes this process easier and more necessary. The major simulation framework used in this thesis will be the **agent-based modeling (ABM)** [113], in which each element of the system is an autonomous agent or individual that interacts with its peers based on certain rules and contact patterns. In particular for epidemic modeling, we can think each agent as a possible host and the interactions as transitions in a compartmental model. In this section, we will discuss the fundamental algorithms for our simulations as well as some general techniques. Specific frameworks will be discussed in their respective chapters.

The ABM approach to epidemic modeling commonly considers the transitions between compartments as Poisson processes. Therefore, we can use the results from Section 3.1.1 to construct a general algorithm to simulate them. This is the well know **Gillespie Algorithm** (GA) [102], constructed with the following steps.

i). Construct a list with all possible events and their respective rates $q_1, q_2, \ldots$.

ii). From the exponential distribution $P(\tau) = Qe^{-Q\tau}$, we select the time step size τ as $\tau = -\ln(u)/Q$, in which Q is the sum of all rates and u is a random number with uniform distribution in the interval $(0, 1]$.

iii). Select the event i to implement with probability q_i/Q.

iv). Update the state of the system based on the sorted event.

v). Increment time as $t \to t + \tau$.

vi). Return to step i.

Despite being general, the GA can be extremely inefficient when handling large systems due to the construction and updating of event lists being computationally costly. In Reference [114], the authors developed algorithms that improved simulations performance at least one order of magnitude. The main algorithm which will be used in this thesis is called **optimized Gillespie algorithm** (OGA), the formal details can be found in Reference [114]. This algorithm is based on a simple but powerful concept: the **phantom processes**.

Briefly, phantom processes are events that do not change the system's state but do contribute for time counting. In epidemiology, these processes usually occurs when an infectious individual spreads the pathogen to others that cannot catch the disease (either because already infected or immunized). In the original GA, these events would not be on the list to be selected. By allowing these events (with certain rates and time steps), we greatly simplify the event list construction and therefore, speed up the simulations. To exemplify, we will describe the OGA to perform simulations of the SIS model on a complex network

i). Starting from a initial configuration, we calculate the number of infected vertices N_i and the sum of their degree N_e. In addition, we create a list with the labels of all infected nodes.

ii). With probability $\dfrac{N_i}{N_i + \lambda N_e}$, an infected vertex, randomly chosen, become susceptible. With complementary probability $\dfrac{\lambda N_e}{N_i + \lambda N_e}$, an infected node is chosen with probability proportional to its degree. Then, we randomly select one of his neighbors with equal chance. If this neighbor is susceptible, we infect it. Otherwise, we proceed to the next step.

iii). The time t is incremented as $t \to t + \tau$, in which τ is drawn from the exponential distribution $P^*(\tau)$, $\tau = \ln(u)/Q^*$, $Q^* = N_i + \lambda N_e$. This new distribution contains the rates for phantom processes.

iv). N_i, N_e and the list of infected vertices are updated.

v). Return to step 2.

We still have to determine which analysis will be carried out and which are the relevant quantities to investigate. We are usually interested in identifying the epidemic threshold and characterizing the system near the critical point. This characterization usually occurs by using some well know scaling laws (as mentioned on the Ising model), which relates the relevant quantities with the system size or distance from the transition point. Some scale laws and critical exponents can be calculated theoretically for some systems and, therefore, can serve as reference for comparison with those found from simulations. Although our focus in this thesis is on computational simulations, we present a simple example to illustrate these theoretical framework.

In Equation (3.18), we saw that the stationary density of infected nodes above the transition ($\lambda > \lambda_c$) is $\rho_s = \dfrac{\lambda k - \mu}{\lambda k}$ and the epidemic threshold is $\lambda_c = \mu/k$. Then,

$$\rho_s = \frac{\lambda k - \mu}{\lambda k} = \frac{1}{k}\frac{\lambda - \mu/k}{\lambda} \sim (\lambda - \lambda_c)^{\beta}. \tag{3.30}$$

Here, $\beta = 1$ is a critical exponent that describe the behavior of ρ_s as λ varies near to the transition point. For the reader interested in more in-depth theoretical frameworks regarding scaling laws and critical exponents, see references [31] and [32].

Returning to computational simulations of SIS and SIR-like models, usually our first interest is to identify the transition point λ_c. The most common techniques are the decaying and spreading simulations.

Decaying simulations

We start with a fully infected population and observe the evolution of ρ (density of infected individuals) as time evolves. Depending on λ, three scenarios can occur:

- Subcritical ($\lambda < \lambda_c$): ρ decay exponentially to zero
- Critical ($\lambda = \lambda_c$): ρ decay as a power law ($\rho \sim t^{-\alpha}$)

- Supercritical ($\lambda > \lambda_c$): ρ converges to a non-zero value after a relaxation time.

Spreading simulations

We start with just one infectious individual and observe the survival probability $P_s(t)$, i.e., the probability that a sample does not fall into the absorbing state until a given time t. The behavior of P_s is qualitatively similar to ρ on decaying simulations, with a different scaling law in ($P_s \sim t^{-\delta}$) in the critical region.

Thus, by inspecting $\rho(t)$ or $P_s(t)$ we can estimate the epidemic threshold. In addition, α and δ are critical exponents that characterizes the transition. In Figure 3.10-(a) and (b), we present decaying and spreading simulations of SIS (10^3 and 10^6 samples, respectively) on a RRN with $m = 4$, $N = 10^6$ and different regimes of λ. It can be observed that both simulations present a similar behavior: exponential decaying for subcritical simulations, convergence to a finite value for supercritical ones and a power-law decay that separates these two phases. It worths notice that both methods yield an epidemic threshold $\lambda_c = 0.345$. Other quantities can be defined and calculated by performing these simulations and can be adapted to the models specificity. Taking a spreading simulation for SIR as example, the clusters size of removed individuals at the end of the outbreak is used to identify the epidemic threshold.

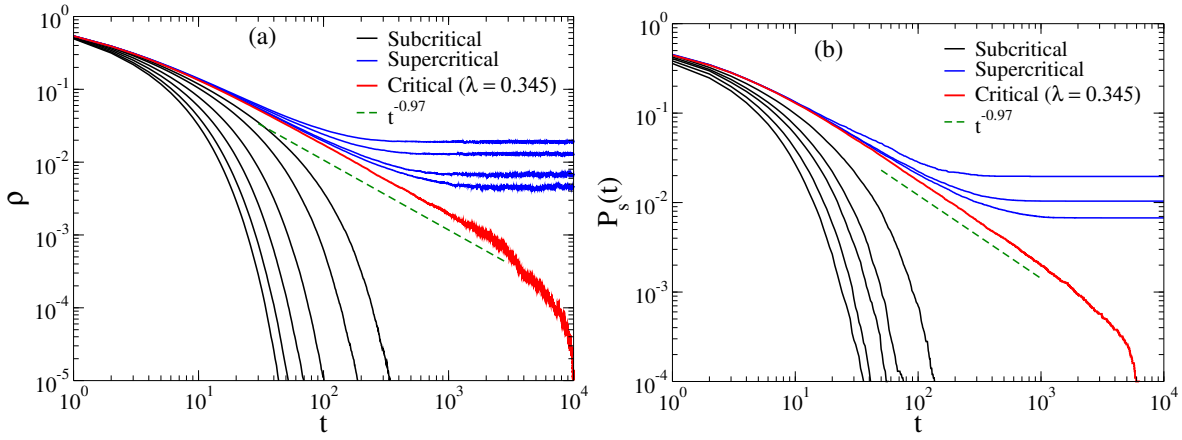


Figure 3.10: (a) Decaying and (b) spreading simulations of SIS model on random regular networks with $m = 4$, $N = 10^6$ and different regimes of λ. The dashed lines are power-laws with exponents $\phi = -0.97$.

Finite size scaling

Many hypothesis and theories on phase transitions are conceived for infinite systems and time. The proper definition of a critical point is only valid in the thermodynamic limit, in which the correlation lengths diverge. To circumvent this, there is a technique to extrapolate properties of the system in the thermodynamic limit by comparing simulations for different system sizes. This technique is called **finite size scaling (FSS)**, initially thought for systems on lattices and later extended to networks, will be briefly discussed here. Near to the criticality, it is observed that the correlation lengths ξ diverges scaling as

$$\xi_{\perp} \sim |\lambda - \lambda_c|^{-\nu_{\perp}}, \tag{3.31}$$

$$\xi_{\parallel} \sim |\lambda - \lambda_c|^{-\nu_{\parallel}}, \tag{3.32}$$

in which $\xi_{\perp}$ and $\xi_{\parallel}$ are the spatial and temporal correlation lengths, and $\nu_{\perp}$ and $\nu_{\parallel}$ are critical exponents. For simulations in finite systems, the correlations have an upper bound $\xi_{\perp} \approx L$, L being the size of a lattice, which makes the transition point different from the critical point in the thermodynamic limit. From now on, we will refer to the critical quantities for finite systems with subscript p while the subscript c will be reserved to the real critical point for $L \to \infty$. Therefore, Equation (3.31) can be used to estimate λ_c for a series of λ_p at different sizes

$$L \sim (\lambda_p(L) - \lambda_c)^{\nu_{\perp}} \to \lambda_p(L) - \lambda_c \sim L^{-1/\nu_{\perp}}. \tag{3.33}$$

From Equaton (3.33) we can find relations for other quantities, such as the critical density ρ:

$$\rho(\lambda_p, N) \sim L^{-\beta/\nu_{\perp}}. \tag{3.34}$$

Although they are not the focus of this thesis, similar scaling relations for other quantities can be calculated by using the temporal correlation length, Equation (3.32). For complex networks, similar relations can be constructed by considering N instead of L. To the interested reader, References [31] and [32] discuss criticality and NEPTs in

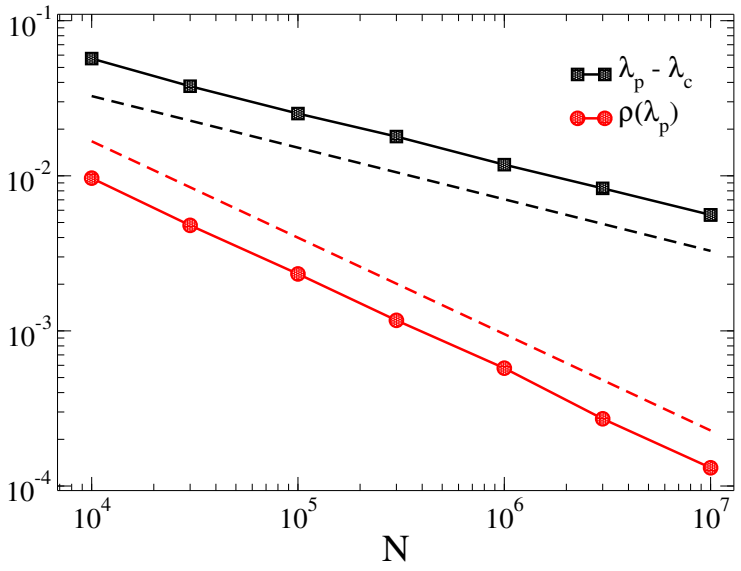


Figure 3.11: Finite size scaling of critical quantities for UCM scale free networks with $\gamma = 2.3$. Dashed lines are power laws with exponents -0.33 (top) and -0.62 (bottom).

detail for several models and substrates. Figure 3.11 illustrates the finite size scaling of λ_p and $\rho(\lambda_p)$ for SIS model on UCM scale free networks with $\gamma = 2.3$ ($\lambda_c = 0$, see Section 3.3.1). By fitting a power law on those points, we obtained the following exponents:-0.33 (epidemic threshold) and -0.62 (critical prevalence).

All these critical exponents can be collected for a certain model, providing a description of the phase transition on a given substrate. It is observed that some models, despite having different microscopic rules, share some symmetries and same critical exponents. These models are aggregated in **universality classes**, such as the Kardar-Parisi-Zhang (KPZ) class [115], that includes deposition phenomena, growing interfaces and random media polymers [116]. The SIS model on regular lattices belongs to Directed Percolation (DP) universality class [31].

Part II

Applications to absorbing state phase transitions

Chapter 4

Simple quasistationary method for simulations of epidemic processes

Related publication:

Simple quasistationary method for simulations of epidemic processes with localized states

Guilherme S. Costa and Silvio C. Ferreira

Computer Physics Communications, 108046 (2021)

During simulations of epidemic processes on random networks, dealing with absorbing states has proven to be a challenge, with several methods being developed to circumvent this difficulty. The quasistationary (QS) analysis is one of them and plays a central role in investigating epidemic processes where localization (when only a subextensive part of the network is active) is present. Several QS methods have been developed, some being very complex computationally and others that, despite being simple, are not able to capture all the details of models with strong localization, such as the aforementioned SIS model on some complex networks. Thus, we developed a QS method that aims to be simple and concomitantly effective in dealing with these behaviors. In this chapter, we start by discussing the QS analysis, describing the most used methods, and detailing the proposed one. Then, we revisit the behavior of SIS on networks, highlighting the localization effects and the activation mechanisms. Finally, we compare the performance of some consolidated QS methods with the proposed method in a myriad of complex networks.

4.1 Quasistationary analysis

During simulations of finite size systems that present an absorbing state, it is possible to show that the system will always visit this state [31, 117] and therefore it will be stuck on it. Thus, if we want to study long term effects of these dynamics, we need to circumvent this problem. A fundamental approach is the **quasistationary (QS) analysis** in which averages at a time t are performed only over a subspace of configurations that did not visit the AS until that time. This concept can be generally extended to encompass perturbations on the system's dynamic in order to prevent the absorbing state [118]. However, these perturbations must be negligible in the thermodynamic limit.

To further develop these concepts, let us define X_t as a Markovian stochastic process in which the number of active[1] individuals n determines the state of the system, including the absorbing state with $n = 0$. Furthermore, we define another similar process X_t^{qs} with some perturbation to get rid of the absorbing state. Them, we can write the dynamical equations for these processes

$$\frac{dP_n}{dt} = \sum_m [W_{n,m}P_m - W_{m,n}P_n] \tag{4.1}$$

$$\frac{dP_n^{\text{qs}}}{dt} = \sum_m [W_{n,m}P_m^{\text{qs}} - W_{m,n}P_n^{\text{qs}}] + F(P_0^{\text{qs}}, P_1^{\text{qs}}, ..., P_N^{\text{qs}}) \tag{4.2}$$

in which P_n and P_n^{qs} are the probability distributions for X_t and X_t^{qs}, respectively, and F is some perturbation to prevent the AS. Since F may contain nonlinear terms, Equation (4.2) is not necessarily a master equation. As aforementioned, it is no easy task to solve these equations analytically. So, we will use a computational approach to calculate P_n^{qs} during the simulations.

We start by setting an initial condition σ_0 to the system and let it evolve during a relaxation time t_{rlx} to allow the convergence to the QS states of the dynamics [117]. Then, during an averaging time t_{av}, we compute P_n^{qs} as

$$P_n^{\text{qs}} = \frac{1}{t_{\text{av}}} \int_{t_{\text{rlx}}}^{t_{\text{av}}+t_{\text{rlx}}} \delta_{s,n(t)} dt \tag{4.3}$$

[1] In epidemiology jargon, we use **infected** and **susceptible** as the possible states. In order to generalize the analysis for other processes, we will adopt the nomenclature active/inactive.

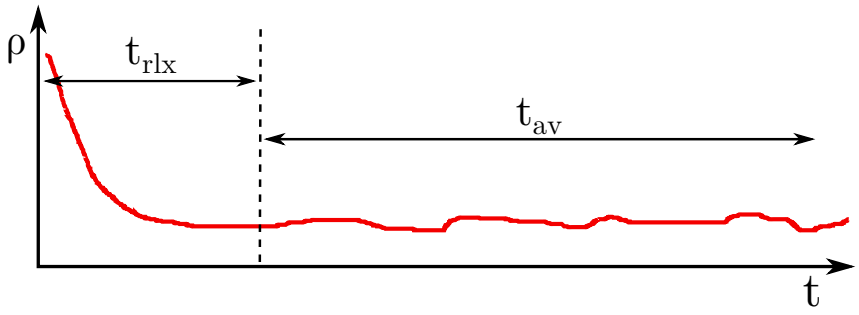


Figure 4.1: Evolution of ρ in a quasistationary simulation, indicating the relaxation and averaging times.

in which $n(t)$ is the number of active individuals at time t and $\delta_{i,j}$ is the Kronecker's delta. These different steps are illustrated in Figure 4.1.

Computationally, P_n^{qs} can be calculated as

$$P_n^{\text{qs}} \to P_n^{\text{qs}} + \frac{\delta_{s,n(t)}\Delta t}{t_{\text{av}}}, \tag{4.4}$$

in which Δt is the time step defined using the aforementioned Gillespie algorithm (Section 3.3.2). From P_n^{qs}, we can calculate some important metrics for the dynamics, such as the moments of this distribution:

$$\langle \rho^k \rangle = \frac{1}{N^k} \sum_{i=1}^{N} i^k P_i^{qs}. \tag{4.5}$$

In particular, the first moment $\langle \rho \rangle$ is the fraction of active individuals, or average QS epidemic prevalence, and act as an order parameter for this dynamic. We can also calculate the dynamic susceptibility [5] as

$$\chi = N \frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle}, \tag{4.6}$$

that measures fluctuations in the order parameter and identify localized and critical regions as the control parameter is varied.

To exemplify, let us consider the SIS model on a scale free network with $N = 10^4$ and $\gamma = 2.3$. In this case, the control parameter is the infection rate λ. Figure 4.2-(a) shows the time series of ρ (after relaxation) for different regions of λ. Note that the fluctuations in those series are small for subcritical and supercritical simulations while they are large for the critical one. So, the susceptibility χ as defined in Equation (4.6) will present a peak in the transition point. Figures 4.2-(b) and (c) show the

behavior of $\langle\rho\rangle$ and χ as function of the control parameter λ for different network sizes. For $\langle\rho\rangle$, it can be noted that the system undergoes a phase transition between an inactive state ($\langle\rho\rangle \approx 0$) to an active state ($\langle\rho\rangle \neq 0$). The transition point can be determined accurately by inspecting the χ curves, which show the aforementioned peaks. Simulations for different sizes shows how the epidemic threshold depends on the system size. In particular for SIS on scale free networks, we saw that the $\lambda_p \to 0$ as $N \to \infty$ [5]; see Section 3.3.1.

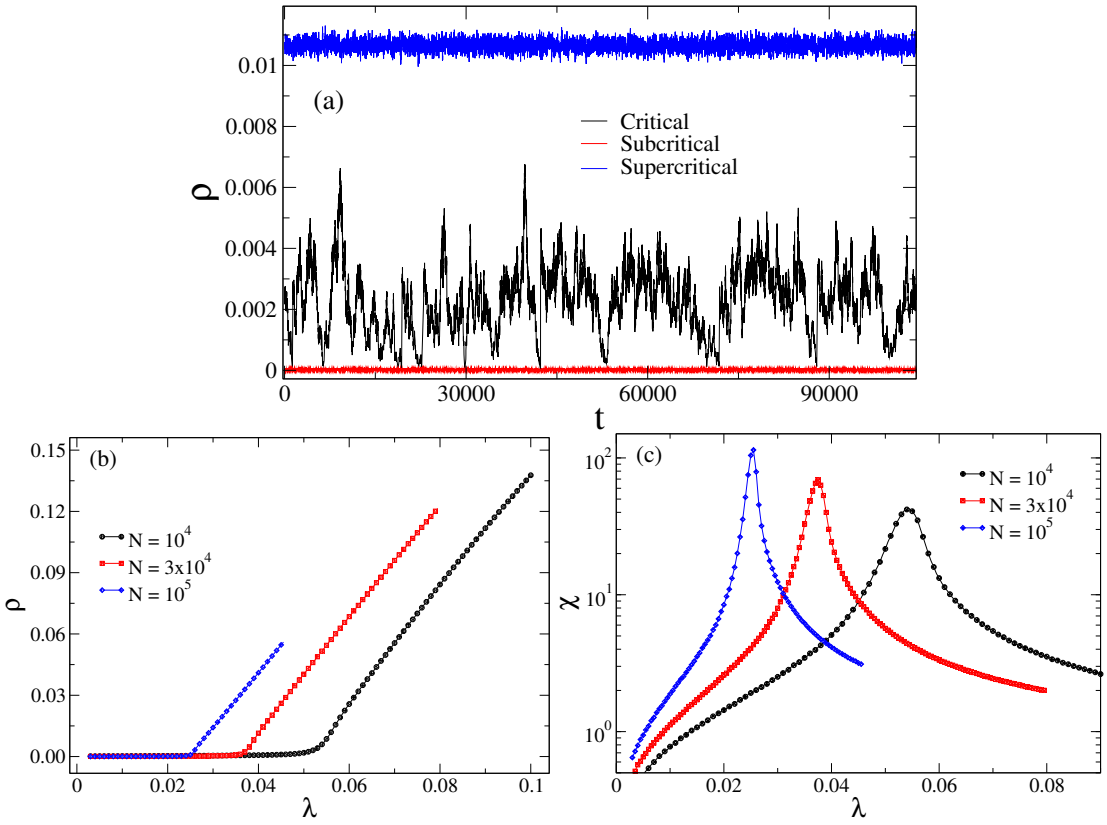


Figure 4.2: Quasistationary simulations of the SIS model on synthetic scale-free networks with $\gamma = 2.3$. (a) Time series of ρ for different regimes: subcritical, critical and supercritical. (b) The order parameter $\langle\rho\rangle$ and (c) susceptibility χ as functions of λ for different network sizes.

The next subsections will be dedicated to discuss different methods to perform the QS analysis.

4.1.1 Standard quasistationary method

The standard quasistationary (SQS) method is the most fundamental technique to perform QS analysis on absorbing phase transitions and consists on restricting the averages over active samples which has not been trapped into the AS up to the current

simulation time [31]. The QS distribution for this dynamic becomes

$$P_n^{\mathrm{qs}} = \lim_{t\to\infty} \frac{P_n(t)}{P_s(t)}, \tag{4.7}$$

in which $P_s(t) = \sum_{n\geq 1} P_n(t) = 1 - P_0(t)$ is the survival probability until a given time t. This strategy have problems when analyzing subcritical or critical series of finite-size systems. Since the dynamic often falls into absorbing states, this method provides small and noisy intervals of stationary data. However, de Oliveira and Dickman [34] proposed a computational method to overcome these limitations. This method consists in jumping to previously visited configurations every time the system falls into the absorbing state, accordingly to the QS distribution. In Reference [119], the authors shows that this method converges to the actual QS state of the system.

Since the QS distribution is not known *a priori*, we must emulate it computationally. This can be done by saving copies of visited active configurations during the simulation and randomly picking one of them every time the system falls into the AS. Two parameters are important for the SQS simulations: the number of saved configurations n_{conf} and the probability per unit of time of saving copies p_{rep}.

The calibration of these parameters is crucial to the SQS performance due to finite computational resources. For small n_{conf}, the dynamic may get biased towards a subset of recent configurations that may not represent the actual QS state while for large n_{conf}, RAM allocation may become a problem and the memory of σ_0 takes too long to be erased. For p_{rep}, we also have problems: if it is small, the convergence to QS state takes too long due to slow loss of memory of σ_0. On the other hand, if it is too large, the simulations can also be biased toward recent configurations and saving them will demand too much computational power. After previous experimentation, we set $n_{\mathrm{conf}} \simeq 100$ and $p_{\mathrm{rep}} = 10^{-2}$ per unit of time. To simulate the SQS method, we follow a relatively simple algorithm:

i). Set a initial configuration σ_0. In our simulations we start with all network nodes infected.

ii). During the simulation with probability per unit of time p_{rep}, we save the system's configuration on a list.

iii). If the list is full, we start replacing its configurations by new ones at random.

Therefore, more probable configurations will appear frequently in the list, which erases the influence of σ_0.

iv). When the system falls into an absorving state, one saved configuration is randomly chosen and used to reactivate the system. Then, the simulation continues until a new visit to an absorbing state occur.

4.1.2 Reflecting boundary condition

The reflecting boundary condition (RBC) method is simple: when the system visits an AS, it jumps to the immediately preceding configuration at a certain rate (usually the inverse of the time unit). Based on RBC, other simple methods can be constructed by considering the random reactivation of nodes. The external weak field (EWF) method, in which vertices are spontaneously reactivated with a small rate (vanishing in the thermodynamic limit), is an widely used variation that was proven to be equivalent to RBC on References [118] and [36]. Other simple variation that was also equivalent to original RBC is the random reactivation of a single vertex when the AS is visited. From now on, every time the RBC method is referred, we are considering the reactivation of one random node when the system falls into absorbing state.

Despite being computationally simple, the RBC and its variations do not reproduce the QS distributions obtained with standard quasistationary method, specially if there are strongly localized effects [118], as addressed in the next section. Targeted reactivation methods may be developed in contrast with random reactivations, accordingly to the dynamic peculiarities. For example the hub reactivation [118, 35], in which we reactivate the node with highest k, is an important one for SIS model due to the key role that hubs and outliers have on this model transition.

4.1.3 Reactivation per activity time

Inspired by simple reactivation dynamics, we propose a method that aims at capturing the localization effects on networks as the SQS method does, but with significantly less computational and algorithmicaly costs. In this method, named reactivation per activity time (RAT), randomly selected vertices are reactivated accordingly to their history of activity. Thus, nodes that were active for longer times are more

probable of being reactivated. During the simulations, the activity time of a node i at time τ is given by

$$T_i^{(\mathrm{a})}(t) = \int_0^t \sigma_i(\tau) d\tau, \tag{4.8}$$

in which $\sigma_i = 1$ if vertex i is active and $\sigma_i = 0$, otherwise. For epidemic models, we set that a node is active if it is infectious and inactive otherwise. The method may be generalized fo different models by adapting the meaning of activity accordingly to their need. The time $T^{(\mathrm{a})}$ of each vertex can be exactly calculated by subtracting the last time that a node became active from the current one when it is inactivated. For the SIS model, this step can be simplified by incrementing $T_i^{(\mathrm{a})}$ by $1/\mu$ every time the vertex i is healed, since $1/\mu$ is the average time that a vertex remains infected and the healing process is spontaneous.

The average number of reactivated nodes depends on the overall average activity, computed as

$$n_{\mathrm{a}}(t) = \frac{1}{t} \sum_{i=1}^{N} T_i^{(\mathrm{a})}, \tag{4.9}$$

in which t is the simulation time. As in the SQS method, the relaxation time must be large enough to erase the influence of σ_0 and allow the convergence of n_{a}. Figure 4.3 shows the evolution of $n_{\mathrm{a}}(t)$ on simulations of critical ($\lambda \approx \lambda_p$) SIS on scale-free networks with $N = 10^6$ and two values of degree exponent: $\gamma = 2.3$ and 2.7. It can be noted that after $t \approx 10^5$ the number of reactivated nodes converges to a constant value.

Computationally, we can only reactivate an integer number of vertices while n_{a} can assume real values. To take this into account, we use the floor $\lfloor x \rfloor$ and ceiling $\lceil x \rceil$ functions, defined as the greatest integer less than or equal to x and least integer greater than or equal to x, respectively. In this way, the number of reactivated nodes will be $\lceil n_{\mathrm{a}} \rceil$ with probability $n_{\mathrm{a}} - \lfloor n_{\mathrm{a}} \rfloor$ and $\lfloor n_{\mathrm{a}} \rfloor$ with probability $\lceil n_{\mathrm{a}} \rceil - n_{\mathrm{a}}$. Although for critical simulations this difference is negligible, for subcritical simulations, in which the number of reactivated particles is small, this difference is relevant.

Vertices are reactivated proportionally to their total activity time T_i. The selection of which vertices will be reactivated can be done using the rejection method, in which

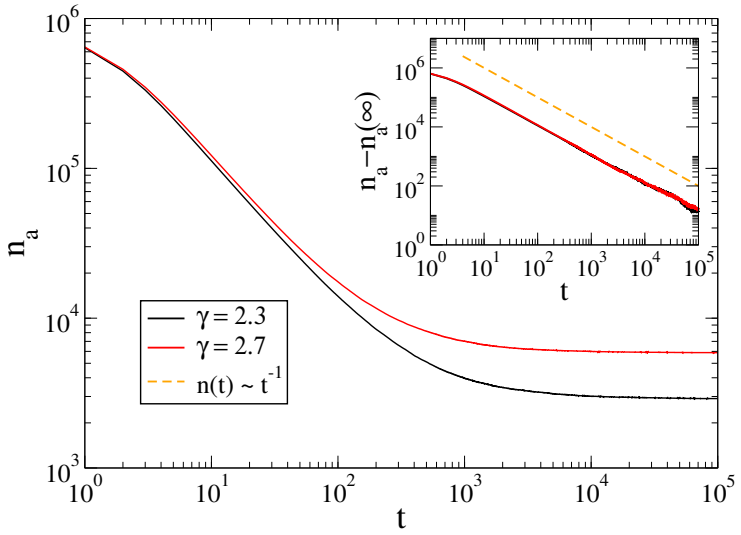


Figure 4.3: Evolution of the average number of active individuals in SIS simulations running on power-law networks with $N = 10^6$ vertices and degree exponents $\gamma = 2.3$ and 2.7. Inset shows the same data subtracted from the stationary value $n_a(\infty)$. The simulation was run for infection rates $\lambda = 0.0115$ and $\lambda = 0.0362$ for $\gamma = 2.3$ and $\gamma = 2.7$, respectively, which give the maximum value of the dynamical susceptibility.

a randomly chosen vertex is accepted with probability

$$q_i = \frac{T_i^{(a)}}{\max_j \left\{ T_j^{(a)} \right\}}. \tag{4.10}$$

This method can be computationally costly if the system often visits the absorbing state and the activation times $T_i^{(a)}$ has a broad distribution. Some optimizations to circumvent this problem are provided on Appendix A.1.

4.2 Activation mechanisms on SIS

The SIS model on power-law networks present remarkable and interesting properties. As previously mentioned, the epidemic threshold is asymptotically null [41, 120] on thermodynamic limit, implying that for any infection rate there will be a nonzero stationary prevalence as $N \to \infty$.

For finite networks, the dynamic near the transition point is characterized by localized activations [99, 38, 121, 122, 123], i.e., only a subextensive portion of the system is active, in contrast with usual phase transitions, that occurs collectively [124].

In the case of random power-law networks, we have two main structures that plays

a central role on these localization: stars (hubs) and maximum k-core subgraphs [89, 125]. The high degree hubs promotes a reinfection mechanism, in which the center infect its leaves and vice-versa. Due to high k, the center may spread the infection to many leaves, virtually guaranteeing that one of them will reinfect it. Figure 4.4 illustrates this mechanism. The epidemic sustained on k-core is somewhat similar: a densely connected core that maintain the infection within it and neighbors. For real networks, more complex scenarios, such as cliques, may rule the epidemic activation [126].

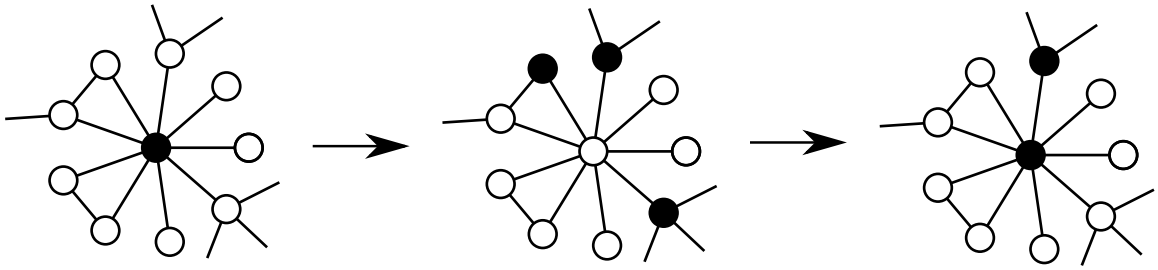


Figure 4.4: Typical SIS dynamic on a star subgraph. Empty nodes are susceptible and black ones are infected.

Particularly for SIS, those different activation mechanisms depend on the heterogeneity level of the network [125]. Considering uncorrelated networks with degree distribution $P(k) \sim k^{-\gamma}$, we have three relevant intervals for the degree exponent:

- $2 < \gamma < 2.5$: networks with hubs that form a densely connected core of high degree. Therefore, this activation is led by the maximum k-core. In our simulations we chose networks with $\gamma = 2.3$ to represent this interval.

- $2.5 < \gamma < 3$: networks with hubs that were not sufficiently connected to form a dense core. Thus, this activation is led by the stars. We chose $\gamma = 2.8$ to represent this range.

- $\gamma > 3$: Similarly to the previous case, in which some isolated hubs are responsible for the localized activation, these transitions are also led by the stars. In addition, these networks may have outliers with degree much larger than the rest and may generate multiple transitions [38]. We chose specifically a network with $\gamma = 3.5$ and a outlier with degree $k = 1837$ to represent this group.

4.3 Comparison of QS methods

Except when mentioned, in all comparisons we performed simulations of the SIS model with all vertices infected as initial condition. For the sake of simplicity we adopted $\mu = 1$ (other values for this parameter just alters the time scale). The averages were computed during an averaging interval $t_{\text{av}} = 3 \times 10^7 \mu^{-1}$ after a relaxation time $t_{\text{rlx}} = 10^7 \mu^{-1}$. For some very supercritical simulation, since the system does not fall into the absorbing state, these simulation times are reduced by one order of magnitude. It worth notice that these parameters are intrinsic to all QS methods and the larger the more accurate. This rule does not hold for specific parameters, such as p_{rep} for SQS.

The simulations were performed on several types of complex networks, with different levels of localization, to verify the robustness of RAT in comparison with SQS and RBC. We started with simple graphs in which the localization is well controlled, the RRNs with hubs. Then, we passed to more complex structures such as synthetic power-law networks, multiplex and real correlated networks.

Random regular networks with a hub

The behavior of the SIS model on RRNs of degree m is predictable. The system undergoes a phase transition at a finite epidemic threshold $\lambda_c \approx \dfrac{1}{m-1}$ [127]. By adding a hub with degree $m \ll K \ll N$ on a RRN, we expect to see a localized activation, in which only the hub and its neighborhood are active, in addition to the global collective one [128]. In Figure 4.5 we have the dynamical susceptibility χ versus the infection rate λ on a RRN ($m = 4$) plus a hub ($K = \sqrt{N}$) for different network sizes, using the RAT method. It can be noticed the two peaks on each curve, representing both activations. The right hand peaks represent the global transition and those moving towards zero represent the localized activation. It is worth remembering that $\lambda_p \approx 1/\sqrt{K}$ for a star subgraph. Thus, $\lambda_p \to 0$ as $N \to \infty$.

Now that we are aware of the typical behavior of SIS model in RRNs with hubs, we can compare the methods. In Figure 4.6 we have a comparison between the three QS methods simulated on the same network. Susceptibility χ and prevalence ρ are shown in Figures 4.6-(a) and (c) for a network with $N = 10^6$ and $K = 10^3$. It can be noted that

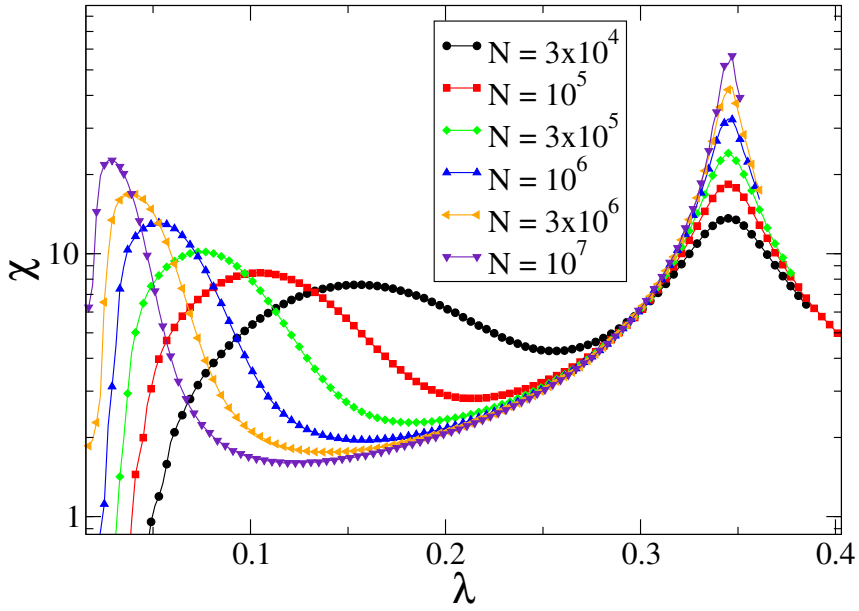


Figure 4.5: Susceptibility χ as a function of the infection rate for RRNs ($m = 4$), of different sizes N (indicated in the legends) plus a hub of degree $K = \sqrt{N}$ using the RAT method.

RAT and SQS curves match each other almost perfectly except in the very subcritical region, in which $\langle\rho\rangle \sim 1/N$. Moreover, the peak related to the localized activation (left one) is distorted by RBC with respect to SQS and RAT. The finite size scaling of critical quantities (λ_p and $\langle\rho(\lambda_p)\rangle$) for the left peak confirm these conclusions, as show in Figures 4.6-(b) and (d). Assuming a scaling law $\lambda_p \sim N^{-\phi}$, we obtain $\phi = 0.21$ for both RAT and SQS whereas $\phi = 0.26$ for RBC. For QS prevalence, we obtain $\langle\rho(\lambda_p)\rangle \sim N^{-1.20}$ for RBC, differing from those of SQS and RAT, $\langle\rho(\lambda_p)\rangle \sim N^{-0.90}$. In the present thesis, these exponents are not intended to determine a universality class, but to indicate different scaling behaviors. Therefore, uncertainties are presented.

Synthetic power-law networks

Considering now the power-law networks discussed in Section 2.3, we investigate the QS methods on synthetic networks constructed using the UCM [94] algorithm with $k_{\min} = 3$. The exponents for $P(k)$ were chosen to encompass the different activation mechanisms of the SIS model presented in Section 4.2: $\gamma = 2.3$ and $\gamma = 2.8$ with structural cutoff $k_c = \sqrt{N}$ and $\gamma = 3.5$ with cutoff $k_c = N$ to favor the occurrence of outliers.

Figures 4.7-(a) and (b) present the susceptibility and QS epidemic prevalence ver-

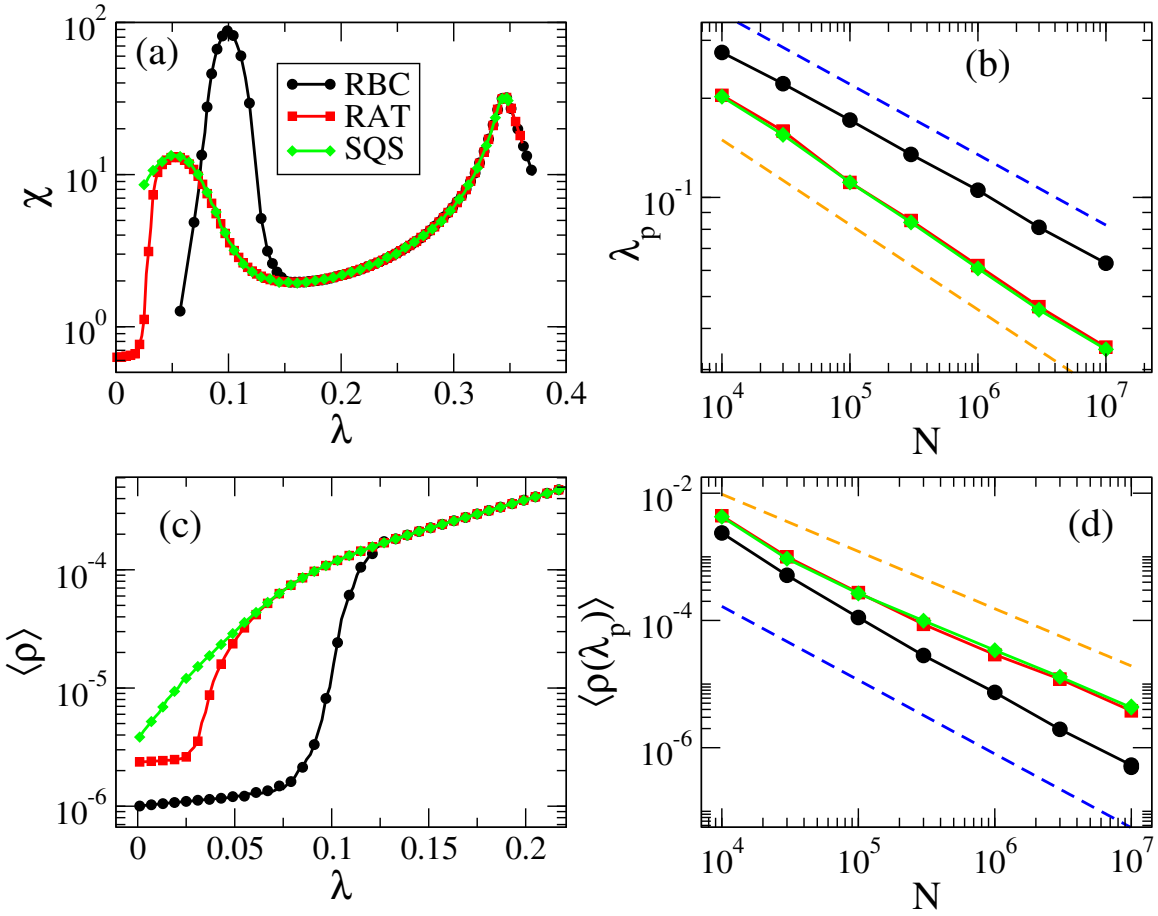


Figure 4.6: Quasistationary quantities for a RRN ($m = 4$) with $N = 10^6$ vertices plus a hub of degree $K = 10^3$ obtained with different QS methods. (a) Susceptibility and (c) prevalence versus infection rate. Finite-size scaling for the (b) position of left susceptibility peak and (d) the corresponding QS prevalences. Dashed lines represent power-laws with exponents -0.27 (top) and -0.21 (bottom) in (b) and -0.90 (top) and -1.2 (bottom) in (d).

sus infection rate computed using different SQS methods on power-law networks with $N = 10^6$ vertices and degree exponent $\gamma = 2.3$ and $\gamma = 2.8$, respectively. Curves for the RAT method deviate from the SQS ones only for the very subcritical phase while the RBC deviates on both subcritical and critical phases. All methods are equivalent in the very supercritical region since absorbing states are not visited for the investigated simulation times. Figure 4.8 presents the FSS of the epidemic thresholds and QS prevalence, considering two values $\gamma = 2.3$ and 2.8. All methods provide equivalent epidemic thresholds with the same scaling for both γ values, while the scaling of the prevalence using RBC deviates from RAT and SQS method; the last two match each other almost perfectly.

To tackle stronger localization effects, simulations on a synthetic network with

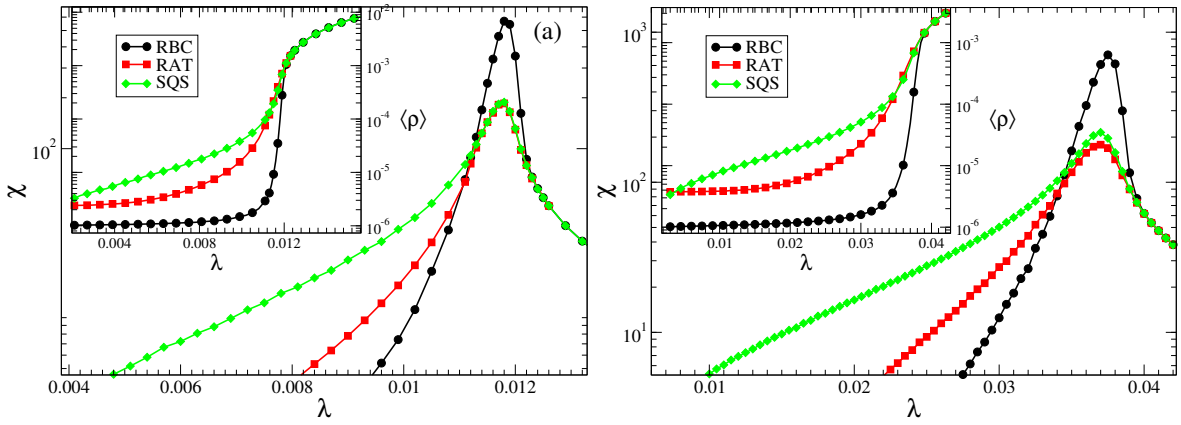


Figure 4.7: Susceptibility versus infection rate for a UCM network with $N = 10^6$ vertices and degree exponent $\gamma = 2.3$, considering three QS methods. Inset shows the QS prevalence as a function of the infection rate.

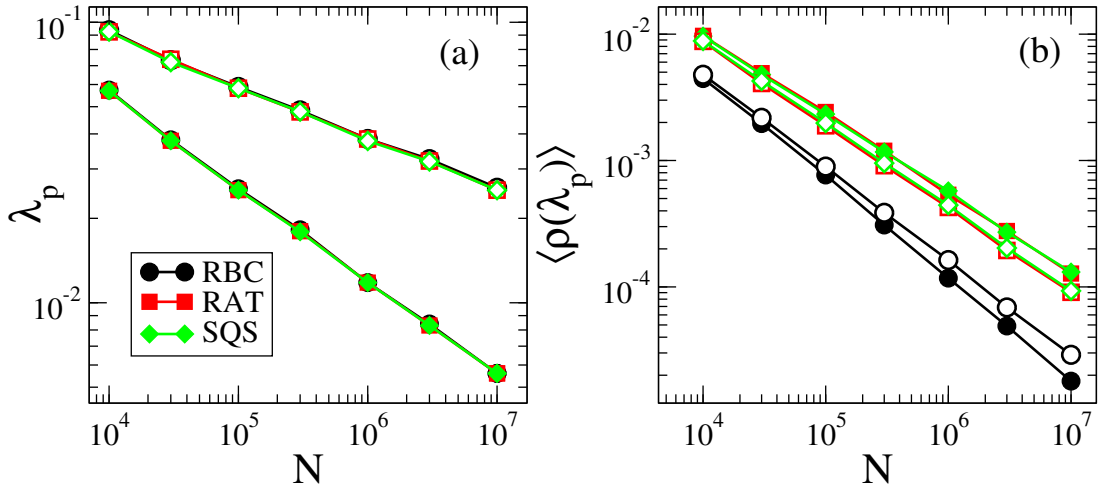


Figure 4.8: Finite-size scaling for the (a) epidemic threshold and (b) the corresponding QS prevalence of the SIS model on UCM networks with $\gamma = 2.3$ (closed symbols) and $\gamma = 2.8$ (open symbols). Averages were computed over 35 networks and error bars are smaller than symbols.

$N = 10^6$ vertices and degree exponent $\gamma = 3.5$ was considered. We selected a sample with an outlier of degree 1837, as can be seen in the degree distribution shown in Figure 4.9(a). Comparing the curves obtained with different methods in Figure 4.9(b), similarly to the case of a RRN plus a hub, RAT and SQS methods are equivalent except in a very subcritical region while RBC deviates substantially in the whole interval of λ where epidemic activity is highly concentrated in the outlier's neighborhood, including the position of the peak which is significantly displaced. All methods match in the secondary peak that corresponds to the global activation of the epidemics since falling into the absorbing state is extremely rare, due to activation of the hub.

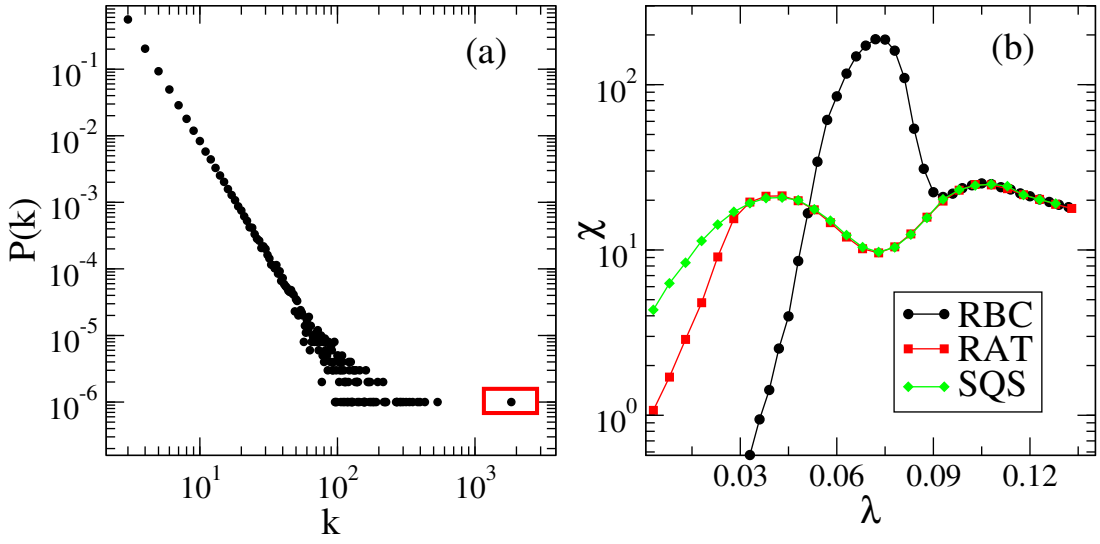


Figure 4.9: (a) Degree distribution of a network with $\gamma = 3.5$ possessing an outlier of degree 1837 highlighted in the box. (b) Susceptibility curves obtained using different QS methods for the network whose degree distribution is shown in (a).

Multiplex networks

The multiple layered structures of multiplex networks combined with the possibility of regulating the heterogeneity degree of each layer separately makes them a perfect substrate for studying localization effects, see Section 2.4 for a brief description of multiplex networks. Additionally to the infection and healing rates λ and μ, the SIS model on multiplex networks have an additional rate η, namely the vertical transmission rate, in which one node in a layer infects itself in another layer. Additional details for the simulation of the SIS model on multiplex networks can be found on Appendix A.2.

In our simulations we considered a three-layered network (layers are fully connected to each other, as in Figure 2.18-(b)) with power-law degree distributions of exponents $\gamma = 2.3$, 2.6, and 2.9. Each layer has $N = 10^4$ vertices. In these simulations, we vary λ while fixing the ratio $\frac{\eta}{\lambda}$. According to Reference [99], multiple transitions, related to the activation of each layer individually, take place for a sufficiently small ratio $\frac{\eta}{\mu} = 0.002$. Figure 4.10 compares the susceptibility curves for this multiplex network. It can be seen that all QS methods capture the multiple transitions approximately at the same values of infection rate. However, as already seen in the previous cases, SQS and RAT present small diferences before the activation of the first layer whereas RBC deviates strongly around the first peak.

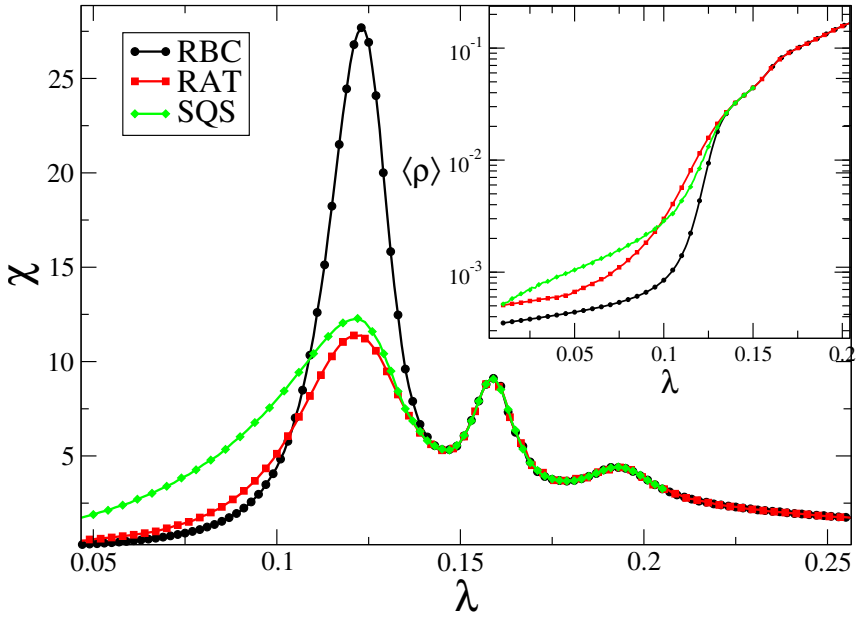


Figure 4.10: Susceptibility as function of the infection rate for a multiplex network with three layers ($\gamma_1 = 2.3$, $\gamma_2 = 2.6$, and $\gamma_3 = 2.9$), $\frac{\eta}{\lambda} = 0.002$ and $N = 10^4$ considering three QS methods. Inset: epidemic prevalence as function of the infection rate.

Real networks

To conclude, we compare the QS methods on SIS simulations on some real correlated networks, in contrast with those generated by UCM. We selected six networks from [85] based on previous exotic behaviors of the susceptibility curves, generated with SQS method. Some properties of these networks can be seen in Table 4.1.

Network	N	$\langle k \rangle$	$\langle k^2 \rangle$	P	Q
USPowerGrid	4941	2.66	10.33	0.003	0.93
GR-QC	4158	6.46	116.1	0.639	0.84
Cora	23166	7.69	182.3	-0.055	0.79
DBLP-Colab	1137114	8.82	366.6	0.096	0.75
USPatents	3764117	8.77	187.8	0.167	0.81
WebStanford	255265	15.2	30897	-0.115	0.92

Table 4.1: Structural properties of real networks. From left to right: name of the network, number of vertices, average degree, second moment, Pearson coefficient, and modularity.

The existence of modular structures with connectivities within modules differing from each other are expected to induce multiple peaks in susceptibility curves, due to localized activation of individual communities [38, 40]. So, this exotic behavior of χ curves is expected from these modular networks. The chosen real networks have a

wide range of sizes, heterogeneity and degree correlations (measured by the Pearson coefficient P).

In Figure 4.11 we show the susceptibility curves for the six chosen real networks. It can be noticed that all QS methods are able to capture the same number peaks and approximately at the same location. However, as reported for synthetic networks, the RBC distorts the first susceptibility peak whereas RAT matches the SQS curves.

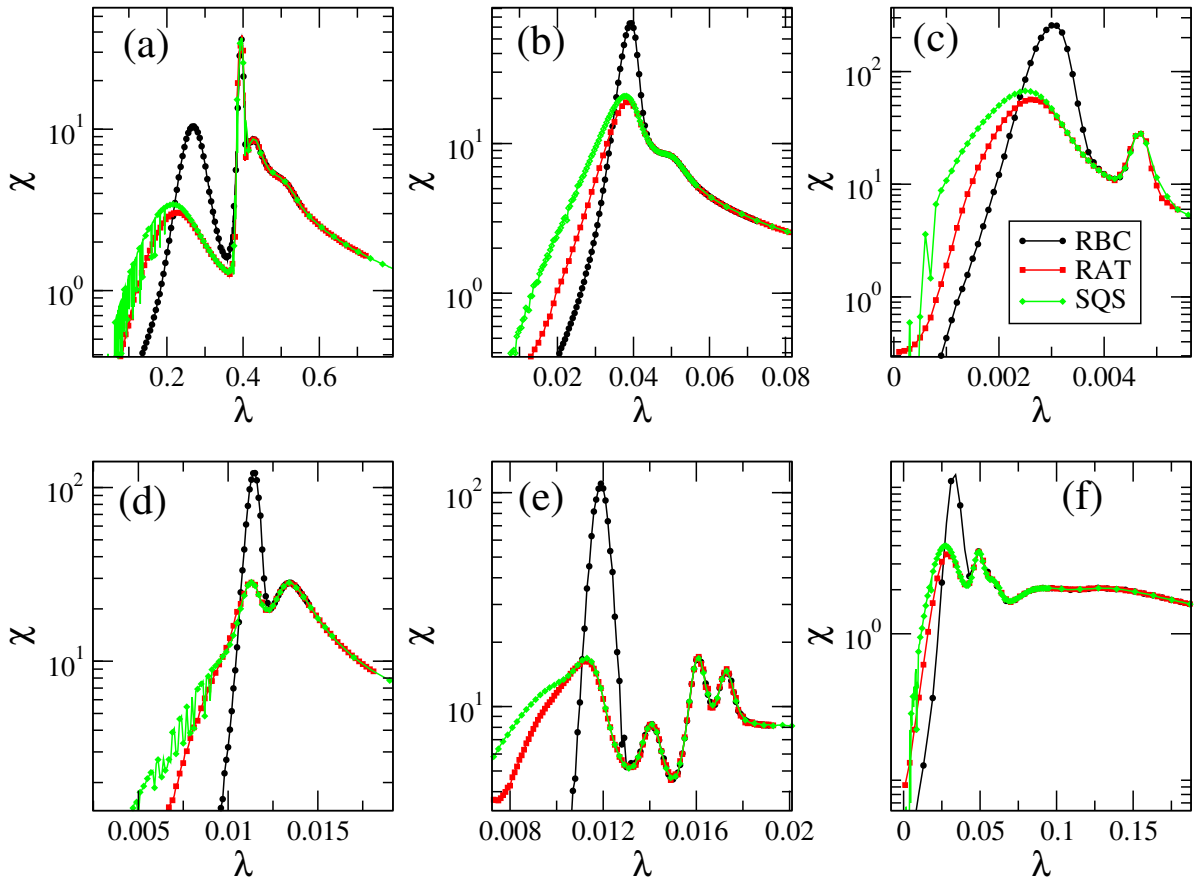


Figure 4.11: Comparison of susceptibility curves obtained for the SIS model on different real networks using different QS methods. Results for the networks given in table 4.1 are presented: (a) USPowerGrid; (b) Cora, (c) WebStanford, (d) DBLP-Colab, (e) USPatents and (f) GR-QC.

4.4 Computer performance of QS methods

In addition to the characterization of transitions, the computational efficiency of the methods is fundamental. In fact, a QS method become uninteresting if it is more complex or inefficient than existing methods, despite being able to reproduce the same localized activation. In terms of implementation and programming, RAT lies

between RBC and SQS. Not so simple as RBC but has no calibration parameters as SQS. Moreover, we analyzed two key ingredients for computational performance: RAM load and CPU times. We performed simulations of the SIS model on UCM networks on a workstation with Intel i5 8400 - 2.80 GHz processor and 8 GB of RAM. Codes were written in Fortran and compiled using a non-commercial version of Intel Fortran for Linux. Codes for QS simulations of SIS using RAT were made publicly available in both Fortran and Python [129].

To evaluate the amount of necessary RAM to simulate the SIS model, we constructed a UCM network with exponent $\gamma = 2.3$ and size $N = 10^7$. For SQS simulations, we choose $n_{\text{conf}} = 100$ copies to be stored. While RBC and RAT needed about 1GB of RAM to store all information, SQS used up to 5 GB, due to $n_{\text{conf}} \times N$ array to store configurations of the system. For RTA, this array is replaced by a smaller one, with only N elements, to store the activity time of each vertice. However, for SIS model, optimizations could have been done to decrease this usage of RAM, since the configuration of system is uniquely defined by the infectious vertices (those who are not infected can only be susceptible). Although this optimization works very well for the SIS model, it is not general. For models with infinitely many absorbing states [130, 131] or more complex epidemic processes [124], the complete array is needed.

The CPU times required for simulations were compared by running a critical ($\lambda = \lambda_p$) SIS dynamics for a total of 10^9 time steps (not time units), after a relaxation time $t_{rlx} = 10^7 \mu^{-1}$, on UCM networks of different sizes. This change in the simulations was necessary to circumvent a natural favoring of RBC, explained as follows: it can be seen in Figure 4.8 that $\langle \rho \rangle$ near the epidemic threshold is much smaller for RBC than RAT or SQS. Since the time step is larger for a lower number of vertices, see OGA time steps in Subsection 3.3.2, less operations were needed to simulate RBC than RAT or SQS for the same time units. However, this also implies that much less configurations are visited during the simulations, leading to reduced statistics in comparison to RAT or SQS for the same averaging time. By considering now the number of time steps, we can guarantee that equivalent statistics are being simulated for all methods. The CPU times presented in Table 4.2 shows that all methods possess similar efficiencies where RBC is the most efficient while RAT outperforms SQS. In all methods, the relaxation

time does not exceed 10% of the whole simulation time.

Size	RBC	RAT	SQS
$N = 10^4$	5.46	6.63	7.04
$N = 10^5$	7.18	8.25	8.79
$N = 10^6$	10.26	11.54	12.44
$N = 10^7$	16.73	18.19	19.21

Table 4.2: CPU times (in minutes) for the QS simulations of critical ($\lambda = \lambda_p$) SIS model on UCM networks with degree exponent $\gamma = 2.3$. The relaxation time is $t_{rlx} = 10^7\mu^{-1}$ and the averages were computed over 10^9 time steps.

Finally, the fact that SQS have two parameters to calibration present additional complications to its simulations. As aforementioned, a poor choice of parameters may lead to poor sampling and trapping the dynamic onto metastable states, leading to spurious behaviors. This poor sampling is particularly common in, but not restricted to, subcritical simulations. Such an effect can be seen in Figures 4.11(a), (c) and (d) where noisy susceptibility curves are observed before the first activation transition.

4.5 Summary and discussions

Dynamical processes with absorbing configurations are relevant for many biological, chemical and physical systems. In particular, the study of these processes on non-regular substrates, such as complex networks [17], constitutes an interdisciplinary issue with applications to real cases [104, 132, 133, 134]. Simulations of dynamics with absorbing configurations on finite-size systems constitute a challenge since the system always visits an absorbing state for sufficiently long times [31]. To circumvent this difficulty, we may rely on QS analysis, that consists of perturbations on the dynamical rules (negligible in the thermodynamic limit), which prevents the system from getting trapped into the AS. There are different methods to avoid the absorbing configurations such as the SQS method [34], which constrains the sampling to active configurations, and the RBC [118], in which the system returns to the pre-absorbing configuration when an AS is visited.

Considering heterogeneous complex networks, the localized effects on subextensive regions imposes further difficulties in the analysis of QS states [38, 123]. Each method has its pros and cons. The SQS is most general and is able to capture localized epidemic transitions [38, 99, 39] but has a high computational cost and algo-

rithmic complexity. The RBC, on the other hand, is extremely simple an low costly but may distort the localized phases. Aiming to soften these drawbacks, we developed a method called RAT (reactivation per activation time), intended to have pros of both RBC and SQS. The method consists of reactivating vertices proportionally to the amount of time they were active during the whole history of the simulation. For epidemic models, the activity is defined as the total time that a vertice remains infected. By setting an appropriate definition of activity, the method can be easily generalized to other dynamical processes.

To validate RAT method, we performed simulations of the SIS model on a myriad of complex networks, characterized by strong localization effects [99, 123, 38] and compared with RBC and SQS. As a general result, we observed that SQS and RAT were equivalent on all analysis, the epidemic thresholds matched, the finite size scaling provided the identical exponents (within error bars) and the general behavior of susceptibility and prevalence curves were the same. The only differences occurs on very subcritical simulations. It is important to note that this difference is not problematic since we are mostly interested in critical characterization of the dynamics. Comparisons with RBC lead to the general conclusion that RBC performs poorly compared to the other two methods. Depending on the network structure, the epidemic threshold or finite size scaling of cricial quantities can differ for the different QS methods. Regarding the computational complexity, we observed that the RAM load of RAT is highly reduced compared to SQS while the CPU times near to the transition point are slightly smaller.

Although the RAT method was extensively tested on a two state dynamic in this chapter, it can be extended to other cases by adapting the method accordingly to the model details. We expect that the RAT method will be extended to QS analyses of many other processes with more complicated dynamics such as sandpile models [135, 136], pair-contact process [137] and systems with bistability which undergo discontinuous transitions [138].

Chapter 5

Heterogeneous mean field theory for two-species symbiotic processes

Related publication:

Heterogeneous mean-field theory for two-species symbiotic processes on networks
Guilherme S. Costa, Marcelo M. de Oliveira and Silvio C. Ferreira
Physical Review E 106, 024302 (2022)

Cooperative or symbiotic processes, in which two or more dynamics evolve on the same substrate and interact with each other, have a very rich behavior in terms of the phase transitions and criticality. Usually, this kind of approach is extensively used in ecological models of competition and cooperation [50, 51]. However, this idea can be expanded to the context of epidemics if we consider two pathogens propagating in the same place and with interactions among them. A simple model to study cooperation is the two-species symbiotic contact process (2SCP) [52], in which two different species spread on a substrate and interact by having a reduced death rate if both occupy the same site. In the recent years, several papers discussing and characterizing the phase transition of this model were out [52, 53, 55, 54]. In order to contribute to the fully understanding of 2SCP, we developed a heterogeneous mean field theory [14] for it. In addition, we validate the results by performing quasistationary simulations on synthetic complex networks. The majority of technical background for this chapter were already discussed in Chapters 2 and 3, refer to it whenever needed.

5.1 Two-species symbiotic contact process

This model is an extension of the original **Contact Process** (CP) [42], one of the extensively studied models to exhibit an APT. Originally constructed on lattices, in the standard CP an individual may duplicate with rate λ (its offspring occupying a random empty neighboring cell) or spontaneously die with rate μ. One can notice that these rules are rather similar to SIS model mentioned on previous chapters. In fact, for lattices, both models are connected through a shift in the rates: $\lambda_{\mathrm{CP}} = \lambda_{\mathrm{SIS}} * k$, in which k is the number of neighbors. The critical behavior of the models are equally similar, both belonging to the Directed Percolation (DP) universality class. However, in complex networks they have very distinct properties [46].

The 2SCP is defined considering two species (A and B) evolving on the same network. Each node can hold at most one individual of each species. The activation process is exactly the same as CP where both species create clones of themselves with rate λ/k at all neighbors that do not carry one individual of its species. A singly occupied node by either A or B becomes vacant with rate μ. If a node contains two species, a reduced symbiotic death rate $\mu_s < \mu$ is adopted, such that the chance of death for both A and B individuals is reduced. All transitions for the model are illustrated in Figure 5.1. Hereafter, we adopt $\mu = 1$ without loss of generality.

We start discussing the phase transition on 2SCP by reviewing the basic homogeneous mean-field theory [52]. Then, we proceed to discuss our development of the heterogeneous one.

5.1.1 Homogeneous mean field

In this mean field theory, we consider that all sites are statistically equivalent. Therefore, the dynamical equations are constructed in terms of probabilities that the node is in a given state: vacant, occupied by a species A, or B, or both species. Taking into account all possible transitions indicated in Figure 5.1, the following set of equations is obtained

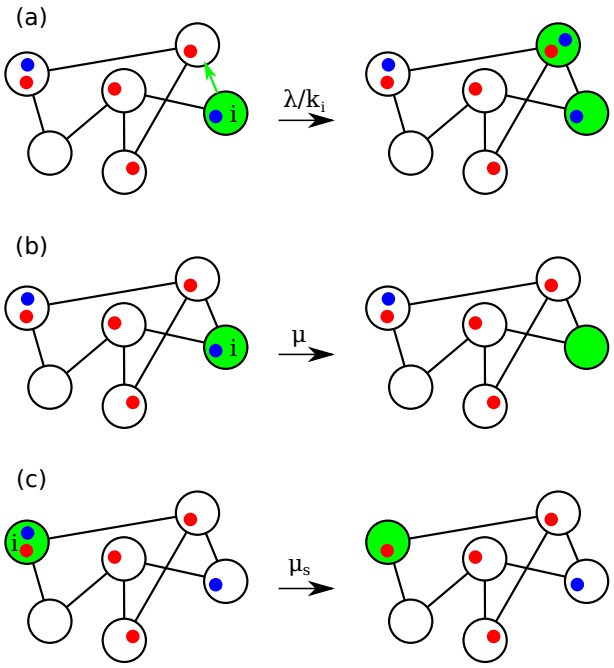


Figure 5.1: Transitions of the 2SCP model on networks where species are represented by red and blue dots while the nodes involved in the transition are depicted in green. (a) Individuals of species A or B in the node i replicate a copy of itself to a neighbor node if allowed (doubly occupation with the same species is forbidden). Spontaneous deaths on nodes that contain (b) one species and (c) both species (symbiosis) happen with rates μ and μ_s, respectively.

$$
\begin{aligned}
\frac{d\rho_0}{dt} &= \rho_A + \rho_B - \lambda\rho_0(\rho_A + \rho_B + 2\rho_{AB}),\\
\frac{d\rho_A}{dt} &= -\rho_A + \rho_{AB}\mu_s + \lambda\rho_0(\rho_{AB} + \rho_A) - \lambda\rho_A(\rho_{AB} + \rho_B),\\
\frac{d\rho_B}{dt} &= -\rho_B + \rho_{AB}\mu_s + \lambda\rho_0(\rho_{AB} + \rho_B) - \lambda\rho_B(\rho_{AB} + \rho_A),\\
\frac{d\rho_{AB}}{dt} &= -2\mu_s\rho_{AB} + 2\lambda\rho_A\rho_B + \lambda\rho_{AB}(\rho_A + \rho_B),
\end{aligned}
$$

in which ρ_0, ρ_A, ρ_B and ρ_{AB} are probabilities that a given node is vacant, occupied by an individual of species A, of species B, or both, respectively. It is important to notice that in the absence of A or B, the equations reduce to standard CP [31]. By considering symmetrical solutions $\rho_A = \rho_B = \rho$ and the closure relation $\rho_A + \rho_B + \rho_{AB} + \rho_0 = 1$, the system is reduced to two independent variables. The stationary solutions are the trivial $\bar{\rho} = \bar{\rho}_{AB} = 0$ and the nontrivial one given by

$$
\bar{\rho} = \frac{\mu_s}{2\lambda(1-\mu_s)}\left[2(1-\mu_s) - \lambda + \sqrt{\lambda^2 - 4\mu_s(1-\mu_s)}\right] \tag{5.1}
$$

and

$$\bar{\rho}_{\mathrm{AB}} = \frac{\lambda\rho^2}{\mu_s - \lambda\rho} \tag{5.2}$$

Analyzing Equation (5.1), we can draw some conclusions regarding the system:

- The nontrivial solution for ρ exists only if $\lambda \geq \sqrt{4\mu_s(1-\mu_s)}$.
- For $\mu_s > 1/2$, $\lambda_c = 1$ is the transition point. Additionally, $\bar{\rho}$ grows continuously.
- For $\mu_s < 1/2$, the critical point $\lambda_c = \sqrt{4\mu_s(1-\mu_s)}$ imply in $\bar{\rho} > 0$, indicating a discontinuous transition.

The discontinuity in solutions for $\mu_s < 1/2$ implies in a bistability region since $\rho = 0$ is also locally stable for $\sqrt{4\mu_s(1-\mu_s)} < \lambda < 1$. The convergence to stationary state depends on the initial condition: for $\rho_{\mathrm{A}}(0), \rho_{\mathrm{B}}(0) \to 1$, the dynamics converges to Equation (5.1) for $\lambda > \lambda^- = \sqrt{4\mu_s(1-\mu_s)}$ while for $\rho_{\mathrm{A}}(0), \rho_{\mathrm{B}}(0) \to 0$, the convergence happens for $\lambda > \lambda^+ = 1$ while the absorbing state remains stable otherwise. The curves $\lambda^-(\mu_s)$ and $\lambda^+(\mu_s)$ are called lower and upper spinodal, respectively. Visual representations of these discontinuity are shown in Figure 5.2. To the interested reader, more accurate homogeneous theories were developed using pairwise interactions [54].

5.1.2 Heterogeneous mean field

In a similar way to SIS degree based theory (Section 3.3.1), we treat vertices within same k as statistically equal. To include this increased resolution on the equations, we standardize our notation as $\rho_k^{\mathrm{X}} \to$ prevalence of individuals of X species (X = A, B, 0 or AB) on sites with degree k. Thus, the equations for the HMF theory become

$$\frac{d\rho_k^0}{dt} = \rho_k^A + \rho_k^B - \lambda\rho_k^0 k \sum_{k'} \frac{(\rho_{k'}^A + \rho_{k'}^B + 2\rho_{k'}^{AB})P(k'|k)}{k'},$$

$$\frac{d\rho_k^A}{dt} = -\rho_k^A + \rho_k^{AB}\mu_s + \lambda\rho_k^0 k \sum_{k'} \frac{(\rho_{k'}^{AB} + \rho_{k'}^A)P(k'|k)}{k'} - \lambda\rho_k^A k \sum_{k'} \frac{(\rho_{k'}^{AB} + \rho_{k'}^B)P(k'|k)}{k'},$$

$$\frac{d\rho_k^B}{dt} = -\rho_k^B + \rho_k^{AB}\mu_s + \lambda\rho_k^0 k \sum_{k'} \frac{(\rho_{k'}^{AB} + \rho_{k'}^B)P(k'|k)}{k'} - \lambda\rho_k^B k \sum_{k'} \frac{(\rho_{k'}^{AB} + \rho_{k'}^A)P(k'|k)}{k'},$$

$$\frac{d\rho_k^{AB}}{dt} = -2\mu_s\rho_k^{AB} + \lambda\rho_k^A k \sum_{k'} \frac{(\rho_{k'}^B + \rho_{k'}^{AB})P(k'|k)}{k'} + \lambda\rho_k^B k \sum_{k'} \frac{(\rho_{k'}^A + \rho_{k'}^{AB})P(k'|k)}{k'}.$$

Each term of these equations is related to a possible event on the model. For ρ_k^0, the first and second terms of the r.h.s of the equation are related to spontaneous death on singly occupied nodes by species A and B, respectively. Third term is related to the creation of a particle on vacant nodes. Passing to ρ_k^{A}, the first term is the spontaneous death of a node that have only species A, the second term is the symbiotic death of species B on a doubly occupied node, third is the replication of species A on vacant nodes and the last is the replication of species B on nodes that contains only A, becoming doubly occupied vertices. Equivalent terms can be found on the equation for ρ_k^{B}. Finally, the dynamic of ρ_k^{AB} has the following terms: the first is symbiotic death of any species on a doubly occupied node, the second is the replication of species B on a node that have only A and the last is the opposite, replication of species A on nodes that have only B.

As in the homogeneous theory, we are looking for symmetric ($\rho_k^{\mathrm{A}} = \rho_k^{\mathrm{B}} = \rho_k$) and stationary solutions ($\frac{d\rho_k^{\mathrm{A}}}{dt} = 0$). For each degree k, we can write the closure equation as $\rho_k^0 + \rho_k^{\mathrm{A}} + \rho_k^{\mathrm{B}} + \rho_k^{\mathrm{AB}} = 1$. In order to advance in the solution, we need an explicit form for the conditional probability $P(k'|k)$. As usual, we consider uncorrelated networks, in which holds Equation (2.12) and sets $P(k'|k) = \frac{k'P(k')}{\langle k\rangle}$. By defining $\beta_k = \frac{\lambda k}{\langle k\rangle}$ and $\varphi = \sum_k(\rho_k^{\mathrm{AB}} + \rho_k)P(k)$, we obtain the following equations for $k = k_{\min}, ..., k_{\max}$

$$\frac{d\rho_k}{dt} - \rho_k + \mu_{\mathrm{s}}\rho_k^{\mathrm{AB}} + (1 - 3\rho_k - \rho_k^{\mathrm{AB}})\beta_k\varphi, \tag{5.3}$$

and

$$\frac{d\rho_k^{\mathrm{AB}}}{dt} = -\mu_s \rho_k^{\mathrm{AB}} + \beta_k \rho_k \varphi. \tag{5.4}$$

Solving these equations in the steady state for ρ_k^{AB} and ρ_k, one obtains

$$\bar{\rho}_k^{\mathrm{AB}} = \frac{\beta_k \rho_k \bar{\varphi}}{\mu_s}, \tag{5.5}$$

and

$$\bar{\rho}_k = \frac{\beta_k \bar{\varphi}}{1 + 2\beta_k \bar{\varphi} + \dfrac{\beta_k^2 \bar{\varphi}^2}{\mu_s}}. \tag{5.6}$$

Substituting on $\bar{\varphi}$ definition, we obtain a self-consistent transcedent equation

$$\bar{\varphi} = \sum_k \left[\left(\frac{\bar{\varphi} \beta_k P(k)}{1 + 2\beta_k \bar{\varphi} + \dfrac{\beta_k^2 \varphi^2}{\mu_s}} \right) \left(1 + \frac{\beta_k \bar{\varphi}}{\mu_s} \right) \right], \tag{5.7}$$

which can be easily solved by bisection method [139] given a degree distribution $P(k)$, providing $\bar{\rho}_k$ and $\bar{\rho}_k^{\mathrm{AB}}$.

The stability loss of the absorbing state $\rho_k = \rho_k^{\mathrm{AB}} = 0$, at λ^+, can be obtained using Equation (5.5) near to the transition point, where both $\rho_k \ll 1$ and $\rho_k^{\mathrm{AB}} \ll 1$. We assume $\rho_k^{\mathrm{AB}} \ll \rho_k$ in Equation (5.6) to obtain the following linearized and closed system for ρ_k

$$\frac{d\rho_k}{dt} = -\rho_k + \frac{\lambda k}{\langle k \rangle} \sum_{k'} \rho_{k'} P(k') + \ldots \simeq \sum_{k'} L_{kk'} \rho_{k'}, \tag{5.8}$$

where

$$L_{kk'} = -\delta_{kk'} + \frac{\lambda k P(k')}{\langle k \rangle} \tag{5.9}$$

is the Jacobian of the linearized system, the same as the single species contact process on uncorrelated networks [140]. The loss of stability of the absorbing state is obtained when the largest eigenvalue of $L_{kk'}$ is zero. In a similar way than SIS calculations on Section 3.3.1, one can show that $\mathbf{v} = k$ is an eigenvector of $L_{kk'}$ with corresponding eigenvalue $\Lambda_0 = 1 - \lambda$. Using the Perron-Frobenius theorem, we can state that Λ_0 is the leading eigenvalue. Therefore, the upper spinodal is $\lambda^+ = 1$.

We define the order parameter as the fraction of nodes occupied by at least one specie, or total prevalence, given by

$$\rho_{\mathrm{T}} = 2\rho + \rho_{\mathrm{AB}}, \tag{5.10}$$

in which

$$\rho = \sum_{k=k_{\min}}^{k_{\max}} \rho_k P(k) \tag{5.11}$$

is the probability that a randomly chosen node is occupied by either A or B species and

$$\rho_{\mathrm{AB}} = \sum_{k=k_{\min}}^{k_{\max}} \rho_k^{\mathrm{AB}} P(k) \tag{5.12}$$

is the probability of double occupation. In Figure 5.2-(a) we present typical hysteresis diagrams for ρ_T as a function of λ using a power-law degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3.5$ and $\mu_s = 0.2$ for two extreme initial conditions: $\rho_{\mathrm{AB}}(0) = 1$ representing a fully occupied substrate and $\rho_{\mathrm{AB}}(0) = 10^{-6}$ which is near to the absorbing state. To compare with the results on uncorrelated networks in Section 5.3, we choose an upper cutoff for the degree distribution as $k_{\max} = \sqrt{N}$, where N is the number of nodes of the network. The hysteresis effect in these curves manifests as bistability regions (hashed area). The phase diagrams in $\lambda \times \mu_{\mathrm{s}}$ parameter space were computed numerically by the numerical solution of Equation (5.7) and presented for degree exponents $\gamma = 3.5$, 2.7 and 2.3 in Figures 5.2-(b) to (d). One can see that the bistability region is close to the homogeneous case for $\gamma = 3.5$, being reduced as the network heterogeneity is increased with smaller values of γ, shrinking in the limit $\gamma \to 2$. The influence of heterogeneity may be thought as follows: the probability of a doubly occupied vertice to replicate both species in one of its neighbors decreases with its degree ($\sim k^{-2}$) due to the randomness in choosing the occupation target. Thus, even hubs being, on average, more active than regular nodes, more hubs dismantle the symbiotic mechanisms by diluting the species in different vertices.

To further investigate this behavior, we performed a continuous approximation for Eq. (5.7) to obtain

$$\bar{\varphi} = \Theta(\bar{\varphi}) = \int_{k_{\min}}^{\infty} \frac{\bar{\varphi}\beta_k(1 + \beta_k\bar{\varphi}/\mu_{\mathrm{s}})P(k)}{1 + 2\beta_k\bar{\varphi} + \beta_k^2\bar{\varphi}^2/\mu_{\mathrm{s}}} dk, \tag{5.13}$$

in the limit $k_{\max} \to \infty$. It can be seen that for $\mu_{\mathrm{s}} \geq 1/2$, $\Theta'(\varphi) > 0$ and $\Theta''(\varphi) < 0$, implying that only a continuous transition is possible, since $\Theta(\varphi)$ is a monotonically

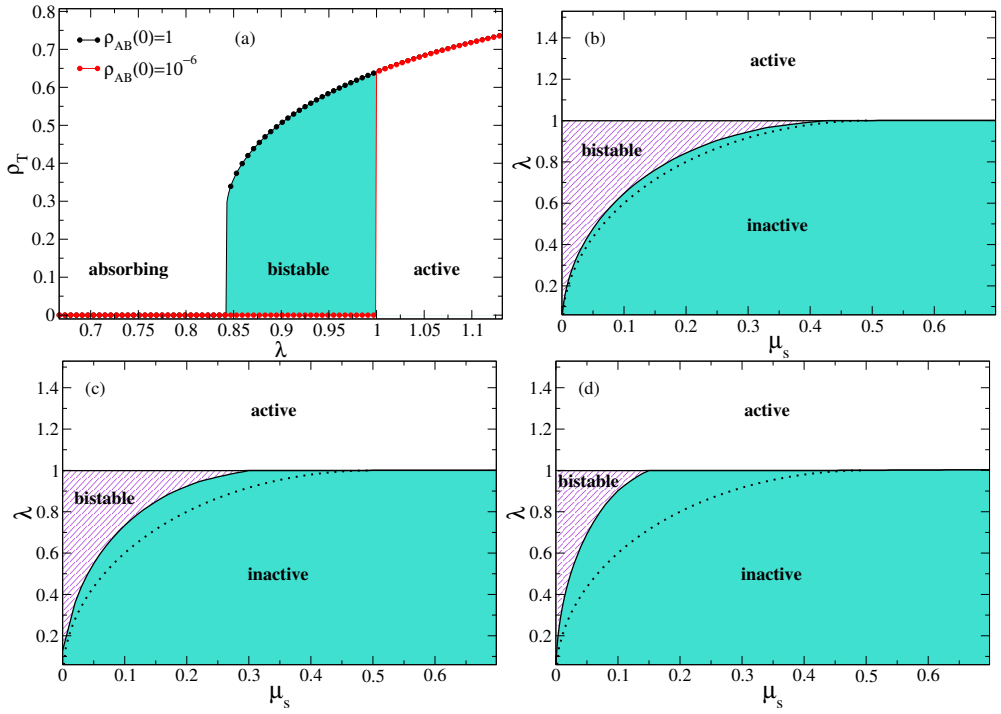


Figure 5.2: HMF theory for 2SCP model on power-law degree distributions with lower and upper cutoffs given by $k_{\min} = 3$ and $k_{\max} = 17320$ ($N = 3 \times 10^8$). (a) Example of hysteresis on 2SCP total prevalence curves for a power law distribution with $\gamma = 3.5$ and $\mu_s = 0.2$. The hashed area indicates the bistability region. Phase diagram for 2SCP considering power-law distributions with (b) $\gamma = 3.5$, (c) $\gamma = 2.7$, and (d) $\gamma = 2.3$. The dotted line represents the lower spinodal for the homogeneous theory $\lambda^- = \sqrt{4\mu_s(1-\mu_s)}$.

increasing function. Thus, if discontinuous transitions occurs, μ_s must be less than $1/2$. Both cases are illustrated in Fig. 5.3 (a) and (b).

By considering the series expansion for small $\bar{\varphi}$, we can find the point when the discontinuous transition turns into a continuous one, i.e., when the gap in lower spinodal vanishes (see Appendix B.1 for detailed calculations):

$$\Theta(\bar{\varphi}) = \lambda\bar{\varphi} + a_{\gamma-1}\bar{\varphi}^{\gamma-1} + a_2\bar{\varphi}^2 + a_3\phi^3, \tag{5.14}$$

where

$$a_2 = \frac{(\gamma-2)^2(1-2\mu_s)}{(\gamma-3)(\gamma-1)\mu_s}\lambda^2, \tag{5.15}$$

$$a_3 = \frac{(\gamma-2)^3(7-8\sqrt{\mu_s})}{(4-\gamma)(\gamma-1)^2\mu_s}\lambda^3, \tag{5.16}$$

$$a_{\gamma-1} = (\gamma-1)\left(\frac{\lambda k_{\min}}{\langle k\rangle\sqrt{\mu_s}}\right)^{\gamma-1}\tilde{\beta} \tag{5.17}$$

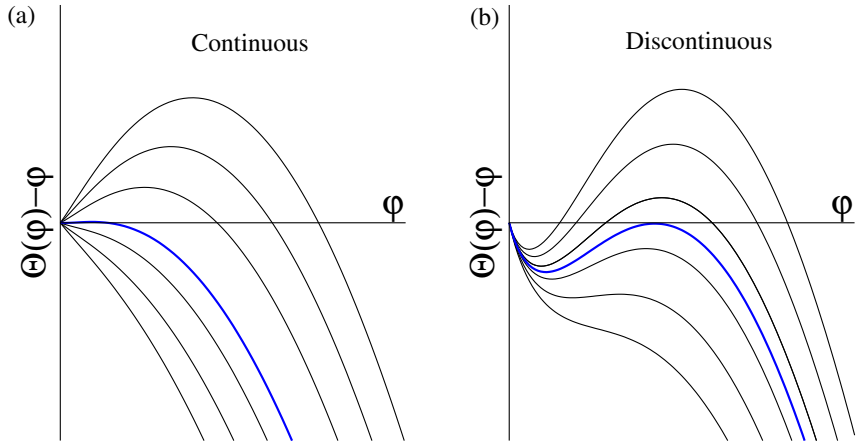


Figure 5.3: Example of a (a) continuous and (b) discontinuous transition through transcritical and pitchfork bifurcations, respectively. The curves correspond to the solution of Eq. (5.7) with (a) $\mu_s = 0.3$ and (b) $\mu_s = 0.05$ for a degree exponent $\gamma = 2.7$. The contagion rate λ is increased from bottom to top. The blue curves are the critical ones.

and

$$\tilde{\beta} = \sum_{n=0}^{\infty} [2(1-\sqrt{\mu_s})]^n \left[\sqrt{\mu_s} + \frac{n-\gamma+2}{n+\gamma-1}\right] \frac{\Gamma(n-\gamma+2)\Gamma(\gamma+n)}{\Gamma(2n+2)}, \tag{5.18}$$

where $\Gamma(x)$ is the Gamma function. The cubic term is negligible for $\gamma < 4$ while the term $\bar{\varphi}^{\gamma-1}$ is negligible otherwise. Solving for the leading term, we obtain the following

$$\bar{\varphi}^* \simeq \begin{cases} \left[\dfrac{(\gamma-1)a_{\gamma-1}}{|a_2|}\right]^{1/(3-\gamma)} & 2 < \gamma < 3 \\ \left[\dfrac{a_2}{(\gamma-1)|a_{\gamma-1}|}\right]^{1/(\gamma-3)} & 3 < \gamma < 4 \\ \dfrac{a_2}{3|a_3|} & \gamma > 4 \end{cases} . \tag{5.19}$$

For $\gamma > 3$, the gap $\bar{\varphi}^*$ vanishes for $a_2 = 0$, which leads to $\mu_s^* = 1/2$. For $2 < \gamma < 3$, the gap vanishes when $a_{\gamma-1} = 0$ or, equivalently, $\tilde{\beta} = 0$. Since this coefficient has a complex dependency on the parameters, we may investigate it numerically. As can be seen in Fig. 5.4, $\tilde{\beta}$ is zero for $0 < \mu_s < 1/2$, depending on the value of γ.

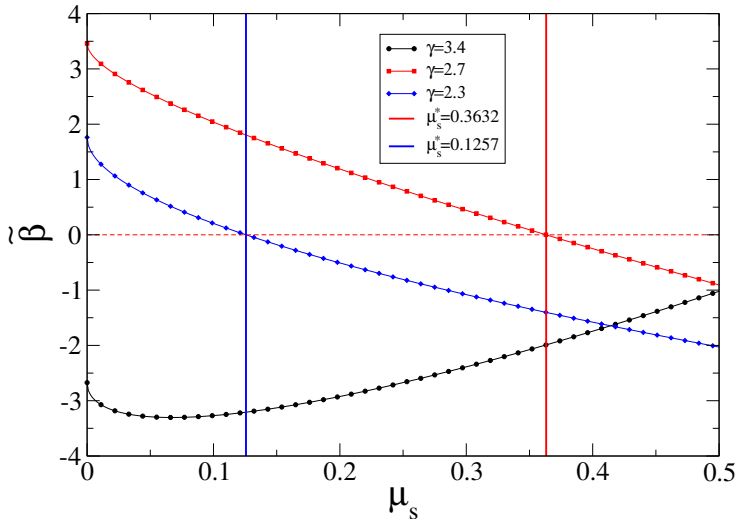


Figure 5.4: Coefficient $\tilde{\beta}$, where $a_{\gamma-1} \propto \tilde{\beta}$, as function of symbiotic parameter μ_s on power-law networks with different values of the degree exponent γ indicated in the legend. Vertical lines are roots of $\tilde{\beta}(\mu_s) = 0$ for $\gamma = 2.7$ and 2.3.

5.2 Numerical analysis and finite size scalings

Stochastic simulations can be performed only in finite size systems. Thus, it is important to understand the finite size dependence of the theory. We investigate the total prevalence curves of different network sizes considering a fully occupied network as initial condition. Figures 5.5-(a) to (c) show the total prevalence as function of infection rate λ for $\mu_s = 0.2$ and different levels of heterogeneity, $\gamma = 2.3$, 2.7 and 3.5. It can be noticed two different finite-size scaling behaviors, depending on the degree exponent. For $\gamma = 2.3$, the transition is discontinuous at an activation threshold $\lambda_p < 1$ for small sizes. However, as N increases, the discontinuity drops towards a continuous transition in the infinite size limit when $\lambda^- \to \lambda_c = 1$ implying in a pseudo threshold for finite sizes. On the other hand, for $\gamma = 2.7$, the discontinuity is sustained in the thermodynamic limit with a threshold converging to $\lambda^- < 1$. Additionally, the convergence can also be seen for $\gamma = 3.5$ but much faster in this case.

Figure 5.6 presents the finite-size scaling for the activation threshold λ_c and the gap discontinuity Δ_p for the same curves shown in Figure 5.5. For $\gamma = 2.3$, the activation thresholds converge to the upper spinodal $\lambda^+ = 1$ and the discontinuity gap decays as a power law $\Delta_p \sim N^{-0.43}$, corroborating the transition continuity as $N \to \infty$. For other degree exponents ($\gamma = 2.7$ and 3.5), the convergence occurs to the

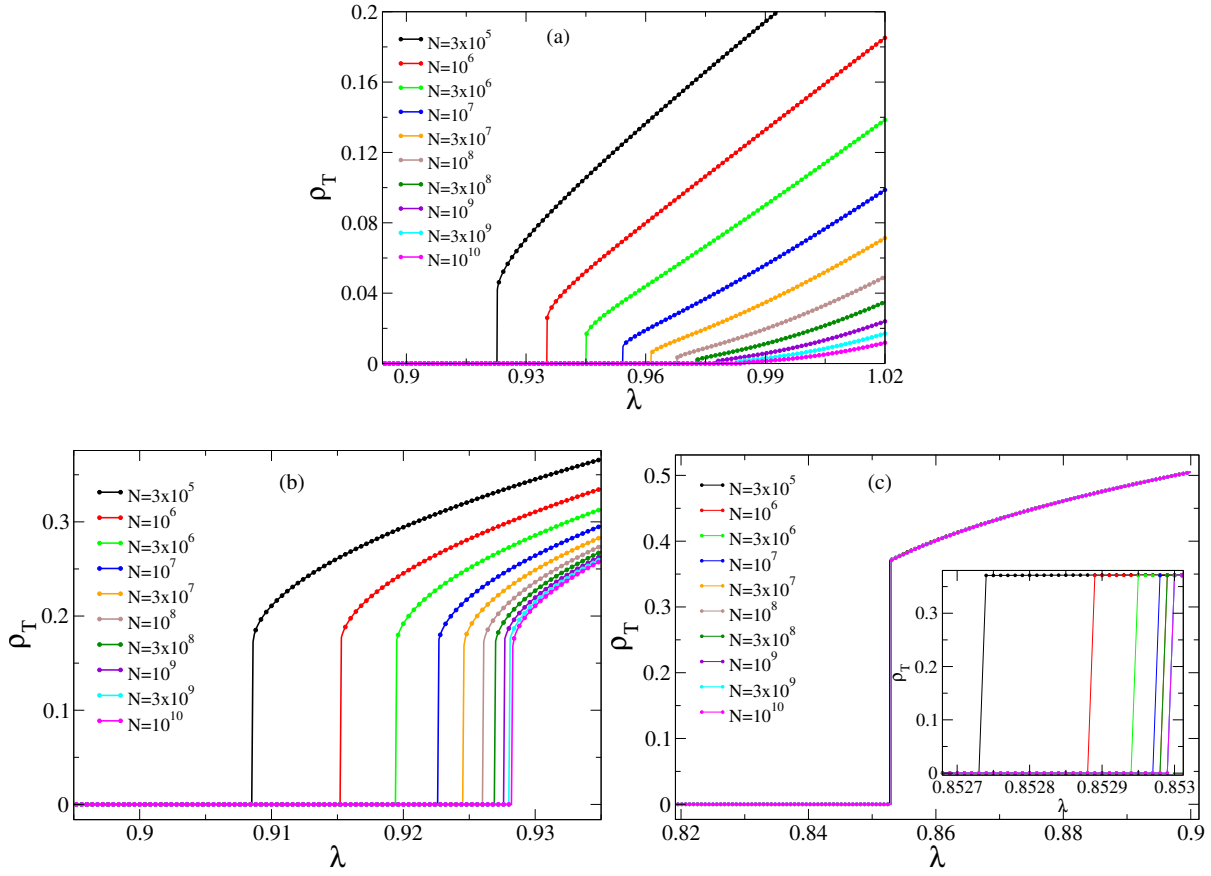


Figure 5.5: Total prevalence ($\rho_T = 2\rho + \rho_{AB}$) in function of λ for $P(k)$representing synthetic scale free networks with $\mu_s = 0.2$ and different sizes. (a) $\gamma = 2.3$, (b) $\gamma = 2.7$. (c) $\gamma = 3.5$

lower spinodals $\lambda^- < 1$, with the discontinuity gap remaining finite.

Delving deeper into the influence of μ_s on the transition behavior of the 2SCP model, we determine the value of the symbiotic coupling μ_s^* that separates the continuous from discontinuous transitions, considering different levels of heterogeneity. As a general behavior, we report that curves for lower values of μ_s tend to a finite value of Δ_p, for higher values it keeps decaying towards 0 within the numerical accuracy of the integration. To estimate the value of μ_s^*, we assume a monotonic approximation using an scaling in the form $\Delta_p(N) = \Delta_p(\infty) + cN^{-b}$ or, more precisely, $z = z_0 + \ln(1 + ce^{-bw})$ with $z = \ln(\Delta_p)$ and $w = \ln(N)$ in double logarithmic form, as can be seen in Figure 5.7-(a). By collecting the values of μ_s^* for several γ, we obtained the phase diagram in the space parameter μ_s versus γ presented in Figure 5.7-(b), in which the theoretical curve is also shown. In general, a very good agreement can be seen between both analysis, validating the method in systems where exact expressions are not available. In addition, Figure 5.7-(b) reinforces the results regarding the

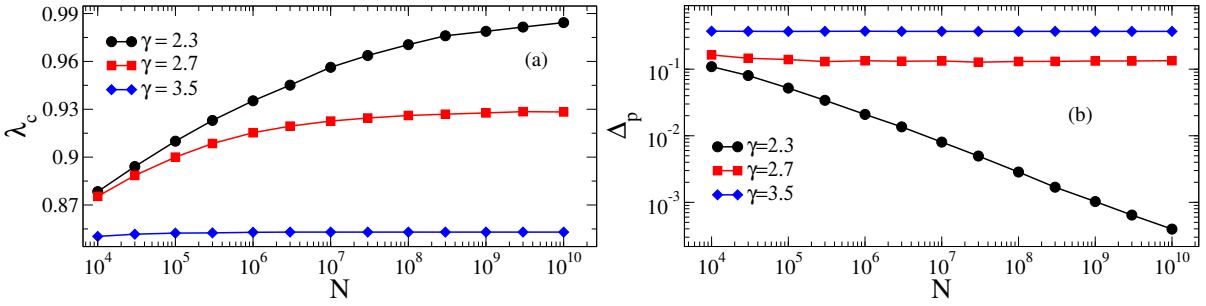


Figure 5.6: Finite-size scaling for (a) activation threshold and (b) discontinuity gap for 2SCP obtained with the HMF theory for power-law degree distributions with different values of γ and $\mu_s = 0.2$ fixed. Lower and upper degree cutoffs $k_{min} = 3$ and $k_{max} = \sqrt{N}$ we adopted.

shrinking of bistability region in the phase parameter λ versus μ_s as the heterogeneity is increased towards $\gamma \to 2$.

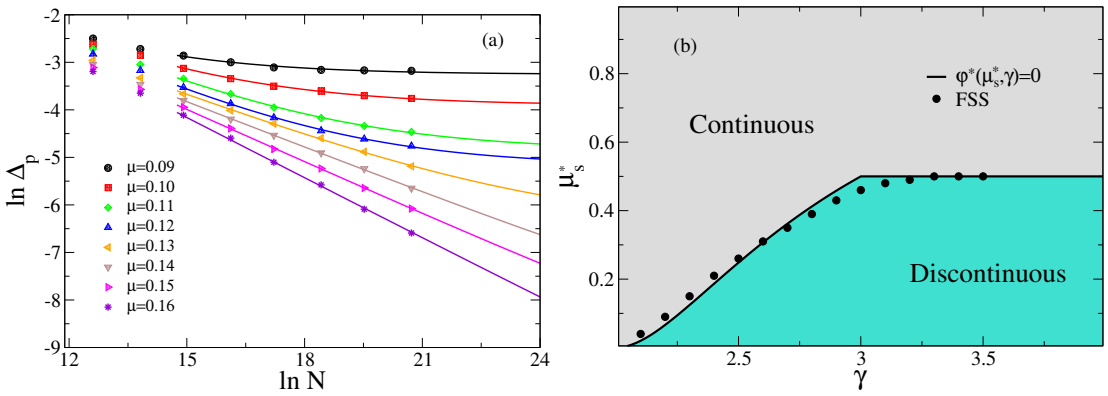


Figure 5.7: (a) Finite-size scaling (FSS) of discontinuity gap (Δ_p) for 2SCP obtained with the HMF theory for power-law degree distributions with $\gamma = 2.3$ and different values of μ_s. Symbols are numerical data obtained from integration of HMF equations and solid lines non-linear regression to perform the FSS; see main text. (b) Symbiosis coupling μ_s^* separating the discontinuous and continuous transitions for different values of the degree exponent, obtained theoretically (solid curve) and with the FSS of Δ_p versus N.

5.3 Quasistationary simulations on complex networks

In order to validate the predictions of HMF theory, we performed quasistationary simulations of 2SCP on annealed and quenched networks with power-law degree distribution (see Section 2.4 for definitions), using the UCM algorithm to construct the latter ones [94]. We performed stochastic simulations using an extension of the aforementioned optimized Gillespie algorithm [54, 114]. To deal with the intrinsic

difficulties of absorbing states in finite-size systems, we adopted an slightly modification of the SQS method (see Section 4.1.1 for details). In the standard quasistationary procedure, replacement of the system's state with a randomly selected configuration of a list occurs whenever the absorbing state is visited. However, in 2SCP exists a non-zero probability that one of the species will go extinct during the dynamic. Since we are interested in studying the symbiotic interactions, do not make sense to simulate states in which only one of the species is absent, despite the system being active. Thus, we will consider "absorbing" any state in which one of the species is extinct. In the present chapter we used $n_{\text{conf}} = 100$, $p_{\text{rep}} = 10^{-2}$, $t_{\text{rlx}} = 10^{6}$, and $t_{\text{av}} = 10^{7}$.

Figure 5.8-(a) show comparisons between total prevalence on HMF theory and QS simulations on annealed power law networks with degree exponent $\gamma = 2.3$ and different sizes. Excluding discrepancies due to computational precision, it can bee seen that the simulations converge to the theoretical predictions as N increases, as expected. Similar results were found for other values of γ. This agreement corroborates the correctness of HMF analysis. Simulations of 2SCP on UCM quenched networks with the same parameters of the annealed case are presented in Figure 5.8-(b). While a similar qualitative behavior is observed, with the reduction of discontinuity gap as N increases, there are expected quantitative differences, mainly an increase of the lower spinodal activation threshold [54]. The same effect can be observed for the single species CP [46, 141] due do dynamical correlations of the vertices.

By performing a finite-size scaling, we get a clearer comparison between these parameters on theoretical and simulation analysis. Figure 5.9 shows the FSS for both Δ_p and λ_c on power law networks with $\gamma = 2.3$ and 2.7. It can be seen a quick convergence of the activation threshold for the annealed simulations to the predicted by HMF theory while the UCM ones deviates from the latter. The discontinuity gap present the same behavior for all three approaches for both investigated exponent. Small differences between the annealed case and the HMF may be attributed to the smaller accuracy of simulations compared to the HMF integrations, for which a small deviation of the activation threshold lead to imprecision in the critical quantities.

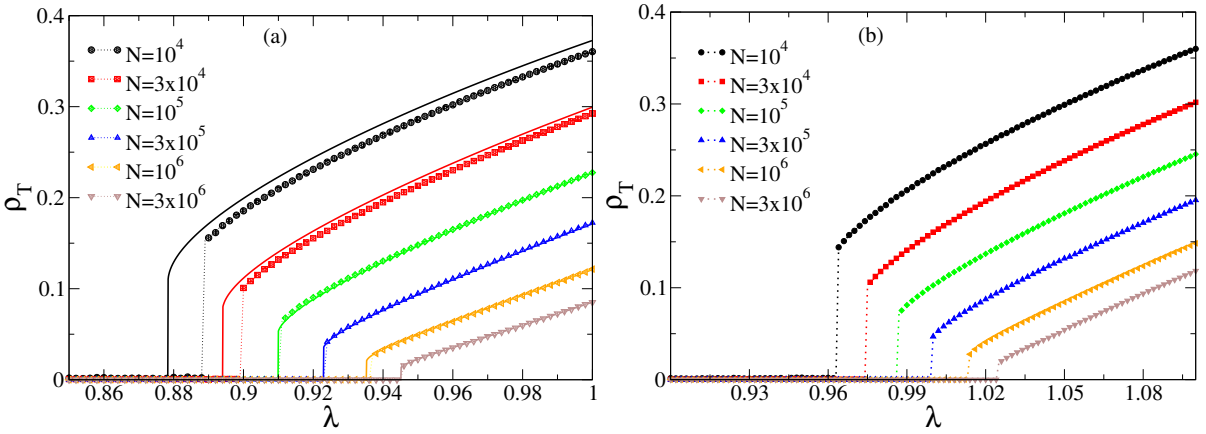


Figure 5.8: (a)Total prevalence curves obtained with QS simulations of 2SCP model on (a) annealed and (b) quenched networks with degree exponent $\gamma = 2.3$ and different sizes (symbols) in comparison with HMF equations integration of the same degree distribution $P(k)$ (solid lines). The lower and upper degree cutoff are $k_{\min} = 3$ and $k_{\max} = \sqrt{N}$ and the symbiosis parameter is $\mu_s = 0.2$. Dots lines are guides to the eyes to visualize the discontinuity in simulation curves.

5.4 Summary and discussions

Coexistence of dynamical processes on top of complex networks is a breakthrough issue that has been investigated in several applied contexts, which demand more complex models and substrates. Due to its simplicity, the two species contact process can be used as a benchmark model to investigate the nature of discontinuous or continuous transition in coexisting dynamics. The existence of continuous and discontinuous transitions in the 2SCP has been already theorized in simpler mean-field theories, such as the homogeneous one [52, 53], in which the discontinuity threshold is fixed and independent of the substrate. Even for more complicated theories, such as the pairwise mean-field theory [54], this threshold is constant. The influence of heterogeneous contact patterns has not been analyzed from theoretical point of view. In this chapter, we tackled this problem by investigating a HMF theory for the 2SCP.

The theory predicts that increased heterogeneity leads to the reduction of bistability regions in the λ versus μ_s phase diagram, where absorbing and active states are locally stable. In terms of network parameters, the bistability region shrinks as the degree exponent γ goes to the lower bound $\gamma = 2$. This weakening may be attributed to the lowered probability to produce doubly occupied nodes among the neighbors of hubs, which are the most active nodes of the network and are more fre-

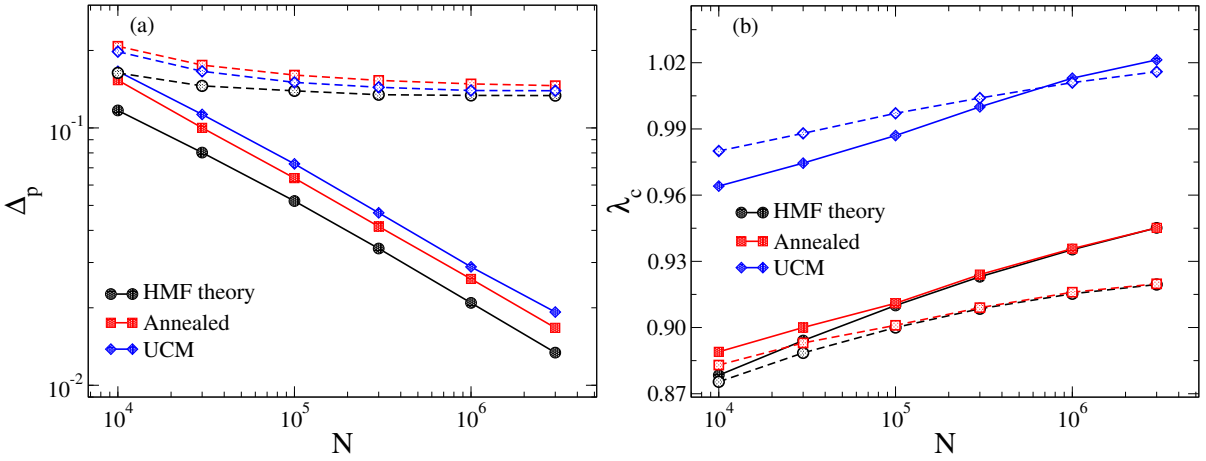


Figure 5.9: Comparison of the finite-size scaling of (a) gap discontinuity Δ_p and (b) activation threshold λ_c obtained in HMF theory and simulations on quenched and annealed networks with $\gamma = 2.3$ (solid lines) and 2.7 (dashed lines). The color sequence is black for HMF theory, red for annealed and blue for UCM networks. The lower and upper cutoff are $k_{\min} = 3$ and $k_{\max} = \sqrt{N}$ and the symbiosis parameter is $\mu_s = 0.2$.

quent the smaller the γ. We performed a finite-size analysis and observed complex and unusual behaviors. Depending on the strength of the symbiotic coupling, a discontinuous transition observed at a finite size gradually converges to a continuous one in the thermodynamic limit, consisting in pseudo discontinuous transition at finite sizes. Our analytical results are backed up by extensive stochastic simulations on both annealed and quenched networks.

We expect our results to stimulate further investigations of the interplay between heterogeneity and bistability in nonequilibrium absorbing-state phase transitions, helping to understand basic phenomena in applied modeling of coexisting dynamics such as interacting diseases [142, 143], and ecological symbiosis [144].

Part III

Spreading phenomena on geographical networks

Chapter 6

Outbreak diversity in epidemic waves through distinct geographical scales

Related publication:

Outbreak diversity in epidemic waves propagating through distinct geographical scales

Guilherme S. Costa, Wesley Cota and Silvio C. Ferreira

Physical Review Research 2 (4), 043306 (2020)

The rapid spreading of COVID-19 mobilized academics from different areas to develop models to understand and predict the behavior of this pathogen. Among the possible approaches, data-driven models, in which a mathematical model is fueled with real data, have proved to be a viable framework to study the epidemic spread of Sars-Cov-2 [29, 28]. In an attempt to contribute to these understanding, we incorporate the main features of COVID-19 pandemic to compartmental models: the asymptomatic phase, in which the individual may spread the disease but does not have symptoms and the lack of robust testing policies in some places. Then, we study the geographical variability and the COVID-19 impacts on continental scales, using the first COVID-19 epidemic wave in Brazil as a study case. In this chapter, we start by discussing aspects of Brazilian demography and concepts regarding its territorial division. Then, we detail the compartmental data-driven model, namely the SEAUCR model. Finally, we analyze the spatiotemporal spreading of the epidemic on different geographical levels.

6.1 Analysis on Brazilian territorial division and demographic

In this section, we discuss aspects of geographical division of Brazil and analyze some demographic data regarding Brazilian cities. Using these geographical divisions, we can study the spreading at different scales, see the influences of these regions on each other and compare the analysis on different spatial resolutions.

The most known geographic divisions of Brazil are the federative states (UF) and municipalities. Each region has a level of independence to adopt sanitary control and, during the COVID-19 pandemic, prophylactic measures to contain the Sars-Cov-2 spreading were adopted heterogeneously in those regions. Moreover, the demographic and developmental structure of Brazil is highly heterogeneous in terms of age pyramid, urban population and density, contact patterns, commuting, etc.

In addition to the aforementioned divisions, the *Instituto Brasileiro de Geografia e Estatística* (IBGE) regulated two more: **Intermediate Regions** and **Immediate Regions** [145]. The immediates regions are groups of municipalities that have large communication and mutual dependencies on each other, whether in search of services, work or goods. The intermediate ones are groups of immediate regions including a metropolis or prominent regional center within this set. To illustrate these regions, in Figure 6.1 we present these geographical divisions for the state of Minas Gerais. The intermediate region of Juiz de Fora, immediate region of Viçosa and municipality of Viçosa are highlighted in the figure.

Figure 6.1: Maps illustrating different geographical divisions of the Minas Gerais state. From left to right: Intermediate regions, immediate regions and municipalities. Highlighted patches represent the intermediate region of Juiz de Fora, immediate region of Viçosa and municipality of Viçosa, respectively.

Considering these geographical concepts, Brazil is divided in:

- 26 federative states and one federal district;
- 133 intermediate regions;
- 510 immediate regions;
- 5570 municipalities (5565 in the 2010 census);

In Table 6.1, we list all federative states of Brazil and their respective abbreviations, which will be used during this text.

Federative state	Abbreviation	Federative State	Abbreviation
Acre	AC	Paraíba	PB
Alagoas	AL	Paraná	PR
Amapá	AP	Pernambuco	PE
Amazonas	AM	Piauí	PI
Bahia	BA	Roraima	RR
Ceará	CE	Rondônia	RO
Distrito Federal	DF	Rio de Janeiro	RJ
Espírito Santo	ES	Rio Grande do Norte	RN
Goiás	GO	Rio Grande do Sul	RS
Maranhão	MA	Santa Catarina	SC
Mato Grosso	MT	São Paulo	SP
Mato Grosso do Sul	MS	Sergipe	SE
Minas Gerais	MG	Tocantins	TO
Pará	PA		

Table 6.1: Federative states of Brazil and their respective abbreviations.

Since the last IBGE census [62], five new municipalities were created in Brazil: Mojuí dos Campos/PA, Balneário Rincão/SC, Pescaria Brava/SC, Pinto Bandeira/RS e Paraíso das Águas/MS. As we did not obtain official data regarding these cities, we consider that confirmed COVID-19 cases in them counts for the municipalities they belonged to.

A variety of public data regarding Brazilian population can be obtained from the 2010 census database [62]. We focus on mobility and urban data, since these aspects are the most important for epidemic spreading [146, 147]. To start, Figure 6.2 presents the histograms for (a) fraction of urban population and (b) urban density. It can be observed that there is a huge dispersion on both histograms, with an approximately

uniform frequency between 50 and 100% urban population in (a) and mostly municipalities in between $1000 - 5000$ inhab/km^2 (some outliers reach 15000 inhab/km^2) in (b). These are footprints of the demographic heterogeneity of Brazil and the needs to take these cities' individualities into account.

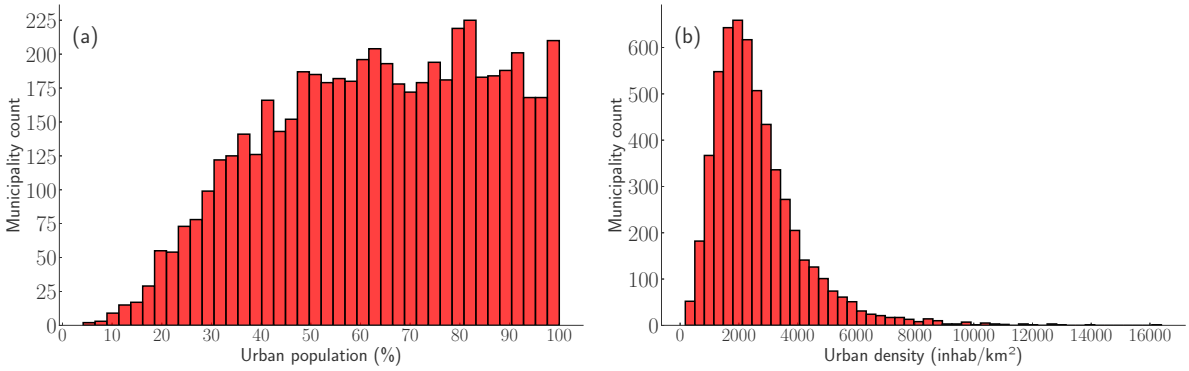


Figure 6.2: Histograms for (a) percentage of urban population and (b) urban density, in inhab/km^2.

Another important demographic aspect regarding epidemic spreading is human mobility between different places. Individuals traveling between cities might get infected and carry out the disease to other locations. To further understand this spreading pattern, we gathered data regarding commuting (individuals who live in a city and travel daily to other cities to work or study) and air traffic [64]. We constructed a mobility matrix W_{ij} representing the recurrent flow of people between cities i and j (in this context, W_{ii} represents the population of i that do not move to other locations). Figure 6.3 illustrates connections between different municipalities. It can be noticed that the states from Northern and West-Center regions present a mobility mesh less connected than the rest of Brazil. This occurs mainly because data lack information regarding river mobility (present in the Northern region) and because of municipality size. For example, the Amazonas state has 62 municipalities while Minas Gerais, with 1/3 of territorial extension of the former, has 853.

To make easier comparisons between cities of different sizes and populations, we can normalize the mobility rates $R_{ij} = \dfrac{W_{ij}}{N_i}$, in which N_i is the population of i. The sum over all municipalities j yields the population fraction that travels recurrently: $R^{(i)}_{\text{out}} = \sum_{j \neq i} R_{ij}$. Thus, R_{ii} is the population fraction that does not travel. We can calculate the number of individuals that arrive at a determined city j, also normalized by the population N_j: $R^{(j)}_{\text{in}} = \dfrac{1}{N_j} \sum_{j \neq i} W_{ij}$. We can analyze the histograms of $R^{(i)}_{\text{out}}$ and

Figure 6.3: Geographical representation of the mobility matrix between different municipalities. If municipalities i and j are connected by a line, then $W_{ij} \neq 0$.

$R_{\text{in}}^{(j)}$ for all municipalities in Brazil. These histograms are shown in Figures 6.4-(a) and (b) respectively. Most of the municipalities have between 0 and 10% of the population commuting daily, presenting outliers up to 30%. Since this mobility is substantial, it may contribute to speed up the spreading of diseases between different regions of Brazil. The histogram for $R_{\text{in}}^{(j)}$ has a similar behavior except for a few outliers, that represent mainly mining parks located in small cities due to natural resources proximity.

6.2 Comparmental model

The construction of the model can be separated in two parts: the hypothesis and assumptions regarding the substrate, i.e., the environment in which the pathogen spreads and the modeling of contagion itself. The former part refers to mobility,

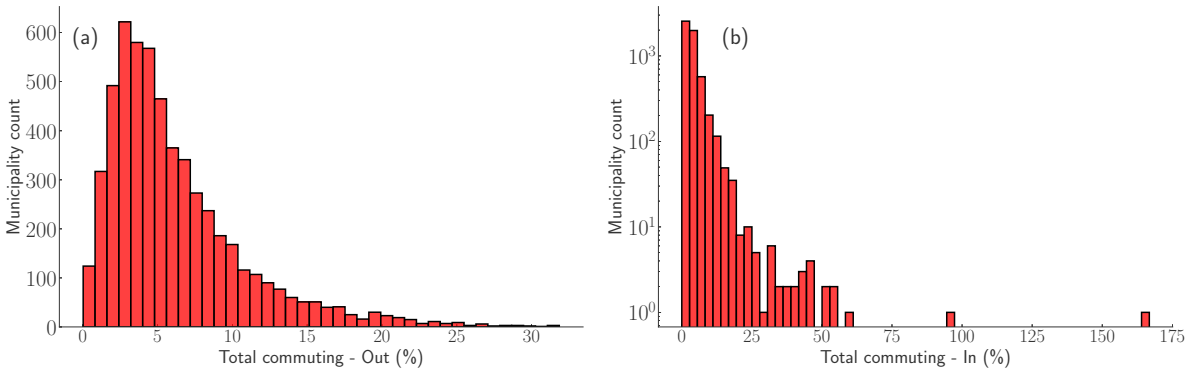


Figure 6.4: Histograms for the (a) population that travels reccurently $R_{\text{out}}^{(i)}$ and (b) fraction of individuals that arrive at a determined city $R_{\text{in}}^{(j)}$. Panel (b) has a semi-log axis to evidence the outliers.

demographics and contact patterns while the later comprehends the construction of compartments and transitions between them.

Starting from the substrate, one of the common hypotheses on epidemic modeling is the homogeneous mixing [148] which considers that all individuals within a given population (a country, a state, a city, etc.) may interact with each other. In a country with continental scales as Brazil, this approach is surely inaccurate when considering that individuals separated apart by more than 4000 km can interact with equal chance, independently of the localization. One way to mitigate this problem is to consider a metapopulation approach, i.e., the division of a bigger population into small groups, increasing the spatial resolution. Within these patches, the individuals may interact or travel to other patches. Figure 6.5 shows a comparison between the homogeneous mixing and the metapopulation of the same population.

The metapopulation framework was initially developed for applications in ecological problems [149, 150], in which an habitat of certain species is fragmented into smaller parts and the interactions between these different pieces are analyzed and characterized. Likewise, a country population is fragmented into smaller portions (municipalities, states, etc.), making it possible to study the interactions among these different geographical scales.

Several studies used this method to improve the homogeneous mixing hypothesis [151, 152, 61, 153]. However, this approach just softens the problem since the mixing hypothesis is still considered within each patch. Therefore, it is more accurate to assume that any two individuals in a patch may get in physical contact than in the

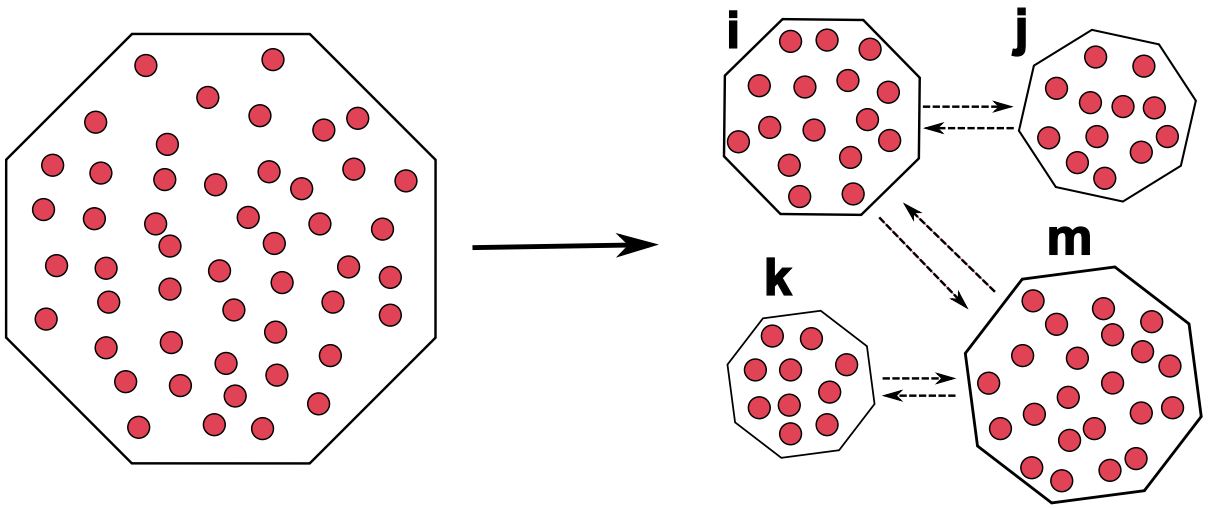


Figure 6.5: Schematic comparison between the homogeneous mixing (left) and the metapopulation (right) for the same population. In each patch *i*,*j*, *k* and *m*, the individuals may interact within the patches or travel to others patches (the movement directions are represented by the dashed arrows).

whole population of a country. Another important advantage of the metapopulation method is the possibility to take the individual mobility between patches into account. Thus, we can study the influence of this mobility on the epidemics and visualize the spatial patterns of spreading. In this work, we group Brazilian population at municipality level, i.e., we consider 5565 patches representing each a municipality of Brazil.

The patches have a population N_i, $i = 1, 2, 3, ...\Omega$, extracted from official data [63] and a number of contacts k_i per day. Information regarding urban density and urban population will be used to modulate k_i according to each municipality. The interactions between urban and rural populations will be implemented implicitly through a contact matrix C_{XY} that quantifies the intensity of interactions between groups X and Y. In our work, we assumed that $C_{uu} = C_{rr} = 1$ and $C_{ur} = C_{ru} = 0.5$ (u = urban, r = rural) to point that groups interact more often within their similar.

By considering ω_i and $(1 - \omega_i)$ the fractions of urban and rural populations, respectively, k_i can be written proportionally to these fractions as

$$k_i \propto g_i = \omega_i^2 C_{uu} + (1 - \omega_i)^2 C_{rr} + 2\omega_i(1 - \omega_i)C_{ur}, \tag{6.1}$$

in which the terms represent the urban-urban, rural-rural and urban-rural interactions, from left to right.

Another important factor to take into account is the urban density of a city ξ_i, recently studied in Reference [147] by using monotonic relations between these quantities ($k_i \propto f(\xi_i)$). In this work, we adopt the same function as in Reference [29]:

$$f(\xi_i) = 2 - \exp(-\xi_i/a), \tag{6.2}$$

in which a is the average urban density from Brazil. Plugging these two contributions, we write $k_i = z f(\xi_i) g_i \langle k \rangle$, in which $\langle k \rangle$ is the average number of contacts and z is the normalization factor, imposed by $\sum_i k_i N_i = \langle k \rangle N_{\text{tot}}$, given by

$$z = \frac{\langle k \rangle N_{\text{tot}}}{\sum_{i=1}^{\Omega} f(\xi_i) g_i N_i}, \tag{6.3}$$

in which N_{tot} is the total population of Brazil. To summarize, our substrate is a network of cities, each one with population N_i given by official data and k_i, modulated by the expressions above. Individuals within patches may interact or travel to other patches with rate R_{ij}.

In the compartmental modeling of an epidemic, we take into account key properties of the pathogen spreading. Among the main features of COVID-19, the transmissibility by **asymptomatic** (who do not develop symptoms) and **presymptomatic** (who have not yet developed symptoms) individuals [25, 65], the protective immunity and, due to testing policies, the undocumented infections [154] are the most prominent ones. To integrate these features into a model, we built a six-compartment model, named SEAUCR model and illustrated in Figure 6.6. They are: Susceptible (S), Exposed (E), Assymptomatic or Pressymptomatic (A), Unconfirmed Symptomatic (U), Confirmed Symptomatic (C) and Removed (R). The transitions and their characteristics are described below:

- Susceptible (S): individual who has not come into contact with the pathogen. By contacting a infectious individual (A, U and C), it become E with rates $\lambda_{\text{A}} k_i$, $\lambda_{\text{U}} k_i$ and $\lambda_{\text{C}} k_i$, respectively.

- Exposed (E): individual with the pathogen but who cannot spread it yet. With rate μ_{A}, it becomes A spontaneously.

- Assymptomatic or Pressymptomatic (A): individual that possesses the pathogen,

spreads it but does not present symptoms, being difficult to track and identify. With rates β_U, β_R and β_C it spontaneously becomes U, R and C, respectively.

- Unconfirmed Symptomatic (U): individual that is infected, spreads the pathogen, present the symptoms but was not diagnosed for COVID-19. With rates α_C and α_R, U spontaneously becomes C or R, respectively.

- Confirmed Symptomatic (C): individuals which were diagnosed for COVID-19. The creation of this compartment reflects the limited testing capacity, especially at the beginning of the pandemics, being impossible to detect all infectious individuals. With rate α_R, it becomes R spontaneously. It has a reduced number of contacts (-30%) to reflect voluntary physical distancing or hospitalization.

- Removed (R): individual that already had contact with the virus and cannot be reinfected.

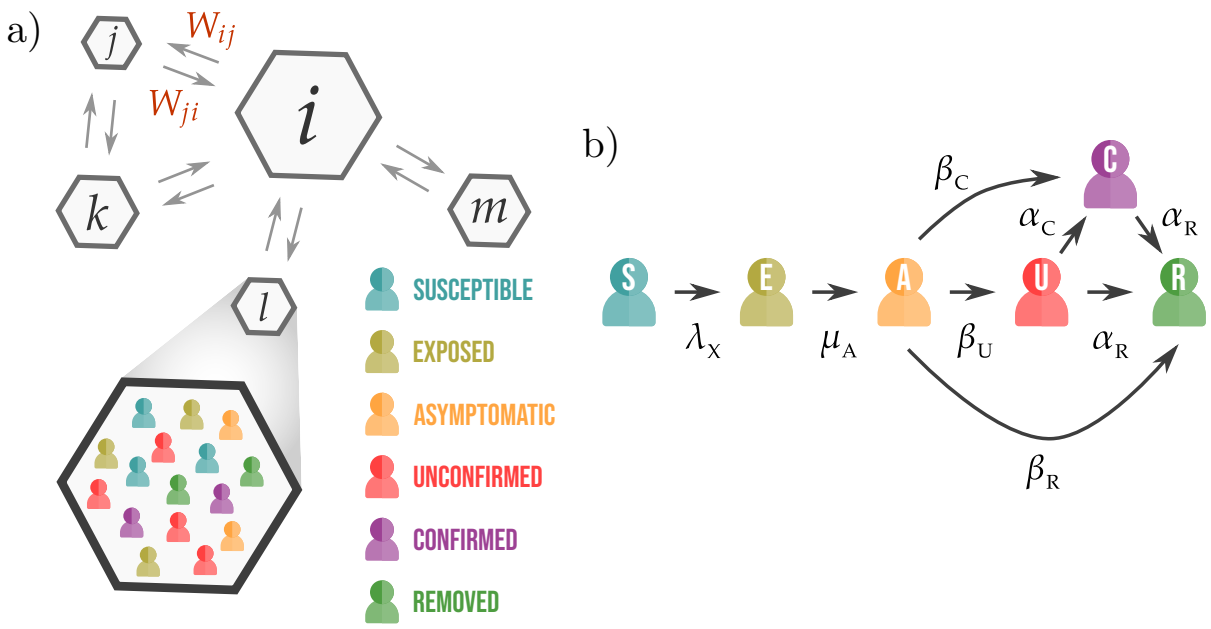


Figure 6.6: Schematic representation of (a) metapopulation structure (b) transitions between different compartiments with respective rates. The epidemic events occur within patches under the homogeneous mixing hypothesis, as illustrated in the zoom of patch l. Transitions from state S to state E occurs by contact between S with A, U or C individuals with rates λ_A, λ_U and λ_C, respectively. All other transitions are spontaneous with indicated rates.

An important question regarding the transition rates refers to the testing rates α_C. The number of confirmed cases have a limiting factor regarding the municipalities

testing capacities, i.e., the cities may supply the demand for testing during a more aggressive spreading with distinct levels. We consider the testing rate as dependent on the number of confirmed cases in the municipalities, assuming two hypotheses: all cities have resources to test and identify few cases and the maximum capacity depends on the population of the city. Thus, we introduced a global parameter ζ as the **testing capacity per inhabitant** and assumed a monotonic function that includes both discussed hypothesis:

$$\alpha_{\mathrm{C}}^{(i)} = \alpha_{\mathrm{C}}^{(0)} \left[1 + \frac{\alpha_{\mathrm{C}}^{(0)} \tilde{n}_i^{\mathrm{U}}}{\zeta N_i} \right]^{-1}, \tag{6.4}$$

in which $\alpha_{\mathrm{C}}^{(0)}$ is test search rate and $\tilde{n}_i^{\mathrm{U}}$ is the number of individuals in compartment U at municipality i. The notation $\tilde{n}_i^{\mathrm{X}}$ represents the number of individuals in compartment X and municipality i from now on.

Another important issue refers to the initial conditions of the model since surveillance counts contains information mostly on confirmed cases (which correspond to C compartment). Thus, we need to estimate initial quantities for the exposed, asymptomatic and preassymptomatic, unconfirmed infected and removed cases. In Chapter 7, we present a method to estimate these compartments using case and death time series. However, at time that the content of the current chapter was conducted, the method was not developed yet and some epidemiological parameters, essential to the method, were not known. Moreover, the focus here is in the geographical spreading across different places and not the precise estimate of the epidemic outbreak. So, we choose three dates $t_{-1} < t_0 < t_1$ and $\tilde{n}_i^{\mathrm{conf}}(t)$ as the number of confirmed cases at day t and municipality i. Considering t_0 as the initial day for the simulations, we assumed that all confirmed cases until t_{-1} have already become removed. So, the number of active confirmed cases in the simulations is $\tilde{n}_i^{\mathrm{conf}}(t_0) - \tilde{n}_i^{\mathrm{conf}}(t_{-1})$.

For other compartments we consider that the number of active confirmed cases from a future date t_1 will depend linearly on E, A and U up to $t_1 - t_0$ previous days. Below we present the used relations between the confirmed cases in the three dates and the estimates for other compartments. The multiplicative factors were calibrated to best fit the curve from t_0 to t_1 considering that $t_0 = 31$ March 2020, $t_{-1} = 24$ March

2020 and $t_1 =$ 04 April 2020.

$$\begin{aligned}
\tilde{n}_i^{\mathrm{C}}(t_0) &= \tilde{n}_i^{\mathrm{conf}}(t_0) - \tilde{n}_i^{\mathrm{conf}}(t_{-1}) \\
\tilde{n}_i^{\mathrm{U}}(t_0) &= \tilde{n}_i^{\mathrm{conf}}(t_1) - \tilde{n}_i^{\mathrm{conf}}(t_0) \\
\tilde{n}_i^{\mathrm{A}}(t_0) &= 2\left[\tilde{n}_i^{\mathrm{conf}}(t_1) - \tilde{n}_i^{\mathrm{conf}}(t_0)\right] \\
\tilde{n}_i^{\mathrm{E}}(t_0) &= 8\left[\tilde{n}_i^{\mathrm{conf}}(t_1) - \tilde{n}_i^{\mathrm{conf}}(t_0)\right] \\
\tilde{n}_i^{\mathrm{R}}(t_0) &= 12\tilde{n}_i^{\mathrm{conf}}(t_{-1}),
\end{aligned} \tag{6.5}$$

We used epidemiological parameters based in Reference [29] which were mined from analyses of early COVID-19 epidemic spreading on Hubei province, China [25, 65, 28]. The incubation time was taken as $\mu_{\mathrm{A}}^{-1} + \beta_{\mathrm{U}}^{-1} = 5.2$ d [65]. We used $\mu_{\mathrm{A}}^{-1} = \beta_{\mathrm{U}}^{-1} = 2.6$ d. The time for a symptomatic individual to be removed was estimated as $\alpha_R^{-1} = 3.2$ d [25, 28]. For the truly asymptomatic individuals, we used the same recovering time of the presymptomatic ones, such that $\beta_{\mathrm{R}}^{-1} = \beta_{\mathrm{U}}^{-1} + \alpha_{\mathrm{R}}^{-1}$. The infection rates $\lambda_{\mathrm{C}} = \lambda_{\mathrm{U}} = \lambda_{\mathrm{A}} = 0.06$ d^{-1} [29, 26, 155] and average number of contacts $\langle k \rangle = 13$ were estimated to reproduce the scaling of the very early exponential growth in Brazil, $\sim \exp(0.29t)$, in reported cases of COVID-19, as can be seen in Figure 6.7.

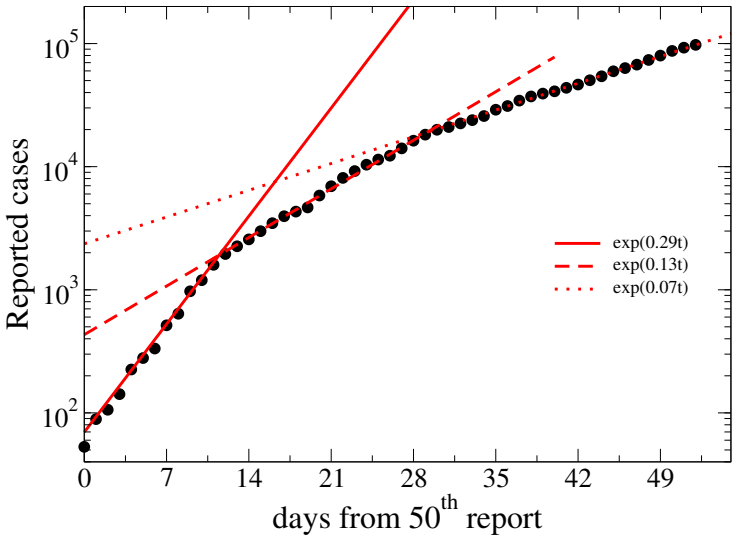


Figure 6.7: Number of cases of COVID-19 reported in Brazil (Data source: [23]). The day zero correspond to the 50th confirmed case.

The contact of the individuals who were confirmed for COVID-19 was depleted to a fraction $b = 0.3$ of their regular contacts. The confirmation rate $1/\alpha_{\mathrm{C}}^{(0)} = 10$ d was calibrated together with the initial conditions to fit the amplitude of confirmed case

curves after 31 Mar 2020 and also the total number of municipalities with confirmed cases in a **weak mitigation scenario**, to be defined in the sequence of this chapter. The testing capacity per inhabitant was estimated as approximately $\zeta = 1/40\ 000$ d^{-1}, from testing data for Brazil at that time [23].Due to the testing targeted only to severe cases of respiratory syndromes adopted in Brazil to the date of investigation, the transition from asymptomatic to confirmed was neglected setting $\beta_C = 0$.

By considering a voluntary quarantine, we assume that U and C individuals will return to their residence city when they are traveling and stay there until becoming R. Since the Sars-Cov-2 was a new pathogen, several questions related to spreading mechanisms, protecting immunity and, more important, the mutations and variants, were not fully understood during the development of this work. The model can be easily modified to include any new aspect of the COVID-19 or changes in the healthy policies during the course of the pandemic or to describe other pathogens.

Defining the simulation scenarios

We defined the different spreading scenarios by setting two parameters M and K that change the mobility between municipalities and the number of daily contacts, respectively. For a given pair (M,K), we have

$$\begin{aligned} k_i &\rightarrow (1-K)k_i \\ W_{ij} &\rightarrow (1-M)W_{ij}. \end{aligned}$$

By using this parameters, we defined three different scenarios: **No mitigation** ($M = 0; K = 0$), scenario in which nothing is done to contain the spreading; **weak mitigation** ($M = 0.4; K = 0.3$), scenario used to perform the calibrations and closest to the Brazil during the considered time window, according to Google Mobility Report [146], and **Moderate mitigation** ($M = 0.8; K = 0.5$).

With the scenarios defined, we use the number of reported cases and cities with at least one documented case to calibrate the empirical parameters of the model considering a weak mitigation scenario during a time window of 21 days from the starting date, 31 March 2020. In Figure 6.8, we show these calibrations. It can be observed that the model fits the data reasonably well during this 21 time window but deviations start to occur after this interval. This happens mainly because of the spreading dy-

namic features, public and prophylactic measures are changing constantly while the model has temporally fixed parameters. In panel 6.8-(b), we set $K = 0.3$ and vary M to check the effects of mobility in the spatial spreading of the pandemic, measured by the number of cities with confirmed cases. It is observed a clear dependency between M and the spreading speed of the pandemic on different municipalities. Note that there are not significant changes in the total number of cases, since they are led by the community transmission on the places with higher epidemic incidences.

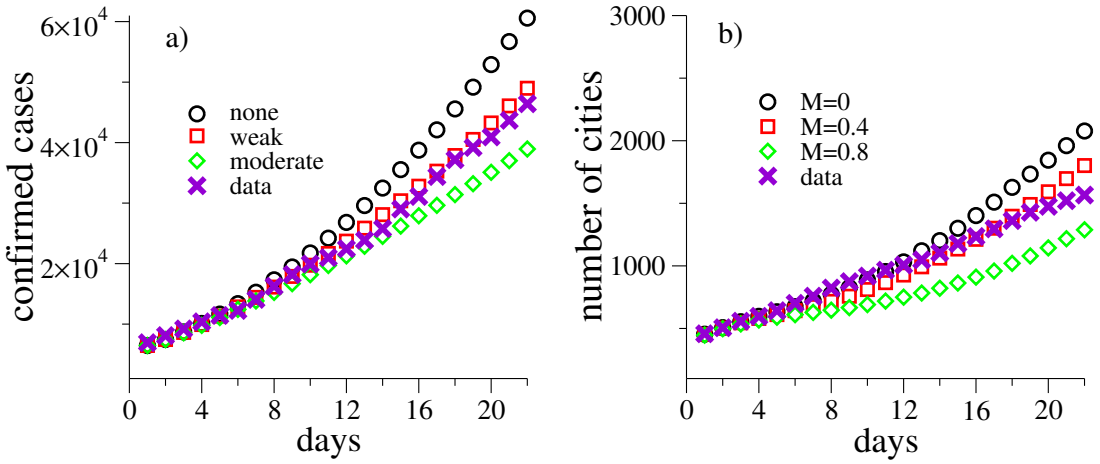


Figure 6.8: Comparison of simulations with reported data for COVID-19 in Brazil for the first three weeks with day 0 corresponding to 31 Mar 2020. (a) The number of confirmed cases is presented for the different mitigation scenarios. (b) Evolution of the number of municipalities with confirmed cases for $K = 0.3$ and different levels of mobility restriction.

A comparative overview can be drawn by plotting the number of municipalities with confirmed cases per federative states. These graphs are shown in Figure 6.9. It can be observed how each state fit into the simulated scenarios. For example, the Amazonas, Roraima, Pernambuco and Piauí are well described by the **no mitigation** scenario while the curves of Maranhão and Ceará fit with the **weak mitigation** scenario.

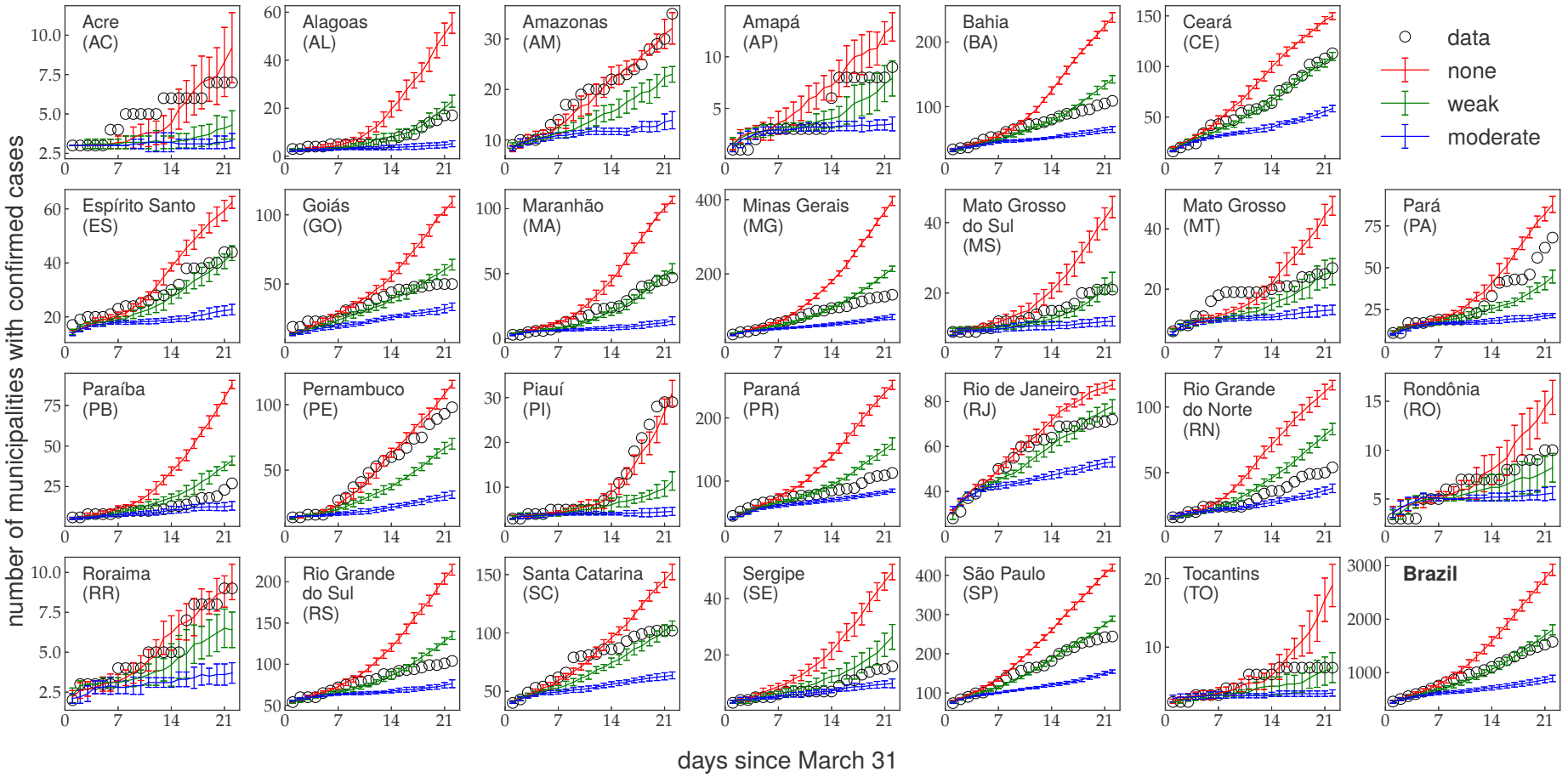


Figure 6.9: Simulations and observations of the number of municipalities with confirmed cases of COVID-19 for the 26 federative states of Brazil in the first three weeks of simulations. Data for the whole country is shown in the bottom right corner.

6.3 Epidemic spreading on multiple scales

In order to study the spatial spreading of the pandemic, we analyzed the three mitigation scenarios on different geographical levels, as discussed in Section 6.1. In Figure 6.10, we present an initial analysis considering the federative states and intermediate regions for the weak mitigation scenario. In each panel, we have the epidemic prevalence ($\rho_U + \rho_C$) for the UFs (thick lines) and the intermediate regions belonging to it (shaded curves). It can be seen that there is a certain variability for the prevalence in the intermediate regions within each state. In BA, CE and SP panels for example, we clearly see a dispersion of the incidence peaks for each region. Prevalence curves for other mitigation scenarios can be found in Figures B.1 and B.2 of Appendix B.2.

To investigate this variability in detail, we constructed Figure 6.11, in which we analyzed the epidemic prevalence for multiple geographic scales, also considering the weak mitigation scenario. In Figures 6.11-(a) and (b), we have the curves for all federative states of Brazil in comparison with the curve for the whole country. It can be noticed that the curves have different behaviors and, in general, the curve for Brazil is not representative of the whole situation. In some states, such as Amazonas and Ceará, the peak occurs earlier while in Tocantins and Piauí, they are delayed.

In Figure 6.11-(c), we increased the spatial resolution of the analysis using Minas Gerais as an example. We compared the prevalence curves for all intermediate regions of MG in comparison with Brazil's and Minas Gerais itself. Again, it can be seen the curve variation, with early peaks such as Belo Horizonte or Uberlândia and regions with later peaks or plateaus, such as Teófilo Otoni. By increasing the resolution even more, we selected regions with later and earlier peaks and plotted the curves of immediate regions within them. In panels 6.11(d) and (e), we observed this same variability occurring in the immediate regions of Belo Horizonte and Teófilo Otoni, with visible dispersion on the prevalence curves. The same effect is observed if we further increase the resolution at municipality level, as observed in panels (f) to (i) of the Figure 6.11. Although we just used the Minas Gerais state in the weak mitigation scenario as object of study, this effect is not exclusive of these parameters, see Figures B.3 to B.6 of Appendix B.2 for similar analyzes of Ceará, Rio de Janeiro, Rio Grande do Sul and São Paulo states.

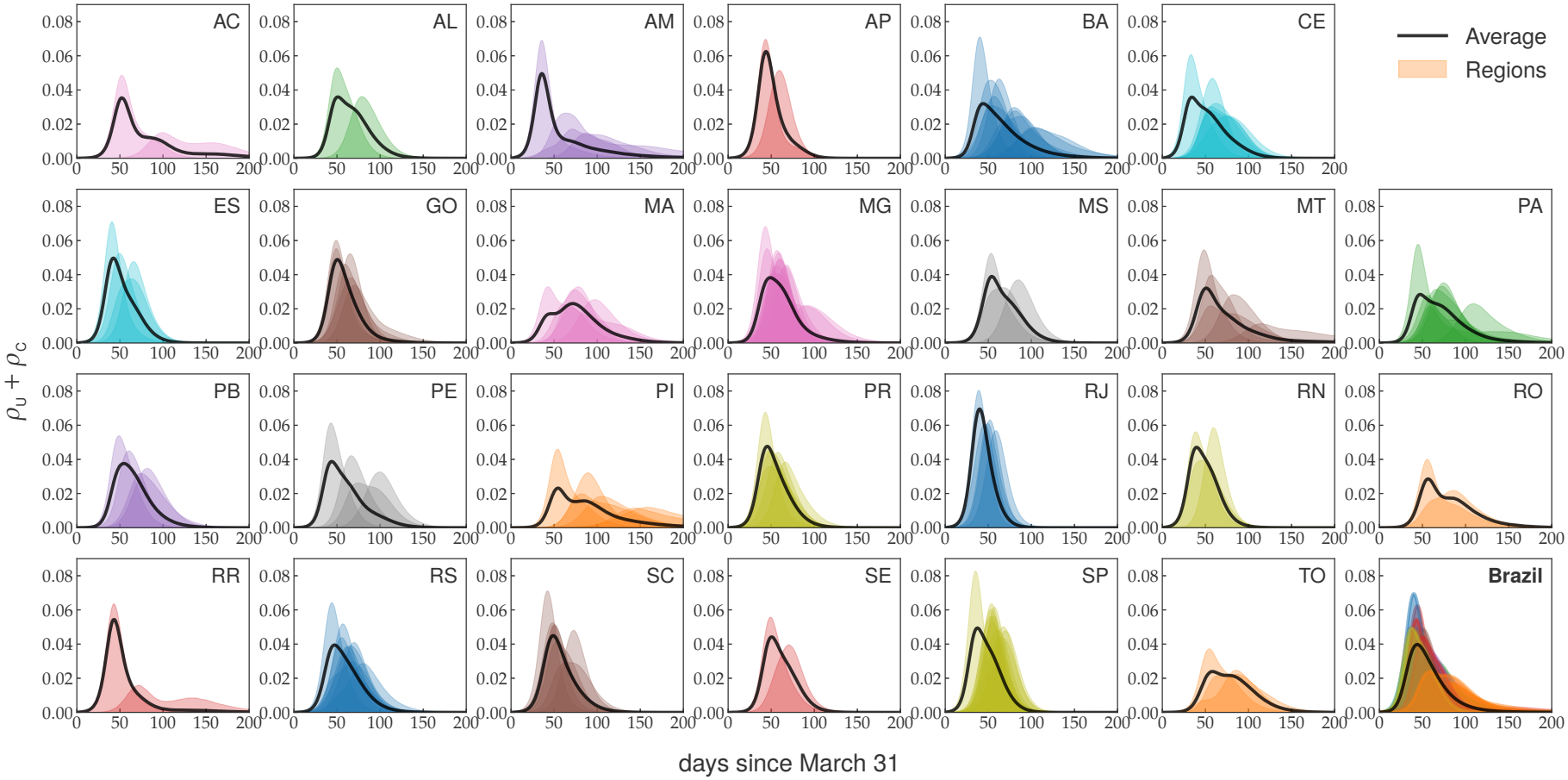


Figure 6.10: Epidemic prevalence of symptomatic individuals for all intermediate regions of Brazil grouped by federative state obtained with a weak mitigation scenario $(M, K) = (0.4; 0.3)$. The shaded plots represent the intermediate regions while the thick lines averages within each state. The bottom right corner plot represents an average over the whole country and shaded plots represent federative states.

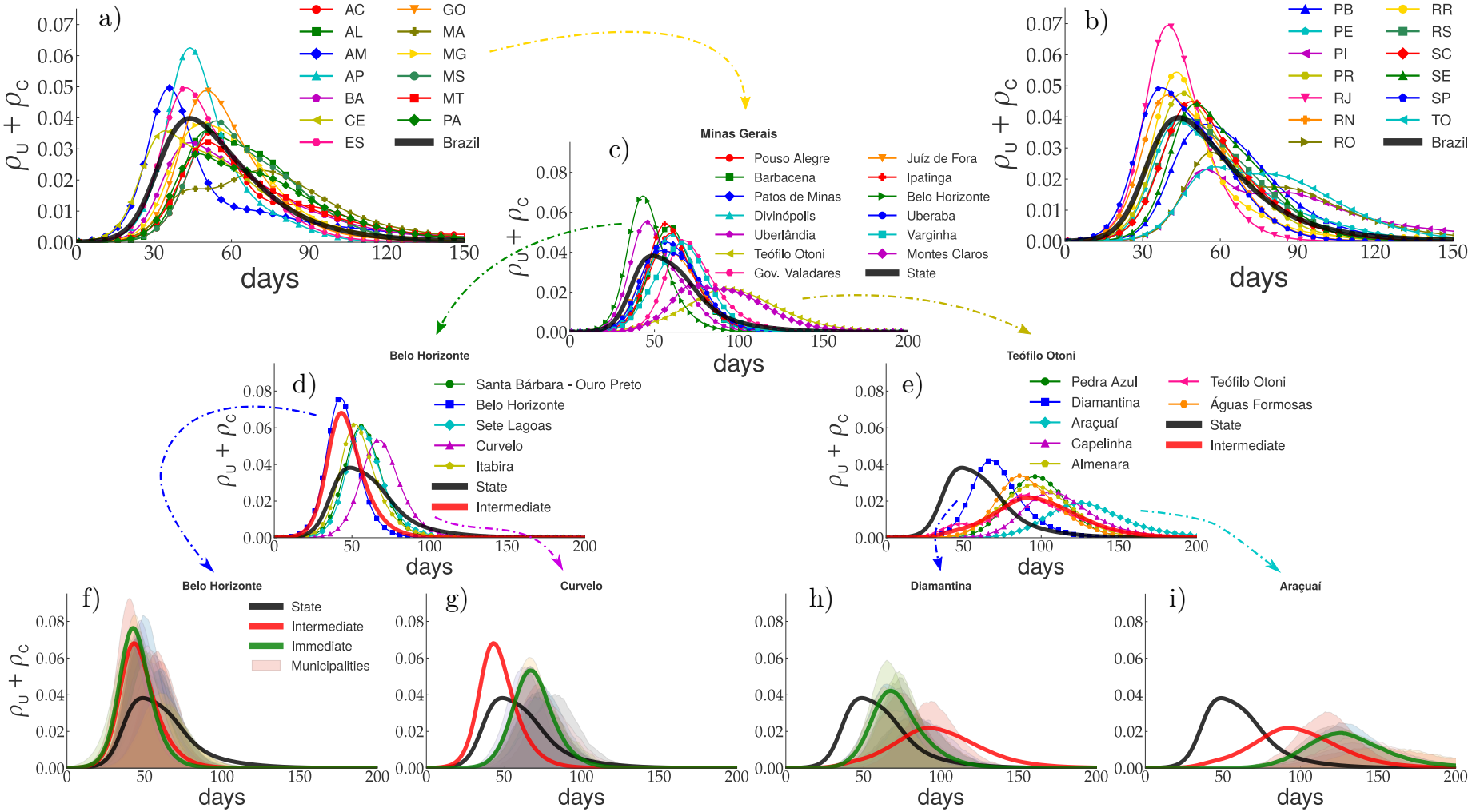


Figure 6.11: Multiscale analysis of epidemic prevalence at several scales of geographical organization, considering a weak mitigation scenario. Epidemic curves averaged with geographical resolution increasing from top to bottom are compared with lower resolution averages. (a),(b) federative states ; (c) intermediate regions ; (d),(e) immediate regions ; (f)-(i) municipalities. The curves presenting the earliest and latest maxima are chosen as representative within each panel (c)-(e). Arrows indicate curves selected for zooming. Day 0 corresponds to 31 Mar 2020. The state of MG was chosen for the higher resolution curves.

In Figure 6.12, we extracted some metrics from the prevalence curves of immediate regions, grouped by federative state. Time for occurrence of the epidemic peak (T - in days), width or duration of the outbreak (τ - in days) and maximum epidemic prevalence (ρ - population fraction). It can be seen a general behavior for all metrics, a dispersion depending on the UF and the presence of outliers (dots outside the whiskers). It can also be noted a bigger epidemic desynchronization in states with sparser municipalities or more disconnected mobility mesh, such as Amazonas and Mato Grosso, see Figure 6.3. Even for smaller states such as São Paulo, the epidemic in some immediate regions can last twice as long as the peak at the capital. Box plots for other mitigation strategies can be found in Figures B.7 and B.8 of Appendix B.2.

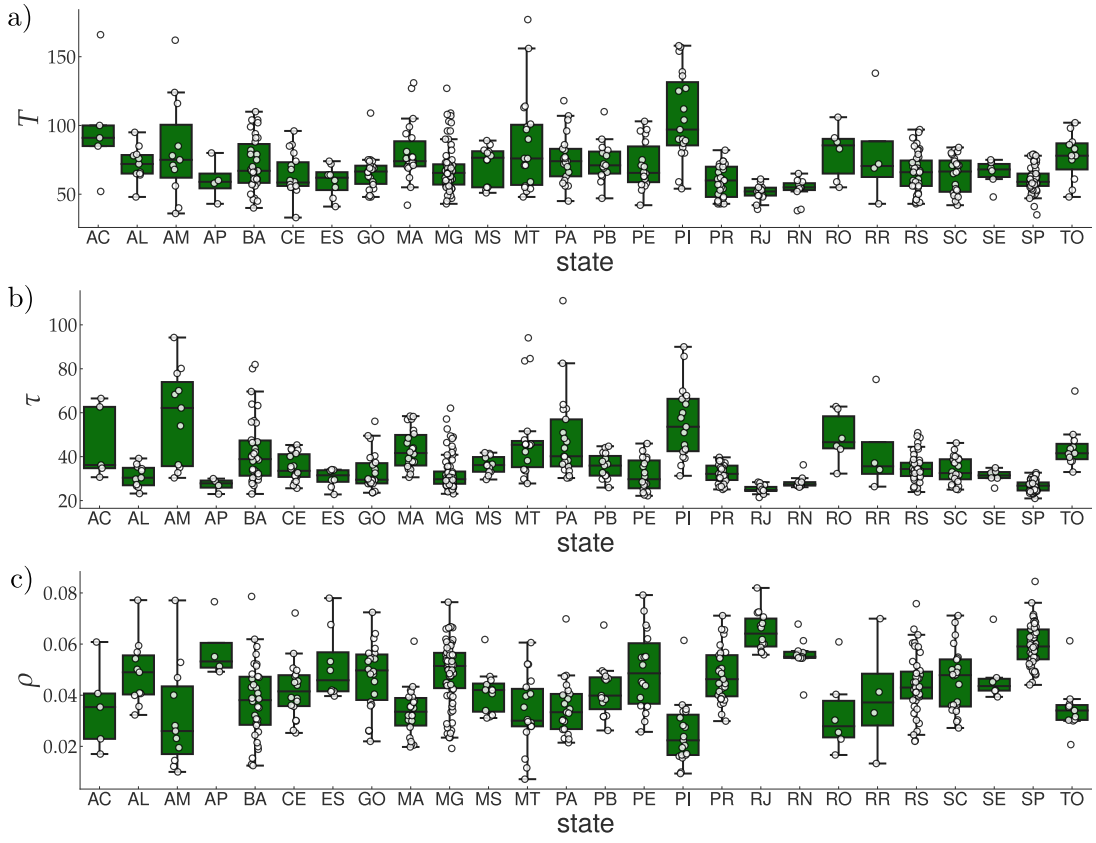


Figure 6.12: Box plots for the peak (a), time T, (b) width τ and (c) prevalence ρ averaged over immediate regions for the 26 federative states using a weak mitigation scenario. Circles are data for different regions.

6.3.1 Epidemic invasion on the countryside

From this initial analysis regarding the variability of epidemic peaks, we could see that, in fact, the pandemic presents a complex spreading pattern toward the country-

side. In order to further analyze this spreading, we constructed some heat maps to show the spatio-temporal evolution of the pandemic, as seen in Figure 6.13. We have maps for the epidemic prevalence in each municipality of Brazil, spaced at three-week intervals and considering the weak mitigation scenario. It can be seen an epidemic wave spreading mainly from the coast to the interior of Brazil as time advances. In many cases, the epidemic starts in big urban centers, such as the city of São Paulo or Manaus, and slowly spreads out to their surroundings. We can see the aforementioned desynchronization of the epidemic. The capitals show a decreasing prevalence while the interior still suffers from the outbreak. Other mitigation scenarios can be found in Figures B.9 and B.10 of Appendix B.2. In addition, links to videos illustrating this spatio-temporal propagation in a precise way can be found in Appendix B.2.

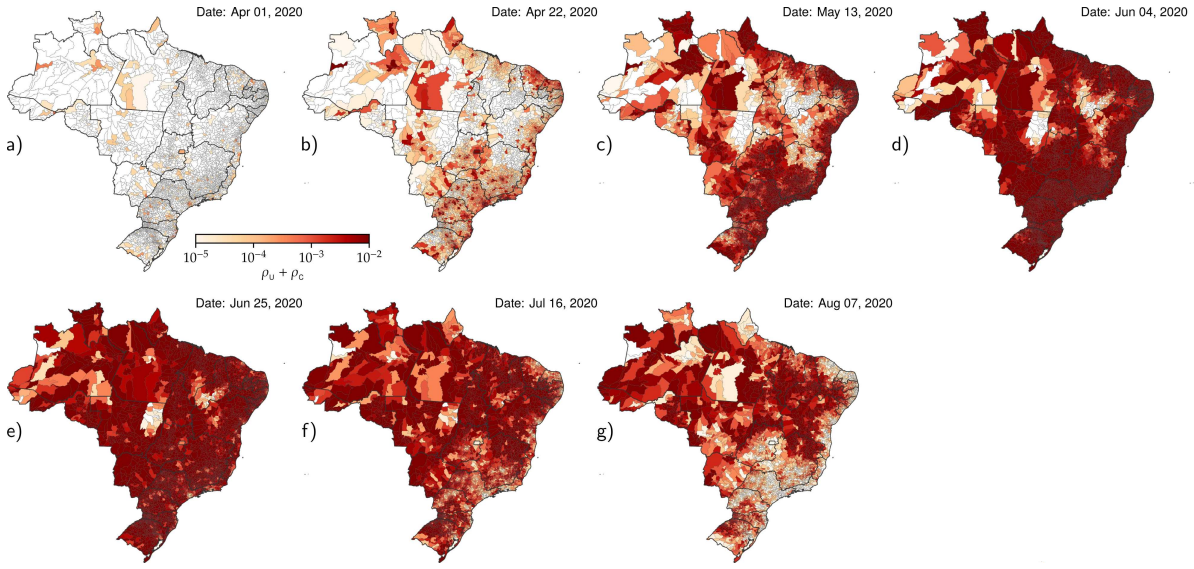


Figure 6.13: a-g) Color maps presenting the evolution of the prevalence of symptomatic cases (U and C) for Brazil in a weak mitigation scenario. Dates of the simulations are shown in the upper right corner of each frame. The darker colors represent higher prevalences in a logarithm scale.

Still in an attempt to characterize the epidemic spreading to the interior of Brazil, in Figure 6.14 we present an analysis of the time T of the epidemic peak for the prevalence curves of immediate regions as a function of their distances D from the capital city of the respective state. We classified the statistical correlation between distance and time of the peak in three groups: **strong correlation** (Pearson coefficient > 0.6 and p-value < 0.02), **moderate correlation** (Pearson coefficient < 0.6 and p-value < 0.02) and **not significant** (p-value > 0.02). Figures 6.14-(a) to(c) illustrates these correlations

for three states: São Paulo (strong), Minas Gerais (moderate) and Paraná (not significant). São Paulo presents a densely connected mobility network and clear correlations between T and D. MG and PR presents moderate or none correlations due to multiple epidemic outbreaks evolving simultaneously, related to different influence areas within the state. In Minas Gerais, for example, the Triângulo Mineiro, Zona da Mata and Norte Mineiro regions are influenced by the neighbor states of São Paulo, Rio de Janeiro and Bahia, respectively. In Figure 6.14-(d), we categorized all federative units of Brazil in these correlation categories. States in the north region (RR, RO, AP, AC, AM, PA and TO) have few immediate regions distributed in large territories and are less connected, see Figure 6.3. Thus, a lack of correlations is not surprising for these states. Moderate and lack of correlation observed in other states such as MG, BA, GO and PR, with larger and better connected territory, are due to existence of multiple important regions that play the role of regional capital cities. The regressions, Pearson coefficients and p-values for all states can be found in Figure B.11 of Appendix B.2.

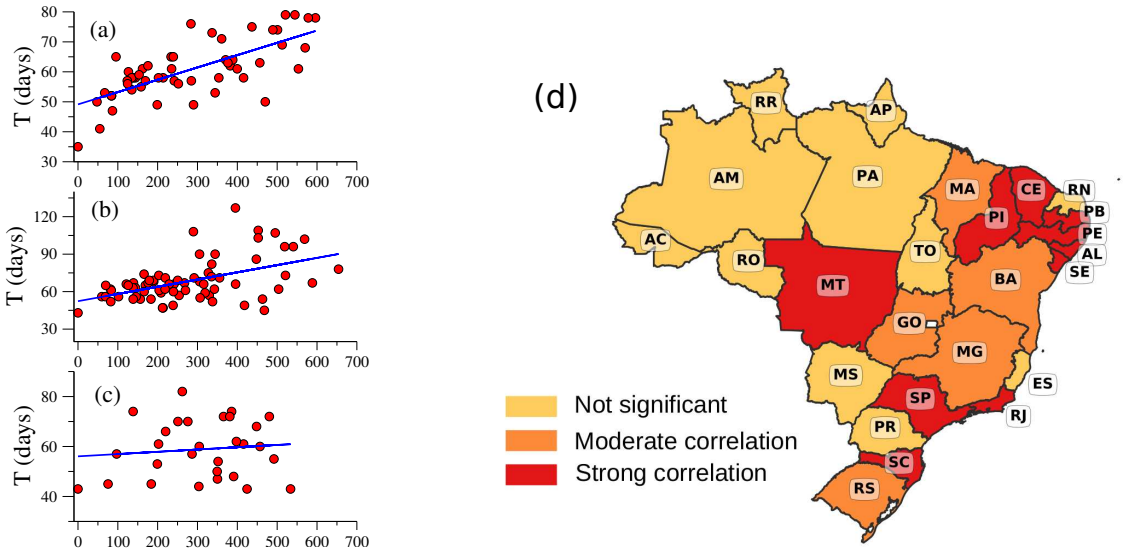


Figure 6.14: Scatter plots of peak time T vs distance from the capital D for the (a) SP, (b) MG and (c) PR states. (d) Map with the classification of typical correlation between peak time T and distance of immediate region to capital city D of each federative state.

6.3.2 Variability on real epidemic data

There was an interval of five months between conceiving the preprint of the paper related to this chapter and the published version [104]. Thus, due to this delay and with the advance of the epidemic in the interior of Brazil, several places attained

the peak of epidemics allowing to verify the predictions of the simulations. Some real incidence curves are presented in Figure 6.15, aggregated into epidemic weeks to minimize fluctuations in the series. Figure 6.15-(a) presents confirmed cases curves for intermediate regions of São Paulo in comparison with the curve for the entire state (solid black line). It can be observed that there are variability for epidemic curves, a result in agreement with the model predictions for São Paulo. The same can be said of Rio de Janeiro (Figure 6.15(b)), where the prediction of small variability in epidemic peaks was confirmed qualitatively except the Petrópolis region, and for Ceará (Figure 6.15(c)). This variability was also observed in smaller scales to intermediate regions. In Figure 6.15-(d), we have immediate regions of Sobral/CE, in which dispersion of the prevalence curves can be observed. At municipalities scale, data is extremely noisy, which makes it difficult to make conclusions.

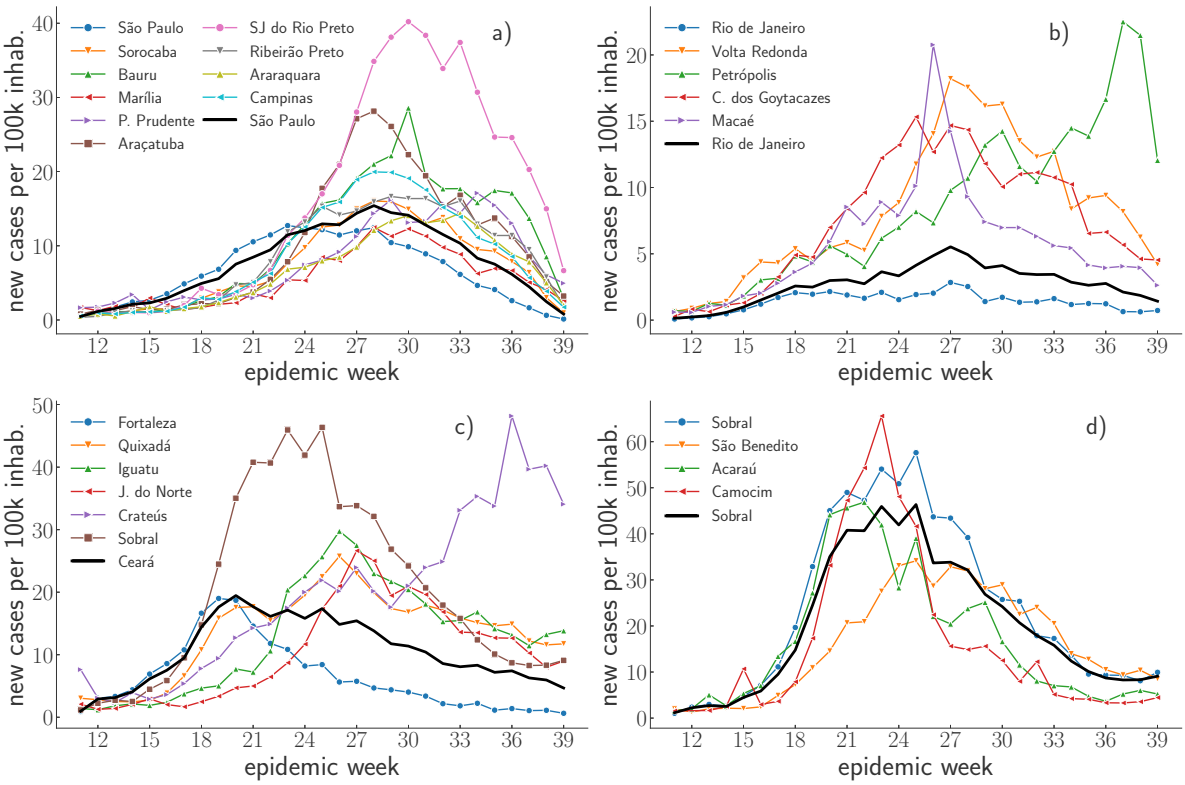


Figure 6.15: Real time series of epidemic incidence per week for intermediate regions of the states of (a) SP, (b) RJ, and (c) CE and for the (d) immediate regions belonging to the intermediate region of Sobral in the state of CE. Thick lines are the incidence for lower resolution scale in the corresponding a)-c) state or d) intermediate region. Epidemic week 12 correspond to March 15-21, 2020.

6.4 Summary and discussions

The wide territorial, demographic, infrastructural diversities of large countries such as Brazil demand modeling of a pandemic with geographical resolution higher than the usual compartmental epidemic models, which can be achieved using the metapopulation framework. We developed a stochastic metapopulation model and applied it to the COVID-19 spreading in Brazilian municipalities during the first epidemic wave. Performing simulations with different hypothetical mitigation scenarios and considering the integration of regions through long-distance recurrent mobility [64], we have identified a high degree of heterogeneity and desynchronization of the epidemic curves between the main metropolitan areas and other inland regions. The diversity of outcomes is observed in several geographical scales, from municipalities to federative units, with these differences being more remarkable the stricter the mitigation strategies. Moreover, we found moderate and strong correlations between the delay of epidemic peak in the countryside with the distance to capital cities for most well connected states. At the time of this work development, during the first wave of COVID-19 in 2020, the pandemic scenario was very different from the current one. Several caracteristics about the pathogen were discovered until now, new variants emerged and vaccines were developed. At that time, our simulations suggested that uniform mitigation measures were not the best strategy. In a city where the epidemic peak would happen later than the capital city of its state, the adoption of strict measures synchronously to the capital will delay the peak even more. Thus, such a place probably would have to extend the measures for longer periods, since once the epicenters start to relax their restrictions, the virus will circulate fast, reaching these vulnerable municipalities, in which most of the population is still susceptible. Therefore, the social and economic impacts would be higher. On the other hand, an eventual collapse of the local health system in the countryside regions may occur after the larger metropolitan areas were under control, opening the possibility to unburden the healthy care system. In this way, the desynchronization of outbreaks could be an asset to design optimal resource allocations.

Chapter 7

Data-driven epidemic model to assess undocumented infections

Related publication:

Data-driven approach in a compartmental epidemic model to assess undocumented infections

Guilherme S. Costa, Wesley Cota and Silvio C. Ferreira

Chaos, Solitons & Fractals, 112520 (2022)

The most stringent epidemic characteristic of the SARS-CoV-2 contagion may be its high transmissivity before the onset of symptoms (presymptomatic) [65, 66, 67] or even for those that never manifest detectable symptoms (asymptomatic) [68, 69]. Thus, the nowcasting and forecasting of COVID-19 spreading, fundamental tools to guide policy makers' decisions, are much more complex to perform, due do unknown fraction of undocumented infections. Whilst Sars-CoV-2 biology and interaction with humans are better understood nowadays, the behavioral aspects of the spreading such as usage of masks, vaccination support remains unpredictable. Aiming to soften these drawbacks, we developed a method based on a compartmental model to estimate the undocumented cases across different regions of Brazil. The method was applied to COVID-19 cases series from Paraná and Espírito Santo, allowing us to measure the heterogeneity level of under-reporting across different places. We start this chapter by detailing the methodology, the compartmental model and some analytical results. Then, we proceed to applications of the method to epidemiological data.

7.1 Compartmental model

We made some changes on the SEAUCR model, described in the last chapter [104], to incorporate additional compartments to improve our methodological approach. The first improvement is to create the Deceased (D) compartment, since under-reporting is intrinsically connected to discrepancies between **case fatality ratio** (CFR), extracted from death and reported cases series [156] and **infectious fatality ratio** (IFR), extrapolated from epidemiological surveys [157, 158]. Another important modification comes from the differentiation between diagnosed asymptomatic C_A and symptomatic individuals C_I, which have different detection rates and outcomes from the infection. The distinction between R, R_A, and R_I were made to ease the computation of detected cases. Figure 7.1 depicts the complete epidemiological chain and transition rates.

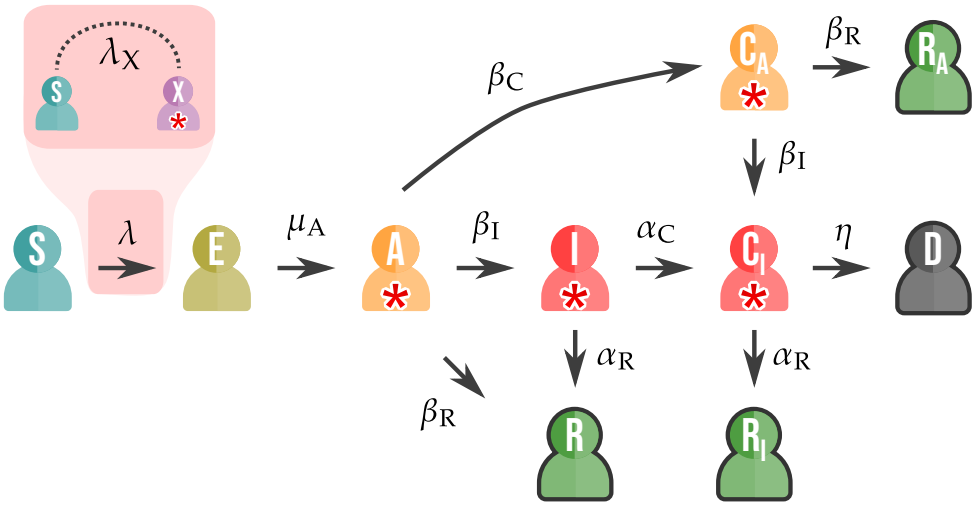


Figure 7.1: Schematic representation of the SEAICRD epidemic model including the following compartments: susceptible (S), exposed (E), asymptomatic (A), symptomatic (I), recovered (R, R_A, and R_I), deceased (D), and confirmed cases (C_A and C_I). The transition and respective rates are indicated by arrows. The infectious compartments are depicted with the symbol $\star$. The infection processes, represented by the dashed line, involve the interaction between susceptible and one of the infectious compartments, happening with rates λ_X, X=A, I, C_A, and C_I, which can depend on the compartment.

Susceptible (S) individuals in contact with infectious ones become exposed (E) with rates λ_A and λ_I, respectively. While the complete model allows the infection by diagnosed individuals C_A and C_I with rates λ_{C_A} and λ_{C_I}, for sake of simplicity, we considered that documented infectious C_A and C_I will isolate themselves, therefore

not contributing to further spreading of the pathogen. All remaining transitions are considered spontaneous. Exposed individuals evolve to asymptomatic (A) state with rate μ_A, which may develop symptoms and become symptomatic (I) with rate β_I, recover with rate β_R (R) or be tested and diagnosed with rate β_C (C_A). If it becomes a symptomatic case, the individual may be diagnosed with rate α_C (C_I) or recover with rate α_R (R). The clinical evolution of diagnosed cases is similar than undocumented ones, with addition that C_I individuals may die with rate η and become deceased (D). By assuming that death series are more accurate regarding under-reporting, we may disregard undocumented deaths. It can be noticed that true asymptomatic and presymptomatic cases, two important epidemic states [68, 69], are covered in the model by routes A→R (C_A →R_A) and A→I (A→C_A →C_I), respectively. We assume the same recovering times for documented and undocumented cases using the following constraints

$$\beta_I^{-1} = \beta_I'^{-1} + \beta_C^{-1}, \tag{7.1}$$

$$\beta_R^{-1} = \beta_R'^{-1} + \beta_C^{-1}, \tag{7.2}$$

$$\alpha_R^{-1} = \alpha_R'^{-1} + \alpha_C^{-1}. \tag{7.3}$$

With the added complexity of new compartments and the impossibility of working with a ABM due to computational and methodological costs, we changed the approach to a set of differential equations under homogeneous mixing hypothesis. In this way, for a constant population $N = \sum_X N_X$, where N_X is the number of individuals in compartment X, we can construct a set of equations to represent this model:

$$\partial_t S = -(\lambda_A A + \lambda_I I + \lambda_{C_A} C_A + \lambda_{C_I} C_I) S\,, \tag{7.4a}$$

$$\partial_t E = (\lambda_A A + \lambda_I I + \lambda_{C_A} C_A + \lambda_{C_I} C_I) S - \mu_A E\,, \tag{7.4b}$$

$$\partial_t A = \mu_A E - (\beta_I + \beta_R + \beta_C) A\,, \tag{7.4c}$$

$$\partial_t I = \beta_I A - (\alpha_R + \alpha_C) I\,, \tag{7.4d}$$

$$\partial_t R = \alpha_R I + \beta_R A\,, \tag{7.4e}$$

$$\partial_t C = \beta_C A + \alpha_C I\,, \tag{7.4f}$$

$$\partial_t C_A = \beta_C A - (\beta_R' + \beta_I') C_A\,, \tag{7.4g}$$

$$\partial_t C_\mathrm{I} = \alpha_\mathrm{C} I + \beta'_\mathrm{I} C_\mathrm{A} - (\alpha'_\mathrm{R} + \eta) C_\mathrm{I}\,, \tag{7.4h}$$

$$\partial_t R_\mathrm{A} = \beta'_\mathrm{R} C_\mathrm{A}\,, \tag{7.4i}$$

$$\partial_t R_\mathrm{I} = \alpha'_\mathrm{R} C_\mathrm{I}\,, \tag{7.4j}$$

$$\partial_t D = \eta C_\mathrm{I}\,, \tag{7.4k}$$

where $X = N_\mathrm{X}/N$, $\mathrm{X} \in \{\mathrm{S}, \mathrm{E}, \ldots, \mathrm{D}\}$, are the corresponding population fraction in compartment X.

7.2 Estimating epidemiological parameters and undocumented infections

The calculation of basic reproductive number R_0, given by Eq. (3.13), is straightforward

$$R_0 = \frac{1}{\beta_\mathrm{I} + \beta_\mathrm{C} + \beta_\mathrm{R}} \left[\lambda_\mathrm{A} + \frac{\beta_\mathrm{C} \lambda_{\mathrm{C_A}}}{\beta'_\mathrm{R} + \beta'_\mathrm{I}} + \frac{\beta_\mathrm{I}}{\alpha_\mathrm{C} + \alpha_\mathrm{R}} \left(\lambda_\mathrm{I} + \frac{\alpha_\mathrm{C} \lambda_{\mathrm{C_I}}}{\eta + \alpha'_\mathrm{R}} \right) \right] . \tag{7.5}$$

By neglecting infection of documented cases $\mathrm{C_A}$ and $\mathrm{C_I}$, the expression simplifies to

$$R_0 = \frac{1}{\beta_\mathrm{I} + \beta_\mathrm{C} + \beta_\mathrm{R}} \left[\lambda_\mathrm{A} + \lambda_\mathrm{I} \frac{\beta_\mathrm{I}}{\alpha_\mathrm{C} + \alpha_\mathrm{R}} \right] . \tag{7.6}$$

For a given set of epidemiological parameters, initial insights of the epidemic spreading may be inferred by calculating R_0 or the effective counterpart $R_t = R_0 \times S$.

Despite the model equations being constructed in terms of transition rates, some parameters are more intuitive if we consider probabilities instead. In particular, the testing rates β_C and α_C are better parameterized as the probabilities p_A and p_I that individuals are diagnosed during infectious phases. These parameters may be written in terms of the compartmental model rates as

$$p_\mathrm{A} = \frac{\beta_\mathrm{C}}{\beta_\mathrm{I} + \beta_\mathrm{C} + \beta_\mathrm{R}} \qquad p_\mathrm{I} = \frac{\alpha_\mathrm{C}}{\alpha_\mathrm{C} + \alpha_\mathrm{R}}. \tag{7.7}$$

Thus, the chance of a given exposed individual being diagnosed can be expressed as

$$\mathcal{P}_\mathrm{C} = p_\mathrm{A} + (1 - p_\mathrm{A}) p_\mathrm{I} \left(\frac{\beta_\mathrm{I}}{\beta_\mathrm{I} + \beta_\mathrm{R}} \right) . \tag{7.8}$$

The first and second terms in the r.h.s of the equation are due to diagnosis during asymptomatic and symptomatic phases, respectively. The chance of recovering without diagnosis is the complementary probability $\mathcal{P}_R = 1 - \mathcal{P}_C$. For ease of notation we will define $\phi = \beta_I/(\beta_I + \beta_R)$. Therefore, we can write the under-reporting coefficient σ_{ur} in terms of the documented and undocumented individuals N_C and N_R, respectively:

$$\sigma_{ur} = \frac{N_R}{N_C} = \frac{1 - \mathcal{P}_C}{\mathcal{P}_C} = \frac{(1 - p_A)(1 - \phi p_I)}{p_A + (1 - p_A)p_I\phi}, \tag{7.9}$$

where

$$N_C = N_{C_A} + N_{C_I} + N_{R_A} + N_{R_I} + N_D. \tag{7.10}$$

Considering the model parameters, we can divide them into two main groups: the biological and social ones. The first can be, at least in principle, estimated on epidemiological surveys [159, 160, 161, 162, 163, 164]. This category includes the parameters μ_A, β_I, β_R, α_R, and η. On the other hand, the social parameters may have a biological component, but behavioral and socioeconomic factors such as social distancing [165], mask wearing [166] and testing policies [167, 168] are determinant for estimating them. All these aspects are very heterogeneously distributed across time and different places therefore requiring a local approach. This category includes λ_A, λ_I, β_C (p_A) and α_C (p_I). It is these parameters that the method aims to estimate on the following paragraphs.

We start by determining the model's IFR (ℓ_{IFR}) considering the probabilities that exposed individuals evolve to death by following the two epidemic chains (E $\to$ A $\to$ I $\to$ C_I $\to$ D) or (E $\to$ A $\to$ C_A $\to$ C_I $\to$ D) which are $p_A\frac{\phi\eta}{\eta+\alpha_R}$ and $(1 - p_A)\phi p_I\frac{\eta}{\eta+\alpha_R}$, respectively. Rearranging the terms, we obtain the following equation for ℓ_{IFR}

$$\ell_{IFR} = [p_A + (1 - p_A)p_I]\frac{\phi\eta}{\eta + \alpha_R}. \tag{7.11}$$

Then, we need to connect the ℓ_{IFR} with CFR (ℓ_{CFR}), which can be extracted from data series. Let us set $\mathcal{C}(t)$ and $\mathcal{D}(t)$ as the cumulative series of confirmed cases and deaths, respectively. To equate ℓ_{CFR}, we need to be careful when defining the time window $[t, t + \Delta\tau]$ as individuals positively diagnosed on day t_{cal} will evolve to death

on the following days. In this way, we need to compare the reported cases on t with obits t_{delay} days ahead. Thus,

$$\ell_{\text{CFR}}(t - t_{\text{delay}}) \equiv \frac{\Delta\mathcal{D}(t)}{\Delta\mathcal{C}(t - t_{\text{delay}})}. \tag{7.12}$$

The CFR can be calculated in terms of the model rates, given by the conditional probability that an infection will evolve to death given that it was diagnosed. By defining $\mathbb{D}$ and $\mathbb{T}$ as the events of death and diagnosis of infected individuals, we have that

$$\ell_{\text{CFR}} \equiv \Pr(\mathbb{D}|\mathbb{T}) = \frac{\Pr(\mathbb{D} \cap \mathbb{T})}{\Pr(\mathbb{T})} = \frac{\Pr(\mathbb{D})}{\Pr(\mathbb{T})} = \frac{\ell_{\text{IFR}}}{\mathcal{P}_{\text{C}}}, \tag{7.13}$$

where we have used the Bayes rule for conditional probabilities, the hypothesis that only diagnosed individuals evolve to death, and Equation (7.8). Rearranging the terms, one obtains

$$\frac{\ell_{\text{IFR}}}{\ell_{\text{CFR}}} = \mathcal{P}_{\text{C}} = p_{\text{A}} + (1 - p_{\text{A}})p_{\text{I}}\phi. \tag{7.14}$$

Equation (7.14) is very handy since it relates the testing rates (or probabilities) with quantities and parameter that, in principle, may be obtained directly from data. Therefore, if the ratio $r = p_{\text{A}}/p_{\text{I}}$ is given, the testing rates can be calculated as

$$p_{\text{I}} = \frac{r + \phi}{2r\phi}\left[1 - \left(1 - \frac{4r\phi\mathcal{P}_{\text{C}}}{(r + \phi)^2}\right)^{1/2}\right]. \tag{7.15}$$

The under-reporting coefficient, which can be expressed as

$$\sigma_{\text{ur}} = \frac{1 - \mathcal{P}_{\text{C}}}{\mathcal{P}_{\text{C}}} = \frac{\ell_{\text{CFR}}}{\ell_{\text{IFR}}} - 1, \tag{7.16}$$

is extracted directly from data using Eq. (7.12).

The estimates for infection rates (λ_{A}; λ_{I}) and the initial conditions, i.e., the number of individuals in each compartment of our model, require more complex approaches, since the relations between these parameters and the data series of reported cases and deaths is non trivial. Thus, we developed an iterative calibration procedure for these estimates, to be described on the next subsection.

7.2.1 Calibration procedure

To perform the calibration we will consider a time interval $[t_{\text{cal}}, t_{\text{cal}} + \Delta\tau]$ for which $\mathcal{C}(t)$ and $\mathcal{D}(t)$will be analyzed. This time window should be short enough to consider that λ_{A} and λ_{I} are approximately constant but sufficiently large to have a reasonable amount of data. We also need to compute ℓ_{CFR} within this time window. Then, we set a given value for r (to be assumed as a parameter of the model) and determine probabilities p_{I} and p_{A} using Equation. (7.15). In order to estimate the other compartments from these parameters, we will rely on an **adiabatic approximation** for the susceptible population, i.e., that S vary much more slowly than other compartments such that variations on it may be neglected during the investigated period ($S(t) \approx S^*$). From there, we can condense our notation by defining $\gamma_{\text{A}} = \lambda_{\text{A}} S^*$ and $\gamma_{\text{I}} = \lambda_{\text{I}} S^*$.

Under these assumptions the SEAICRD model provides a closed linear system $\dot{\mathbf{X}} = \mathbb{J}\mathbf{X}$ for compartments $\mathbf{X} = (E, A, I)$, the Jacobian given by

$$\mathbb{J} = \begin{bmatrix} -\mu_{\text{A}} & \gamma_{\text{A}} & \gamma_{\text{I}} \\ \mu_{\text{A}} & -(\beta_{\text{I}} + \beta_{\text{R}} + \beta_{\text{C}}) & 0 \\ 0 & \beta_{\text{I}} & -(\alpha_{\text{R}} + \alpha_{\text{C}}) \end{bmatrix}. \tag{7.17}$$

By assuming that the solution of this linear system is ruled by the leading term $\mathbf{X} \sim \mathbf{v}_1 \exp[\Lambda_1(t - t_{\text{cal}})]$ where Λ_1 is the largest eigenvalue of $\mathbb{J}$ and $\mathbf{v}_1 = (v_{\text{E}}, v_{\text{A}}, v_{\text{I}})$ is its corresponding eigenvector, we can provide two relations among initial conditions (E^*, A^*, I^*)

$$\frac{E^*}{A*} \approx \frac{v_{\text{E}}}{v_{\text{A}}}, \tag{7.18}$$

$$\frac{I^*}{A^*} \approx \frac{v_{\text{I}}}{v_{\text{A}}}. \tag{7.19}$$

Then, by integrating Equation (7.4f), we can connect the initial conditions for (E^*, A^*, I^*) with the increment of confirmed cases $\Delta\mathcal{C}$ during the interval $\Delta\tau$,

$$\Delta\mathcal{C} \approx (\beta_{\text{C}} A^* + \alpha_{\text{C}} I^*) \frac{e^{\Lambda_1 \Delta\tau} - 1}{\Lambda_1}. \tag{7.20}$$

If assuming that $\Lambda_1 \Delta\tau \ll 1$ we can further simplify the equation, writing it as

$$\beta_{\text{C}} A^* + \alpha_{\text{C}} I^* \approx \frac{\Delta\mathcal{C}}{\Delta\tau}. \tag{7.21}$$

Thus, we can write I^* as

$$I^* \approx \frac{\Delta \mathcal{C}}{\Delta \tau} \left(\frac{v_{\mathrm{I}}}{\alpha_{\mathrm{C}} v_{\mathrm{I}} + \beta_{\mathrm{C}} v_{\mathrm{A}}} \right). \tag{7.22}$$

The initial conditions for compartments E^* and A^* are straightforwardly calculated by using Equations (7.13) and (7.14). The estimates for the R compartment require knowledge of a wider data interval, since its compartment is a possible end for some epidemic routes. Therefore, the number of individuals in it accumulates over time. For pathogens that induces waning immunity or mutates relatively fast, such as Influenza [169] and COVID-19, it is sufficient to have data regarding outbreaks of a specific variant or of the same interval as the reinfection period of the pathogen. We will set this "initial outbreak" occurring at time t_{tr}. Then, we calculate the average σ_{ur} and number of reported cases N_{C}^* during the time window $[t_{\mathrm{tr}}, t_{\mathrm{cal}}]$ to find $N_{\mathrm{R}}^* = N_{\mathrm{C}}^* \times \langle \sigma_{\mathrm{ur}} \rangle$. Finally, the susceptible population is determined as

$$N_{\mathrm{S}}^* = N - \sum_{\mathrm{X} \neq \mathrm{S}} N_{\mathrm{X}}. \tag{7.23}$$

The last hurdle of this method is to estimate the infection rates λ_{A} and λ_{I} which we will do through an iterative and self-consistent method relied in everything we've done so far. This calibration proceed as follows

i). Start with initial guesses for $\gamma_{\mathrm{A}} = \lambda_{\mathrm{A}} S^*$ and $\gamma_{\mathrm{I}} = \lambda_{\mathrm{I}} S^*$.

ii). From this guess, calculate the initial condition of the dynamic by following the steps discussed above. Calculation of Λ_1 and $\mathbf{v}_1$ can be made using the power method [139].

iii). With the initial conditions, integrate Equations (7.4b) to (7.4f) during the interval $[t_{\mathrm{cal}}, t_{\mathrm{cal}} + \Delta\tau]$.

iv). Calculate the dispersion with respect to reported cases counts as

$$\Omega(\gamma_{\mathrm{I}}, \gamma_{\mathrm{A}}) = \int_{t_{\mathrm{cal}}}^{t_{\mathrm{cal}}+\Delta\tau} \left[\mathcal{C}(t) - C(t) \right]^2 dt. \tag{7.24}$$

v). Construct new guesses for γ_{A} and γ_{I} using the multi-parametric bissection method [139].

vi). Repeat steps (ii) to (v) in order to minimize $\Omega(\gamma_I, \gamma_A)$.

It is important to remark that the method's performance is dependent on data quality and good estimates of epidemiological parameters. Changes in the pathogen dynamic such as the arising of new variants or the development of medical treatments will possibly alter these parameters. However, the method was conceived to easily incorporate these changes by including new transitions or updating the existing ones.

7.2.2 Infection fatality ratio and vaccination

The IFR is the most important parameter of our method, due to directly influencing the under-reporting index. Brazilian vaccination started in mid-February 2021, having an overlapping time window with our analysis. Thus, to estimate ℓ_{IFR} more accurately, we need to take into account the impacts of vaccination on it. Verity et al [170] reported an age-dependent IFR for COVID-19, yielding an exponential increase with age and average 0.68% for Brazil. The fraction of population who completed the vaccination scheme was extracted from surveillance systems and is publicly available at Reference [156]. In addition, demographic data for Brazilian states were obtained from IBGE [145]. These data is available in Table 7.1

Age group	IFR (%)	Percentage of population			
		PR	ES	AM	SP
0-4	1.61×10^{-3}	6.84	7.07	9.59	6.56
5-9	1.61×10^{-3}	6.68	6.98	9.67	6.57
10-14	6.95×10^{-3}	6.53	6.72	9.48	6.38
15-19	6.95×10^{-3}	7.01	7.17	9.22	6.75
20-24	0.0309	7.89	7.73	9.63	7.44
25-29	0.0309	8.00	7.78	8.79	7.58
30-34	0.0844	7.73	8.31	8.14	8.23
35-39	0.0844	7.57	8.21	7.65	8.32
40-44	0.161	7.20	7.51	6.69	7.66
45-49	0.161	6.72	6.41	5.42	6.78
50-54	0.595	6.47	5.99	4.41	6.27
55-59	0.595	5.82	5.47	3.52	5.68
60-64	1.93	4.81	4.67	2.70	4.89
65-69	1.93	3.81	3.62	1.95	3.84
70-74	4.28	2.79	2.45	1.30	2.80
75+	7.8	4.05	3.83	1.77	4.16

Table 7.1: Estimated infection fatality ratio of COVID-19 per age group and percentage of population for some states from Brazil whom belongs to certain age group.

The vaccination strategy followed in Brazil was a decreasing age prioritization one, in which people were vaccinated in age descending order. We considered $N_g = 16$ age groups, in which $g = 1$ corresponds to ≥ 75 years old, $g = 2$ to $70 - 74$ years and so on. Thus, the vaccines were distributed according to this sequence. The average vaccine free IFR is given by

$$\ell_{\text{IFR}} = \sum_{g=1}^{N_g} \ell_g n_g, \tag{7.25}$$

in which ℓ_g and n_g are the IFR and population fraction in the age group g, respectively. Considering x as the total fraction of vaccinated population, the lower group g^* who got vaccine shots is given by

$$\sum_{g=1}^{g^*} f_g n_g < x < \sum_{g=1}^{g^*+1} f_g n_g, \tag{7.26}$$

in which f_g is the fraction of g who got vaccinated. Finally, we consider r_g as the IFR reduction of the vaccinated population g. Thus, the corrected IFR becomes

$$\ell_{\text{IFR}} = \sum_{g=1}^{N_g} \ell_g n_g - \sum_{g=1}^{g^*} \ell_g n_g f_g (1 - r_g). \tag{7.27}$$

For the sake of simplicity, we assumed uniform IFR reductions and vaccination coverage among all groups, which implies $f_g = f = 0.85$ and $r_g = r = 0.05$ (95% effectiveness against death). These choices are consistent with typical protection rates of vaccines used in Brazil. Figure. 7.2 presents the IFR as a function of time for four Brazilian states. The lower IFR for Amazonas's state reflects its young population, as can be seen in Table 7.1. Similar patterns are observed for the other states. To account for fluctuations in the IFR, we consider ℓ_{IFR} draw for a uniform distribution with average calculated by Equation (7.27) and 30% of uncertainty in it. It is worth noting that this is a simple approach that aims at being qualitatively rather than quantitatively accurate.

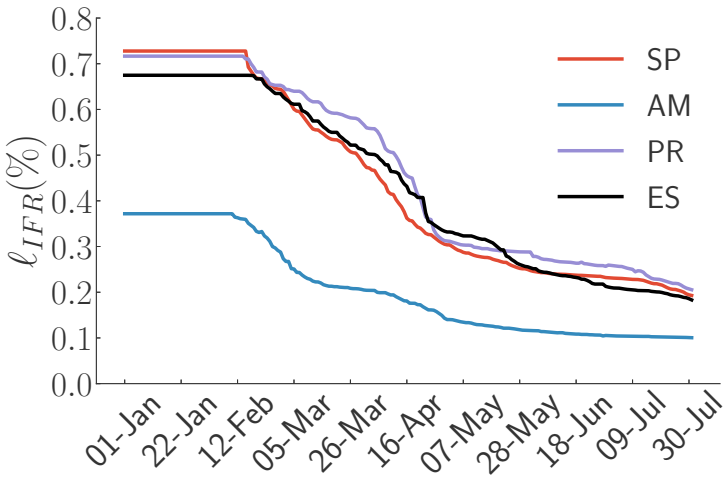


Figure 7.2: Infection fatality ratio as a function of time, estimated for São Paulo (SP), Amazonas (AM), Paraná (PR), and Espírito Santo (ES) states considering their demographics and vaccination rates.

7.3 Estimates for the under-reporting coefficient

The first application for our method was to estimate the under-reporting coefficient for some reported cases data. Two main types of daily count series were available for Brazil, which will be refereed as **Type-I** and **Type-II** from now on:

- Type-I: count series using release dates provided by epidemic surveillance departments of Brazilian federation units which are aggregated and publicly available for all Brazillian municipalities [156]. These data do not yield the date of diagnosis and may present uncontrolled bias caused by reporting delays and should be used with care.

- Type-II: contain dates of diagnosis and first symptoms onset. In this work, we use the publicly available Type-II data for Paraná (PR) [171] and Espirito Santo (ES) [172] states.

The aggregated data used in this chapter is available on [173]. In addition to count series of deaths and reported cases, another important input to SEIAICRD are the epidemiological parameters. Those who have estimates through epidemic surveys are provided in Table 7.2.

Parameter	Transition	Values	References
μ_{A}^{-1}	$E \to A$	$G(6.4, 2.0)$ days	[159, 29]
β_{I}^{-1}	$A \to I$	$G(6.4, 2.0)$ days	[159, 29]
α_{R}^{-1}	$I \to R$	$G(6.4, 2.0)$ days	[25]
β_{R}^{-1}	$C_{\mathrm{A}} \to R_{\mathrm{A}}$	$\beta_{\mathrm{I}}^{-1} + \alpha_{\mathrm{R}}^{-1}$	[29, 104]

Table 7.2: Epidemiological parameters for SEAICRD model. $G(a, b)$ implies that the parameter was draw from a Gamma distribution with shape and rate parameters a and b, respectively (mean $= a/b$; variance = a/b^2).

7.3.1 Preliminary analysis

We calculated σ_{ur} for Type-I count series of two capital cities of Brazil: Manaus and São Paulo, which were severely impacted by COVID-19 second infection wave [174]. As aforementioned, the time series of confirmed cases and deaths have a typical delay associated with them, i.e., the reported cases of a given day are usually not related to deaths on the same day. In order to ease this discrepancy, we shift the series until peaks in cases and deaths coincide. Figure 7.3 shows both time series. In those curves, the delay was estimated as $t_{\mathrm{delay}} = 7$ days for Manaus and 9 days for São Paulo. It is important to notice that this shift does not need to be performed empirically if data regarding this delay is available within the count reports.

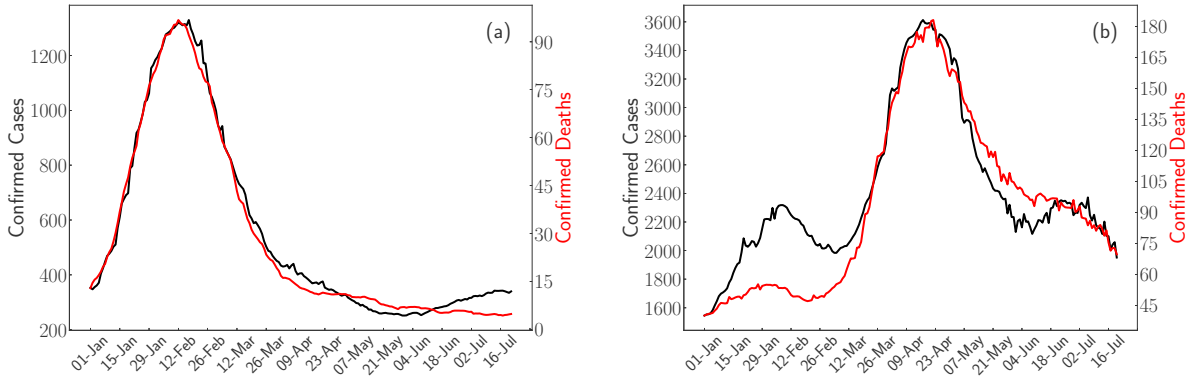


Figure 7.3: Time series of confirmed cases (left axis, black curves) and deaths (right axis, red curves) for cities of (a) Manaus/AM and (b) São Paulo/SP. Time series of death counts are shifted in (a) 7 and (b) 9 days, respectively.

Figure 7.4 show the underreporting coefficient for Manaus and São Paulo within a confidence interval of 95%. It can be seen that the underreporting for Manaus remains high throughout the analyzed period, ranging from 10 to 25. For São Paulo,

the under-reporting is relatively low in the first months but increases considerably in the final analyzed period.

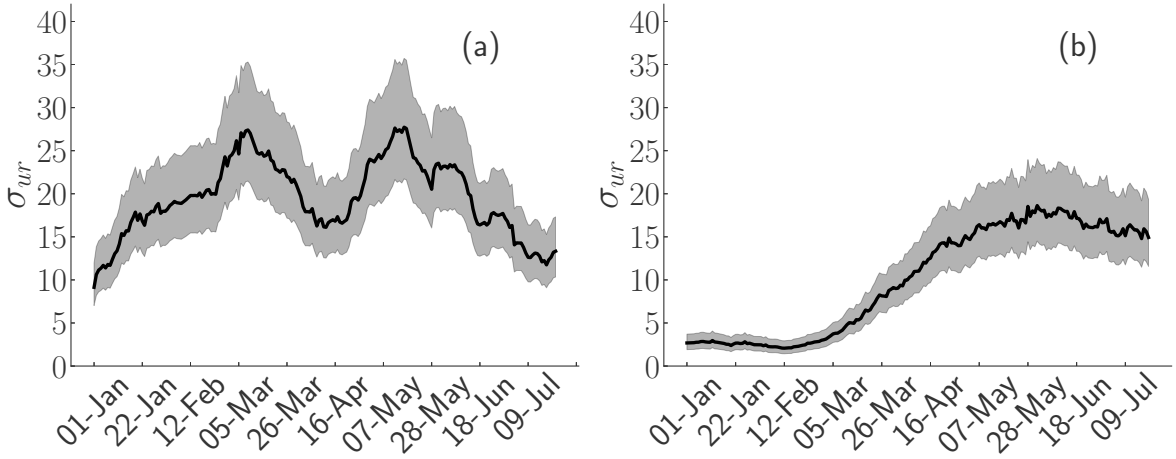


Figure 7.4: Evolution of σ_{ur} for the capital cities of (b) Manaus and (c) São Paulo estimated using moving time windows of 3 weeks for type-I count series (see main text) as notified by state surveillance departments [156]. The 95% interval of confidence is shown in the shaded region.

7.3.2 PR and ES states

Moving on to Type-II report series, we found that for PR and ES states, $t_{\text{delay}} \approx 10$ and 20 days, respectively. Then, we computed the under-reporting coefficient by aggregating data of municipality level into immediate regions (see Section 6.1) starting from January 2021, that coincides with the spreading of P1 variant in Brazil [175]. In order to not pollute the figures, in graphs that do not have a confidence interval on σ_{ur}, we assumed that the IFR is the average value of ℓ_{IFR}. The evolution of σ_{ur} for the aggregated data on PR and ES states jointly to two immediate regions are shown in Figures 7.5-(a) (PR) and (b) (ES). The main outcome is the substantial variation of the undocumented infections along the time and across different localities. For example, in Nova Venécia-ES, the under-reporting fluctuates from approximately 2 to higher than 10 while in Vitória-ES we have a smoother curve reaching $\sigma_{\text{ur}} \approx 10$. The evolution of σ_{ur} for each immediate region of PR and ES within confidence intervals can be found in Figures B.12 and B.14 of Appendix B.3. By calculating the CFR for those time series, we can see that ℓ_{CFR} and σ_{ur} present different temporal correlations, as can be seen by comparing Figure 7.5-(c) with the under-reporting for PR and ES states, despite the proportionality depicted in Equation (7.14).

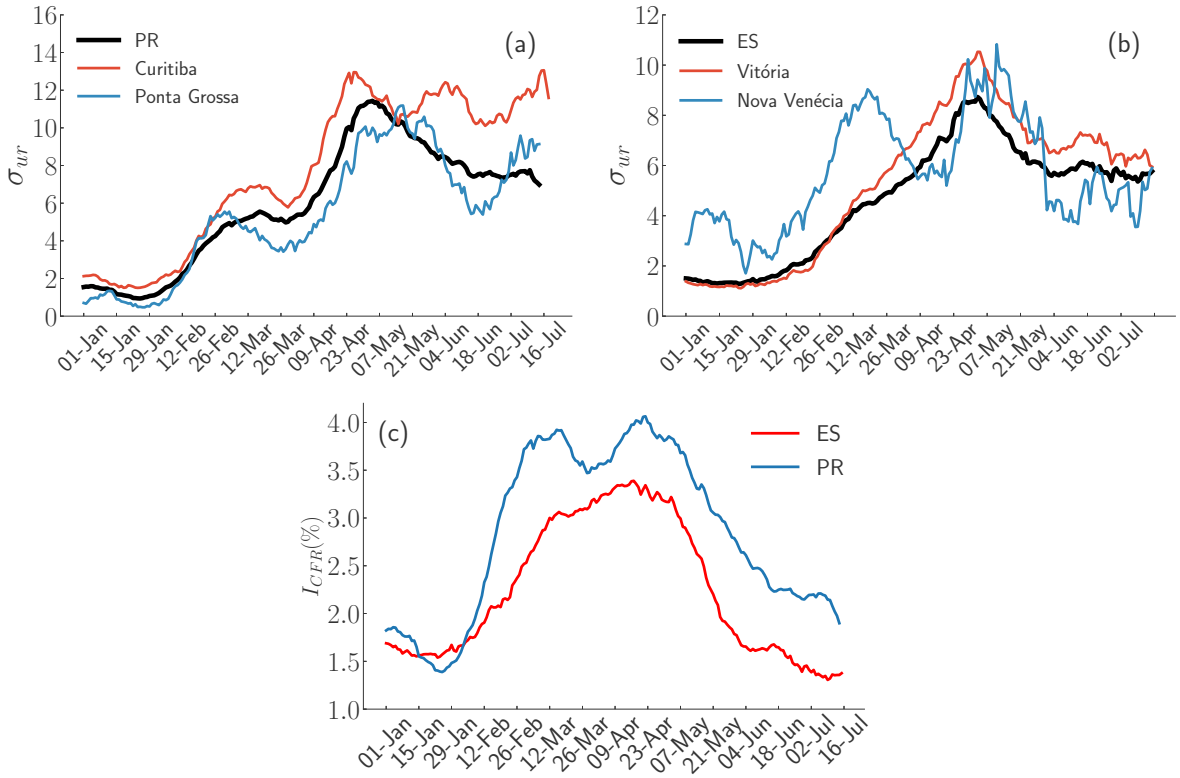


Figure 7.5: Evolution of under-reporting coefficient for (a) PR and (b) ES states with time windows of 3 weeks. Two immediate regions of each state are presented in the corresponding panels. (c) Evolution of the CFR computed using delays $t_{\text{delay}} = 10$ d and 20 d for PR and ES states, respectively.

To visualize the heterogeneity on σ_{ur} in more detail, we collected the under-reporting coefficients for all immediate regions of PR and ES in two periods of the pandemic: around mid-January, when the CFR was low (mild spreading) and around mid-April, when CFR was high (intense spreading). These coefficients are shown on Figures 7.6-(a) for PR regions and (b) for ES ones. Averaged coefficients for the whole states were also shown. The space-time variability on σ_{ur}, which can be differ largely in a same time interval, reflects the outbreak diversities across different localities due to numerous factors such as unequal responses to pandemics, demographic and economic heterogeneity, levels of adherence on NPIs, among others. It is important to note that this desynchronization were also seen in Chapter 6 [104] and later observed in Reference [174] for the first epidemic wave in Brazil.

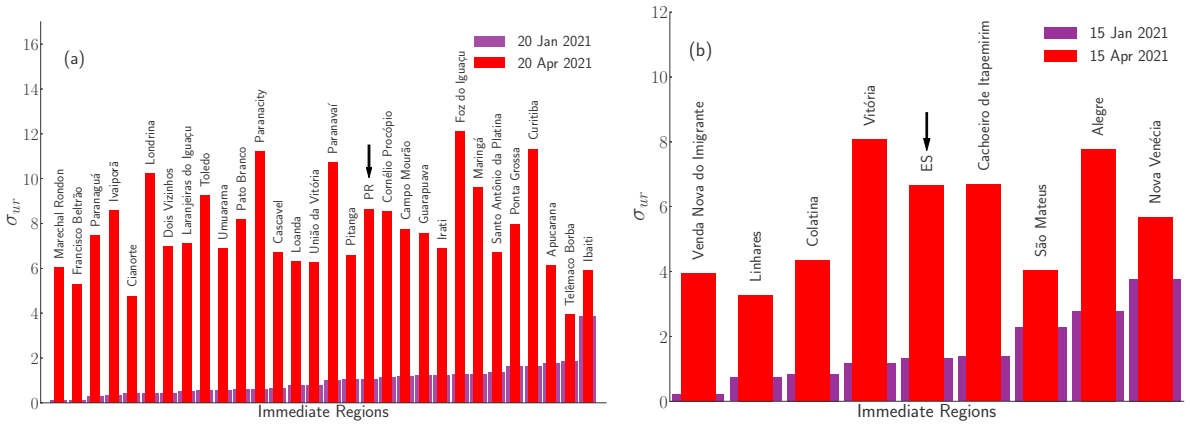


Figure 7.6: Under-reporting coefficients for all immediate regions of (d) PR and (e) ES and the for states (indicated by arrows) computed when the CFR is low (Jan 2021) and high (April 2021).

7.4 Assessing undocumented compartments and initial conditions

To exemplify the calibration method, we analyzed the reported cases for PR and ES states starting from January 2021, using a time window of 14 days and presenting a forecast of one week to verify the robustness of the procedure. With σ_{ur} previously calculated and $r = 0.1$, we performed the calibration described previously in order to find λ_{A} and λ_{I} that fits the data on the 14 day time window. For the sake of simplicity, we assumed that $\lambda_{\text{A}} = \lambda_{\text{I}}$, implying a single parameter to adjust the data. Typical calibration curves for PR and ES states are shown on Figures 7.7(a) and (b) for different times using a 14-day moving window. The initial day for each calibration is indicated on top left of the panels. The goodness of fit is quantitatively verified considering two simple statistical regression analysis: the Pearson coefficient r^2 and mean absolute percent error (MAPE) $M_{\%}$ [176]. The MAPE is calculated from $t = 0$ to $t = T = 21$ d as

$$M_{\%} = \frac{100}{T} \sum_{t=1}^{T} \left| \frac{\tilde{C}(t) - \tilde{\mathcal{C}}(t)}{\tilde{\mathcal{C}}(t)} \right|, \tag{7.28}$$

where $\tilde{C}(t) = C(t) - C(0)$ and $\tilde{\mathcal{C}}(t) = \mathcal{C}(t) - \mathcal{C}(0)$. As a general result, all Pearson coefficients are statistically significant for the regressions while the MAPE values are at most of order of 10%, implying that the method performs very well, reproducing the short-term progression of cumulative case count time series within the confidence

interval. Calibration for some immediate regions can be seen in Figure B.13 of Appendix B.3.

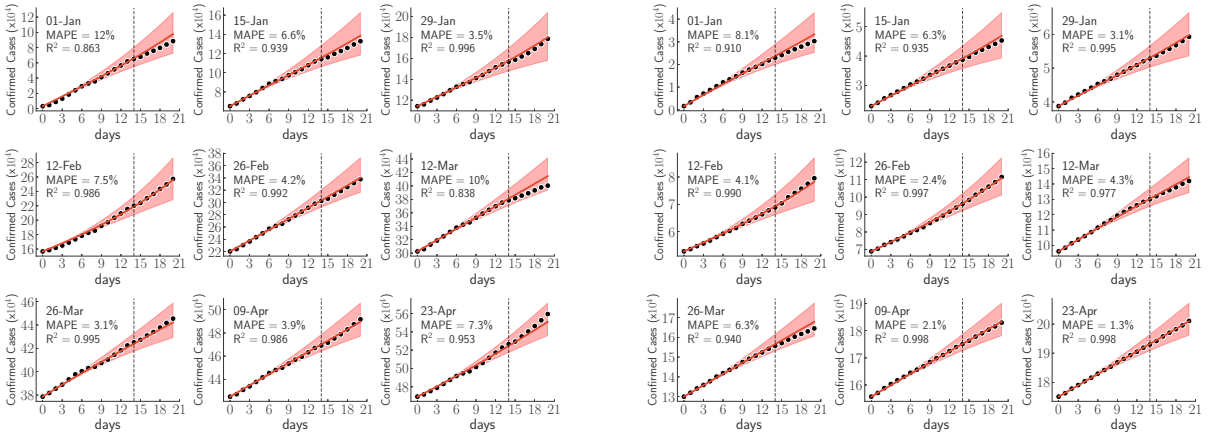


Figure 7.7: Calibration curves for PR (left) and ES (right) states in different time windows of 14 days indicated by the vertical lines. Initial day is in the top of each panel. One week of forecasting is also shown. Symbols are the cumulative cases' counts while lines with shaded regions represent the calibrated curves and the corresponding confidence interval of 95%.

From these calibrations, we can estimate the evolution of undocumented epidemic compartments (exposed, asymptomatic and symptomatic) together with the effective reproduction number R_t, as can be seen in Figures 7.8-(a) and (b) for PR state. Starting in January 1^{st} 2021, we performed calibrations within a 14-day moving window up to the end of May 2021. Regarding the undocumented compartments, it is important to notice that the total amount of infected individuals is much larger than the number of confirmed cases at a given day, being even higher than σ_{ur} estimated previously. This happens because the under-reporting coefficient is measuring the ratio between infections that end being documented (C_{A}, C_{I}, R_{A}, R_{I} and D) and those who ends undocumented (R) while the individuals at compartments E, A and I are still part of the epidemic chain and may or may not be documented in the future. Moreover, we see a general increasing in the quantity of individuals in undocumented compartments, an expected behavior when considering that the vaccination leads to milder infections and lower the seeking for medical care and testing. In addition, we analyzed the R_t for the same time window, as show in Figure 7.8-(b). It can be seen that the mean value of R_t oscillate between approximately 0.9 and 1.2 following the variations and inflections in case count series.

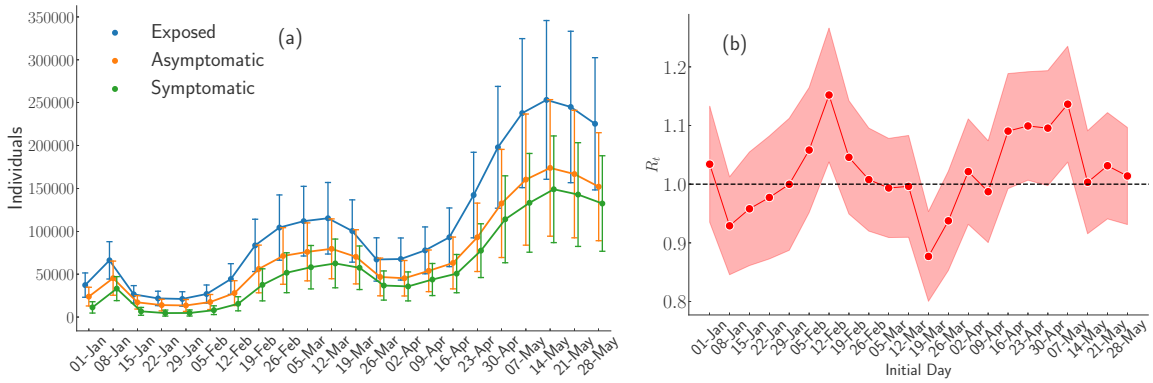


Figure 7.8: (a) Evolution of the undocumented compartments (exposed, asymptomatic and symptomatic) for the PR state since 1 January 2021. (b) Evolution of effective reproductive number computed for the PR state.

7.5 Summary and discussions

The COVID-19 pandemic led to unprecedented efforts gathering different fronts of human society in an attempt to minimize the spreading and damage caused by SARS-CoV-2. From attempting to predict the diseases' spread, to gathering and treating data, by performing epidemiological surveys or acting on the front line to fight the virus, many professionals have spent the last two years working on impacts of COVID-19. In particular, the publicly available data of reported diagnosed infections and deaths is usually incomplete, despite the efforts of many. Delays, limited capacity in testing and the unpredictable human behavior make it difficult to gather some types of data. Moreover, these limitations occur heterogeneously across different places and at different moments, making it difficult to apply general frameworks. However, this opens new venues for improving existing tools or creating new ones to extract implicit information from data. A promising framework is the data-driven approach [30, 29, 104], in which mathematical and mechanistic models are supplied by real world data, allowing us to make predictions which are not explicitly available.

We followed a data-driven approach based on a compartmental model with the main goal to estimate the presymptomatic and asymptomatic infections, that are difficult to be detected without heavy epidemic surveillance protocols. Our method allows to estimate the undocumented infections using some biological parameters, such as the case fatality ratio (inferred from controlled studies) and some empiric parameters, such as the probability of individuals looking for testing. As a case study,

we applied the method to epidemic series of deaths and reported cases on two Brazilian states, Paraná (PR) and Espírito Santo (ES), where days of symptoms onset were available, during the outbreak driven by Gamma variant (lineage P.1) on the first semester of 2021. Our main parameter was the under-reporting coefficient σ_{ur}, defined as the ratio between number of infections that ended diagnosed or not. Our analysis reports a large variability on σ_{ur} across space and time, result of heterogeneous and desynchronized public policies and outbreaks along Brazilian municipalities [104, 175]. Moreover, the method allow to estimate the number of individuals on other epidemic compartments, such as the asymptomatic and exposed ones and the basic and effective reproductive numbers R_0 and R_t.

The knowledge about SARS-CoV-2 evolves each day along with mutations and variants of the virus itself. Thus, some information used in this chapter may be outdated in some time. However, the framework developed on it is extremely general and can easily be adapted to include any new information regarding the virus. By updating the parameters we maintain the model accurate. In general, the performance of the model relies on good estimates for the biological parameters and high quality of data. Even without epidemic changes, the framework can be improved to include stratified data in the form of age contact matrices [178] or metapopulation approaches [29, 104].

Part IV

Concluding remarks

Chapter 8

Summary, forthcoming research and prospects

The use of complex networks for modeling interacting systems such as social interactions, urban centers, metabolic and ecological chains, among others, shows the interdisciplinarity of this framework. Allied to the tools of dynamical processes, a range of systems and problems can be addressed within this approach. From connections with non-equilibrium statistical mechanics, this same framework can be applied to fundamental and theoretical problems in physics. Among this myriads of applications, the study of spreading processes, whether information or epidemics, is a recent subject that has become increasingly important in the face of fake news or the outbreak of COVID-19. Within this context, computer simulations constitute an important part of validating theories and understanding phenomena that would be very difficult to do otherwise, either because of the large number of interactions among its elements or the lack of closed analytical solutions. Thus, the advance of network science is closely related to the advances of computing, due to greater processing powers implying in modeling larger systems with more complex interactions.

After an extensive review of networks and dynamical processes in Chapters 2 and 3, in which we discussed properties, metrics, characterization and models, in Chapter 4 we present a quasistationary method, namely reactivation per activity time (RAT), with the objective of adding the pros of other methods, in particular the simplicity of reflecting boundary condition and the exactness of standard quasistationary. The RAT

method lays on the reactivation of multiple vertices whenever the system falls into the absorbing state. Reactivated vertices are chosen proportionally to their activity time. For the susceptible-infected-susceptible (SIS) model, which we used to validate the method, a vertex is considered active whenever it is infectious. By defining a suitable activity time, the RAT method can be applied to other models. To validate the method, we performed simulations of the SIS model on a variety of complex networks, in particular those with localized activation, in which simpler methods, as the reflecting boundary condition, may fail. As a general result, the RAT method was able to correctly identify and characterize the transitions, providing the same exponents as the standard quasistationary but with lower computational and algorithmic cost. We expect that RAT will ease the study of more complex models with absorbing state transitions in the future.

We also developed an heterogeneous mean-field (HMF) theory for the two species symbiotic contact process (2SCP), a recently proposed cooperation process with a rich and complex phase diagram. This model was partially studied by using homogeneous theories but its behavior on highly heterogeneous substrates was not fully known. By constructing the HMF equations and performing numerical analysis of 2SCP on complex networks with a power law degree distribution $P(k) \sim k^{-\gamma}$, we observed a reduction of the bistability region in the infection rate λ versus symbiotic healing rate μ_s phase diagram as γ decreases. In addition, we found discontinuous transitions on finite systems that evolves to a continuous one as the system's size increases. Both behaviors were not found in previous theories. Simulations of the model on annealed and quenched power law networks confirmed the HMF theory results.

Moving on to applications, in Chapter 6 we developed a compartmental epidemic model to study the spreading of early COVID-19 in Brazil at municipality level, using a data-driven model with metapopulations. By considering each patch as a municipality and inputting demographic and mobility data into the model, we designed different mitigation strategies to investigate spreading scenarios. We were able to visualize the spatio-temporal patterns in the spreading, observing the desinchronization of the outbreaks and epidemic waves propagating from capital cities to the countryside, in general. The diversity of outbreaks were observed in different geographical scales, ranging from municipalities to federative units, with these differences being

less intense the milder the mitigation scenario. We expect that these analysis could help shape containment strategies for future outbreaks. It is important to note that these framework can be adapted to almost any infectious disease peculiarities.

In Chapter 7, we developed a method to estimate the undocumented infections on epidemic models, applying it to COVID-19 spreading. By constructing an epidemic compartmental model, we estimated the level of under-reporting on different geographical scales of Paraná and Espírito Santo states during a time window that coincides with the outbreak of P.1 lineage in Brazil. In addition, the method was able to estimate the number of individuals on some undocumented compartments, such as exposed, assymptomatic and presymptomatic ones. Finally, by performing a calibration of reported cases, we were able to present a short-time forecast in addition to the calculation effective reproductive number R_t.

All the works developed in this thesis emphasize the versatility of the dynamic processes framework jointly with complex networks to solve theoretical and fundamental problems as well as contribute to practical applications. As forthcoming research, we developed a study on epidemic spreading on spatial networks, initially thought to compose this thesis but abdicated for sake of text size and time. The introduction of spatial structures may be responsible for the emergence of relevant quenched disorder in complex networks. This concept, imported from condensed matter physics, refers to the existence of regions that, due to the presence of impurities or defects, behave differently from the rest of the system and end up influencing the global properties [45]. In complex networks, we have the same idea, one or several sets of nodes that behave differently from the whole, changing the dynamics in these networks. One of the most interesting effects of this quenched disorder when dealing with non-equilibrium phase transitions is the existence of critical extended regions, in which there is a slow decay (non-universal power laws) of the order parameter, in contrast to the usual critical points of phase transitions. This effect is called the Griffiths phase [179] and is a potential mechanism to explain the vast existence of criticality in real systems [10].

Among the various models of spatial networks, the one proposed by Rozenfeld et al. [180] is interesting because it mixes the ordering of a lattice with the heterogeneity of power law networks. In this model, we assign to each vertex of the network a

position on a lattice and k_i empty stubs, draw from a given distribution $P(k)$. Then, we connect each node with its k_i nearest available neighbors, i.e., neighbors who didn't connect all their k_j stubs. By embedding an heterogeneous network on a d dimensional Euclidian lattice, we generate a substrate with some peculiar properties such as hierarchical organization ($C(k) \sim k^{-\omega}$, $\omega > 0$), disassortative mixing ($K_{\mathrm{nn}}(k)$ decreasing on k) and fractal dimension ($\langle \ell \rangle \sim L \sim N^{1/d_f}$, $d_f \in \mathbb{R}$), see Figures 8.1(a) to (d) for graphics of these metrics. It is expected that the lattices ordering allied with the heterogeneity of a scale-free network produces rare regions on the system and, therefore, generates quenched disorder. Preliminary simulations of the SIS model on scale free networks embedded on square lattices show the existence of persistent Griffiths phases, i.e., the extended critical region is relevant on the thermodynamic limit.

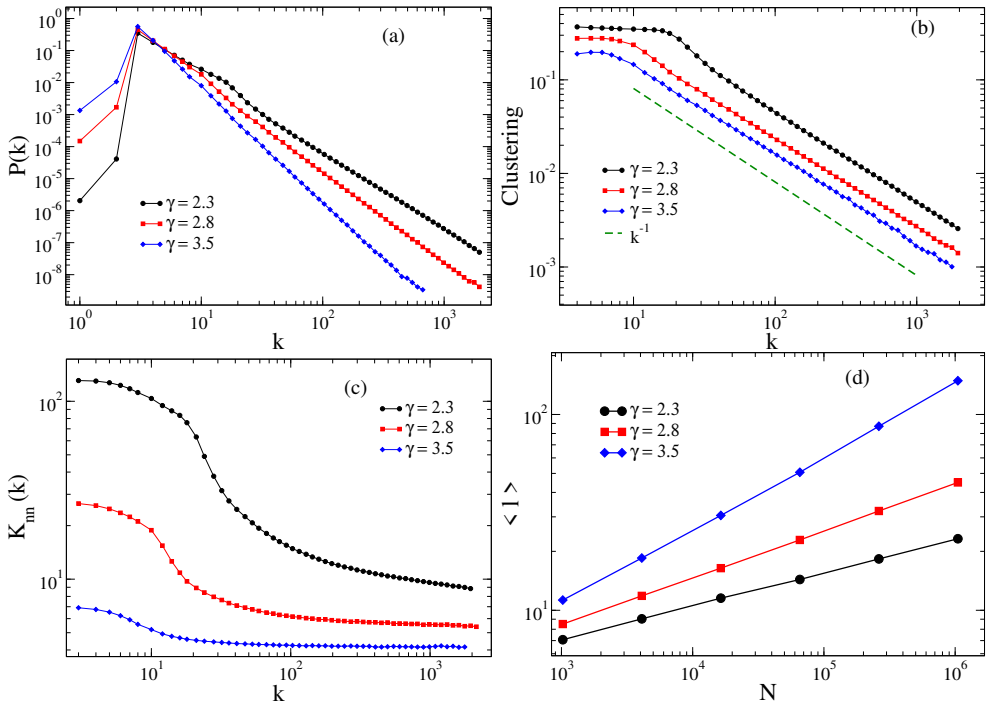


Figure 8.1: Structural properties of scale-free networks embedded on square lattices. (a) Degree distribution, (b) Clustering coefficient, (c) Average nearest neighbor degree, and (d) Average shortest path.

Appendix A

Algorithms

A.1 Rejection method optimization for RAT

To increase the performance of RAT during the selection of reactivated nodes, we divide the vertices into Q groups $\mathcal{G}_g$, $g = 1, \ldots, Q$ according to some property related to its activity. Thus, the total activity time of each group, given by

$$T_g = \sum_{j \in \mathcal{G}_g} \left(T_j^{(\mathrm{a})} \right), \tag{A.1}$$

is computed and one of the groups is selected with probability

$$P_g = \frac{T_g}{\sum_{g\prime=1}^{Q} T_{g\prime}}. \tag{A.2}$$

Finally, a randomly selected vertex $i \in \mathcal{G}_g$ is accepted with probability

$$q_i = \frac{T_i^{\mathrm{a}}}{\max_{j \in \mathcal{G}_g} \{ T_j^{\mathrm{a}} \}}. \tag{A.3}$$

The number of groups and criteria to divide them should be adapted to the dynamical process and network under investigation.

The vertex degree is a natural choice for the SIS model. So, in the RTA simulations, vertices were labeled such that $k_1 \leq k_2 \leq, \cdots, \leq k_N$ form groups of equal size $M =$

N/Q given by

$$\mathcal{G}_g = \{k_i | (g-1)M + 1 \leq i \leq gM\},\ g = 1, \cdots, Q. \tag{A.4}$$

In the case of multiplex networks, the binning can be done separately for each layer. This approach will reduce the number of rejections by gathering vertices of similar degree in the same bin. Except for the RRNS with a hub, in which we consider the single outlier separately, $Q = 100$ was adopted for all simulations in this paper.

A.2 Optmized Gillespie algorithm for multiplex networks

Let us consider a network with $\alpha = 1, 2, \cdots, m$ layers; each layer connected to all others. The OGA for SIS dynamics on multilayer networks runs as follows. With probability

$$p = \frac{\mu N_{\text{inf}}}{(m-1)\eta N_{\text{inf}} + \mu N_{\text{inf}} + \lambda N_{\text{e}}}, \tag{A.5}$$

an infected vertex selected at random is healed. With probability

$$q = \frac{(m-1)\eta N_{\text{inf}}}{(m-1)\eta N_{\text{inf}} + \mu N_{\text{inf}} + \lambda N_{\text{e}}}, \tag{A.6}$$

an infected vertex is randomly chosen and activates itself in another layer, also chosen at random. Finally, with probability $w = 1 - q - p$ we perform the intralayer infection process of the SIS model described for a single layer. The time increment Δt is

$$\Delta t = \frac{1}{(m-1)\eta N_{\text{inf}} + \mu N_{\text{inf}} + \lambda N_{\text{e}}}. \tag{A.7}$$

Note that a vertex active in multiple layers have to be considered accordingly such that

$$N_{\text{inf}} = \sum_{\alpha=1}^{m} \sum_{i=1}^{N} \sigma_{i\alpha} \text{ and } N_{\text{e}} = \sum_{\alpha=1}^{m} \sum_{i=1}^{N} k_{i\alpha}\sigma_{i\alpha}, \tag{A.8}$$

in which $\sigma_{i\alpha}$ and $k_{i\alpha}$ indicate, respectively, the states and degree of vertex i in the layer α.

A.3 Computational simulation of SEAUCR metapopulation model

This algorithm is based on the aforementioned Gillespie Algorithm [114], which consists in calculate the total rates of the events and define the details step by step. To simulate the SEAUCR model with metapopulations, we proceed as follows

(i) We select at random which group of events will occur.

(ii) Within the selected group, we draw the specific event.

(iii) We select the municipality in which the event will take place randomly.

(iv) Among the individuals within the selected municipality, we select their residence place.

(v) The event is implemented.

It is important to notice that in each simulation step we need to calculate several transition rates in order to emulate the probabilities from the model. These rates are discussed bellow.

Definitions

We define n_{ij}^{X} as the number of individuals in compartment X (= S, E, A, U, C, or R) within patch j and have i as residence place. Therefore, the conditions $i = j$ and $i \neq j$ indicates that the individuals are residents or nonresidents of i, respectively. By summing over all municipalities j connected to i (including itself), we define $\tilde{n}_{i}^{X}$ as the number of individuals in city i and compartment X. We can perform another summation over all municipalities i and define N_X as the number of individuals on a determined compartment X. The absence of superscript X on these variables (n_{ij}, $\tilde{n}_i$, etc.) denote a sum over all compartments. The terms **resident** and **nonresident** will be used to identify individuals that are within patch i and resides or not in i, respectively.

Group of events

We divided the simulation into two big event groups: mobility, that represent individuals traveling between patches, and epidemic events, that represent the transitions on the SEAUCR model. The total rate of mobility is

$$D_{\text{tot}} = T_R^{-1} \sum_i \tilde{n}_{ii} R_i + T_R^{-1} \sum_i (\tilde{n}_i - \tilde{n}_{ii}), \tag{A.9}$$

in which T_R is the average time that an individual stay away from residence patch. First and second terms of the r.h.s. of the equation represents individuals leaving and returning to their residence patches, respectively. The total rate of epidemic transitions is $E_{\text{tot}} = \sum_{\mathcal{T}} \Gamma_{\mathcal{T}}$, in which $\mathcal{T}$ is one of the transitions and $\Gamma_{\mathcal{T}}$ is the total rate of the respective transition $\mathcal{T}$, defined in the Table A.1.

Table A.1: Total transition rates for different types of epidemic events.

$\mathcal{T}$	$\Gamma_{\mathcal{T}}$	$\mathcal{T}$	$\Gamma_{\mathcal{T}}$
E→A	$\mu_{\text{A}} N_{\text{E}}$	A→U	$\beta_{\text{U}} N_{\text{A}}$
A→R	$\beta_{\text{R}} N_{\text{A}}$	U→R	$\alpha_{\text{R}} N_{\text{U}}$
U→C	$\sum_{i=1}^{\Omega} \alpha_{\text{C}}^{(i)} \tilde{n}_i^{\text{U}}$	C→R	$\alpha_{\text{R}} N_{\text{C}}$
S→E	$\sum_{i=1}^{\Omega} \frac{k_i \tilde{n}_i^{\text{S}}}{\tilde{n}_i} \left(\lambda_{\text{A}} \tilde{n}_i^{\text{A}} + \lambda_{\text{U}} \tilde{n}_i^{\text{U}} + b \lambda_{\text{C}} \tilde{n}_i^{\text{C}}\right)$		

In each time step, we select one mobility of epidemic event with probabilities $P_{\text{mob}} = D_{\text{tot}}/(D_{\text{tot}} + E_{\text{tot}})$ and $P_{\text{epi}} = 1 - P_{\text{mob}}$, respectively. The time is increased as a Poisson process of mean time $1/(D_{\text{tot}} + E_{\text{tot}})$, discussed in Section 3.1.1.

Mobility events

If a mobility event was chosen, we proceed as follows.

(i) Choose the patch i from where an individual will depart, proportionally to the total mobility rate of the patch, which occurs with probability

$$Q_i = \frac{\tilde{n}_{ii} R_i + (\tilde{n}_i - \tilde{n}_{ii})}{\sum_j \left[\tilde{n}_{jj} R_j + (\tilde{n}_j - \tilde{n}_{jj})\right]} \tag{A.10}$$

(ii) Choose the state X of the individual of patch i that will move with probability $P_i^X = \tilde{n}_i^X / \sum_Y \tilde{n}_i^Y$.

(iii) Select if the individual to move is a resident or nonresident of patch i with probabilities $n_{ii}^X / \tilde{n}_i^X$ and $1 - n_{ii}^X / \tilde{n}_i^X$, respectively.

(iv) For residents: If the individual is not symptomatic (S, E, A or R) the destination patch is chosen proportionally of R_{ij}. Otherwise, nothing happens. The movement is implementd as $n_{ij}^X \to n_{ij}^X + 1$ e $n_{ii}^X \to n_{ii}^X - 1$. Even if nothing happens, the simulated time is incremented by Δt.

(v) For nonresidents: one patch j is chosen proportionally to n_{ji}^X and the individual returns to j. The movement is implemented as $n_{ji}^X \to n_{ji}^X - 1$ and $n_{jj}^X \to n_{jj}^X + 1$. This movement happens even if the nonresident is in the states U or C.

Epidemic events

If a epidemic event was chosen, we proceed as follows.

(i) One of the transitions $\mathcal{T}$ is chosen proportionally to $\Gamma_{\mathcal{T}}$.

(ii) Choose the patch i in which will occur the transition $\mathcal{T}$, proportionally to $\Gamma_{\mathcal{T}}^i$.

(iii) Since individuals of several municipalities may travel to the selected patch i, we choose the residence place j of the individual proportionally to the populations n_{ji}^X and update $n_{ji}^X \to n_{ji}^X - 1$ and $n_{ji}^Y \to n_{ji}^Y + 1$. Notice that this rule is valid for both residents ($i = j$) and nonresidents ($i \neq j$).

Optimizations

During the simulations, most of the computational time was spent in mobility events of individuals S and R, due to the large amount of individuals in those compartments in the beginning and ending of the simulation, respectively. To circumvent this problem, we implement the mobility of S deterministically: At the beginning of each simulation day, we calculate the amount of susceptible and removed individuals that must move between each pair of cities i and j and implement this movement synchronously. After a half a day (0.5 d), we move back all the nonresidents that still

remain on S and R. The simulations yield the same results as the exact procedure within the statistical uncertainties.

Appendix B

Supplementary material

B.1 Chapter 5 - Continuous approximation for $\bar{\varphi}$

Starting from Eq (5.13), we may complete squares in the denominator and takes a binomial series expansion to obtain

$$\Theta(\bar{\varphi}) = \int_{k_{\min}}^{\infty} \frac{\beta_k \bar{\varphi}(\sqrt{\mu_s} + \bar{\varphi}\beta_k/\sqrt{\mu_s})}{\sqrt{\mu_s}(1+\bar{\varphi}\beta_k/\sqrt{\mu_s})^2} \sum_{n=0}^{\infty} \left[\frac{2(1-\sqrt{\mu_s})\bar{\varphi}\beta_k/\sqrt{\mu_s}}{(1+\bar{\varphi}\beta_k/\sqrt{\mu_s})^2}\right]^n P(k)dk. \tag{B.1}$$

The sum is converging since we are considering $\zeta = 2\bar{\varphi}(\beta_k/\sqrt{\mu_s})/(1+\bar{\varphi}\beta_k/\sqrt{\mu_s})^2 \leq 1/2$ for the series expansion of $(1-\zeta)^{-1}$. In addition, by considering a normalized power-law distribution $P(k) = Ak^{-\gamma}$ with $A = (\gamma-1)k_{\min}^{\gamma-1}$, the change of variable $x = \lambda k/(\sqrt{\mu_s}\langle k\rangle)$ leads to

$$\Theta(\bar{\varphi}) = \alpha \sum_{n=0}^{\infty} [2(1-\sqrt{\mu_s})]^n \bar{\varphi}^{n+1} \left[\sqrt{\mu_s}\int_{x_0}^{\infty} \frac{x^{-\gamma+n+1}}{(1+\bar{\varphi}x)^{2n+2}} + \bar{\varphi}\frac{x^{-\gamma+n+2}}{(1+\bar{\varphi}x)^{2n+2}}\right], \tag{B.2}$$

where $x_0 = \lambda k_{\min}/\sqrt{\mu_s\langle k\rangle}$ and $\alpha = (\gamma-1)\left(k_{\min}\lambda/\sqrt{\mu_s}\langle k\rangle\right)^{\gamma-1}$. The integrals in Eq. (B.2) can be expressed in terms of Gauss hypergeometric functions $F(a,b,c;x)$ [181], to obtain

$$\begin{aligned}\Theta(\bar{\varphi}) =& \alpha x_0^{-\gamma+1} \sum_{n=0}^{\infty} [2(1-\sqrt{\mu_s})]^n (\bar{\varphi}x_0)^{-(n+1)} \left[\frac{\sqrt{\mu_s}}{n+\gamma} F\left(2n+2, n+\gamma, n+\gamma+1; \frac{-1}{\bar{\varphi}x_0}\right)\right. \\ &\left. + \frac{\bar{\varphi}x_0}{n+\gamma-1} F\left(2n+2, n+\gamma-1, n+\gamma; \frac{-1}{\bar{\varphi}x_0}\right)\right].\end{aligned} \tag{B.3}$$

Finally, we take the asymptotic expansion for $F(a,b,c;z)$ for $z \to \infty$ up order $\bar{\varphi}^3$ to obtain

$$\Theta(\bar{\varphi}) = \lambda\bar{\varphi} + a_{\gamma-1}\bar{\varphi}^{\gamma-1} + \frac{(\gamma-2)^2(1-2\mu_s)}{(\gamma-3)(\gamma-1)\mu_s}\lambda^2\bar{\varphi}^2 + \frac{(\gamma-2)^3(7-8\sqrt{\mu_s})}{(4-\gamma)(\gamma-1)^2\mu_s}\lambda^3\bar{\varphi}^3 + \cdots, \tag{B.4}$$

where

$$a_{\gamma-1} = (\gamma-1)\left(\frac{\lambda k_{\min}}{\langle k\rangle\sqrt{\mu_s}}\right)^{\gamma-1}\tilde{\beta} \tag{B.5}$$

and

$$\tilde{\beta} = \sum_{n=0}^{\infty}[2(1-\sqrt{\mu_s})]^n\left[\sqrt{\mu_s}+\frac{n-\gamma+2}{n+\gamma-1}\right]\frac{\Gamma(n-\gamma+2)\Gamma(\gamma+n)}{\Gamma(2n+2)}, \tag{B.6}$$

where $\Gamma(x)$ is the Gamma function.

B.2 Chapter 6

Videos showing the daily evolution of the symptomatic prevalence of all Brazilian municipalities can be found at the link `https://arxiv.org/src/2011.03380v2/anc`. Simulation dates are shown in the upper right corner for all images. Darker colors represent higher prevalence, on a logarithmic scale

- No mitigation scenario: Video_SI-1_none.mp4
- Weak mitigation scenario: Video_SI-1_weak.mp4
- Moderate mitigation scenario: Video_SI-1_moderate.mp4

Similar videos were hosted on Youtube and can be found in the links bellow (online only):

- No mitigation scenario
- Weak mitigation scenario
- Moderate mitigation scenario

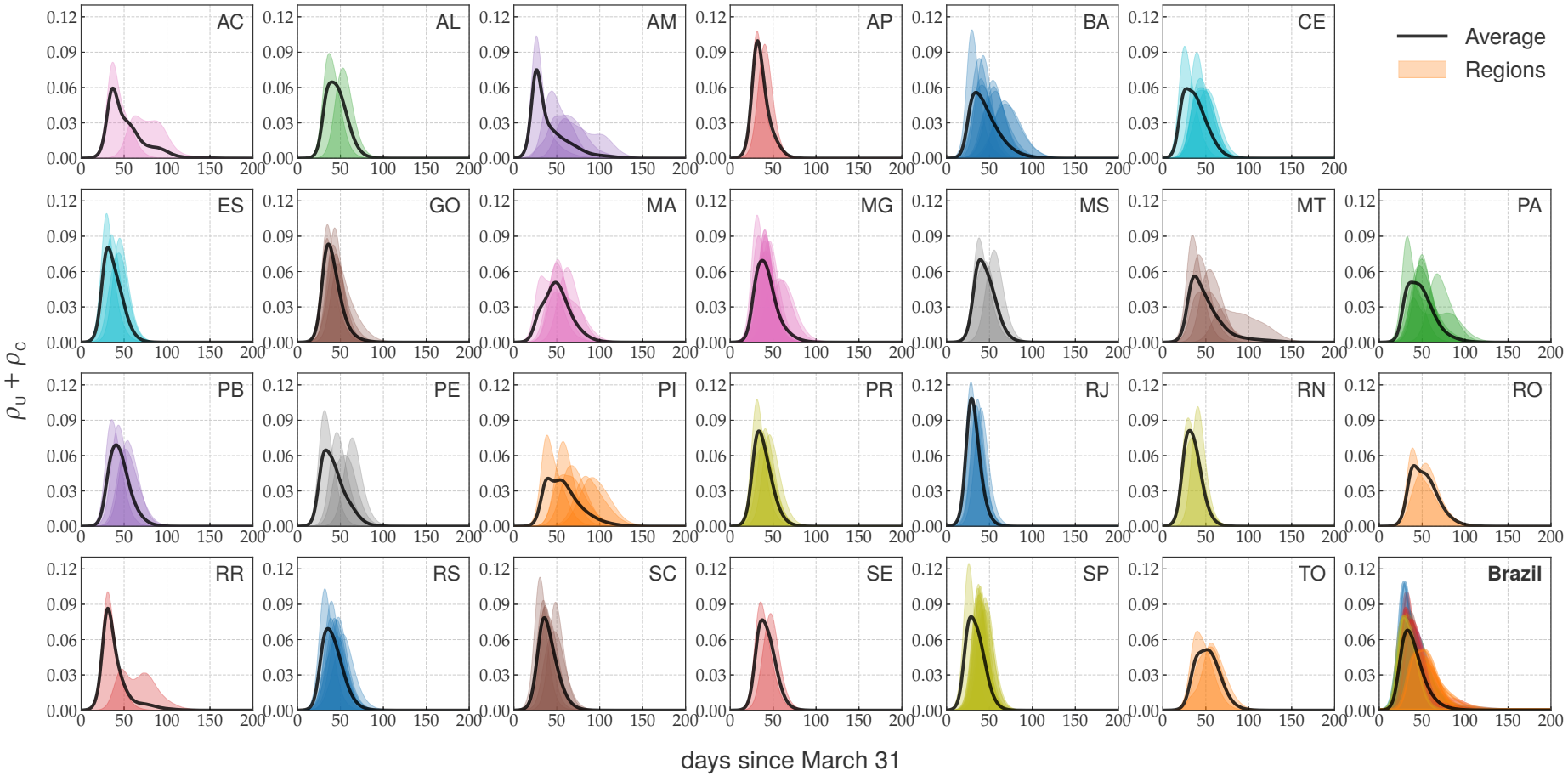


Figure B.1: Epidemic prevalence of symptomatic individuals for all intermediate regions of Brazil grouped by federative state without mitigation obtained with parameters $(M, K) = (0, 0)$. The shaded plots represent the intermediate regions while the thick lines averages within each state. The bottom right corner plot represents an average over the whole country and shaded plots represent federative states.

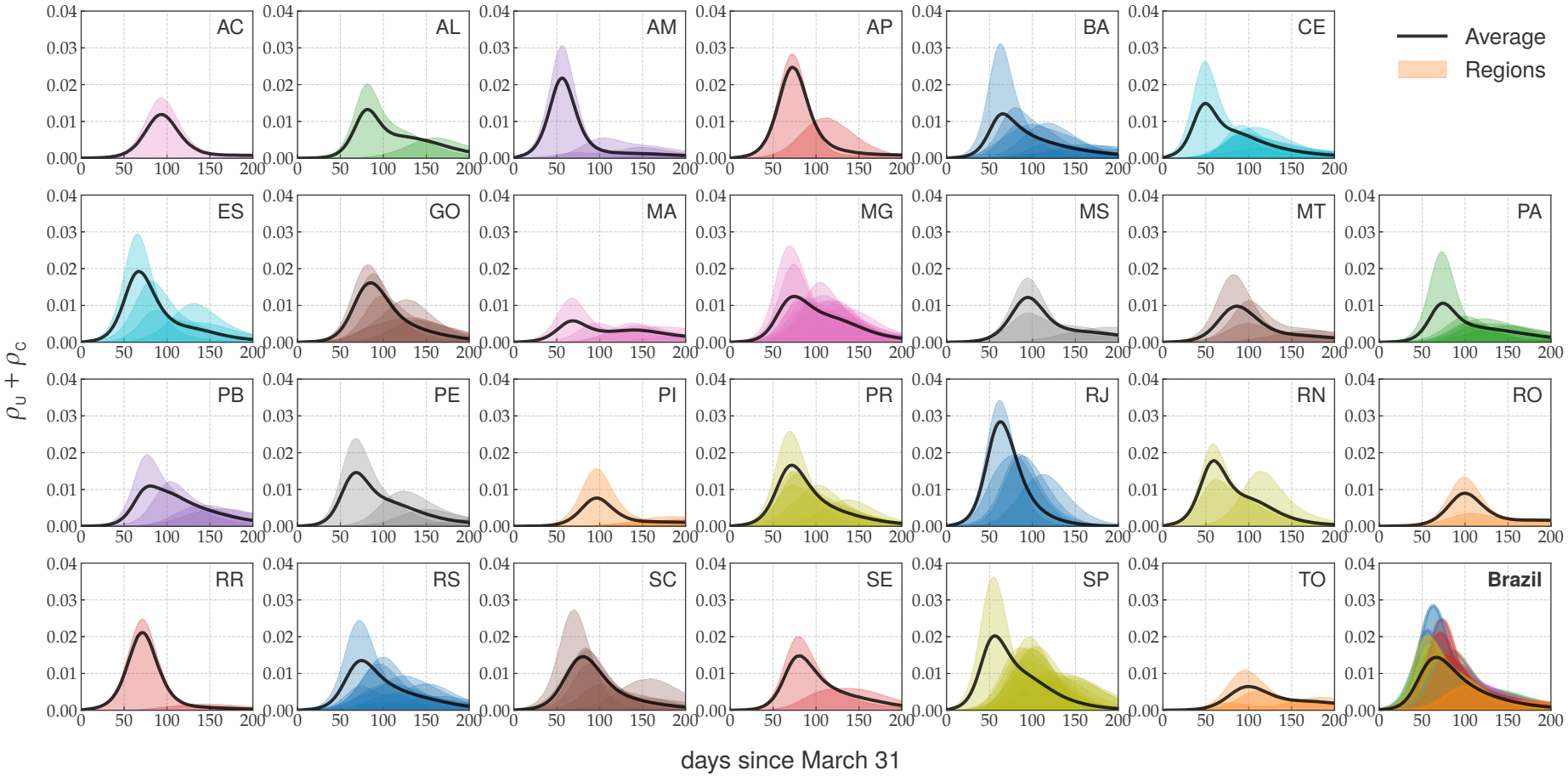


Figure B.2: Epidemic prevalence of symptomatic individuals for all intermediate regions of Brazil grouped by federative state obtained with moderate mitigation parameters $(M, K) = (0.8, 0.5)$. The shaded plots represent the intermediate regions while the thick lines averages within each state. The bottom right corner plot represents an average over the whole country and shaded plots represent federative states.

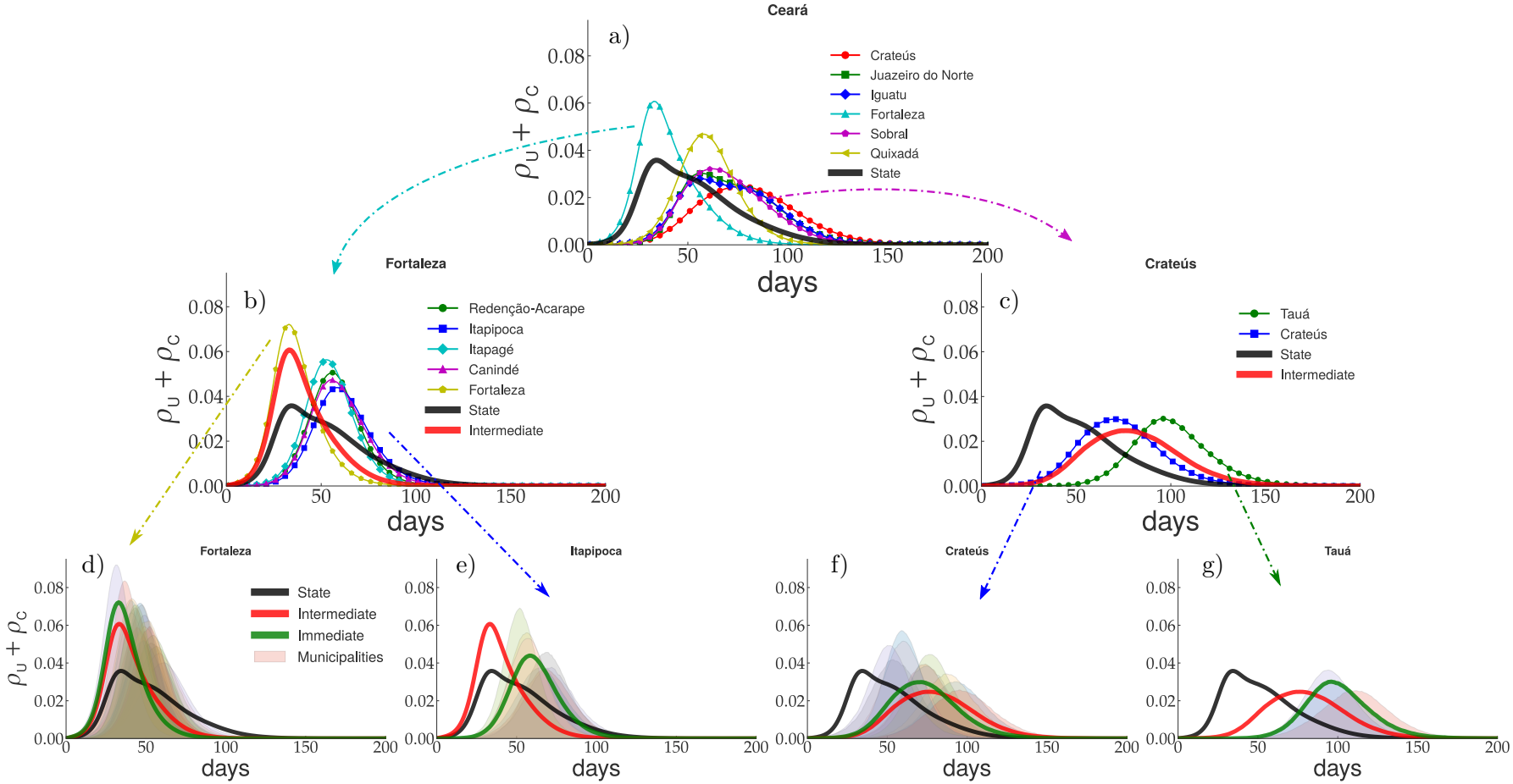


Figure B.3: Multi-scale analysis of the epidemic prevalence of symptomatic individuals for the CE state at several scales of geographical organization considering a weak mitigation with parameters $(M, K) = (0.4, 0.3)$. Epidemic curves averaged with geographical resolution increasing from top to bottom are compared with lower resolution averages: a) intermediate regions, b,c) immediate regions, d-g) municipalities. The curves presenting the earliest and latest maxima are chosen as representative within each panel. Arrows indicate curves selected for zooming. Day 0 represents 31 March 2020.

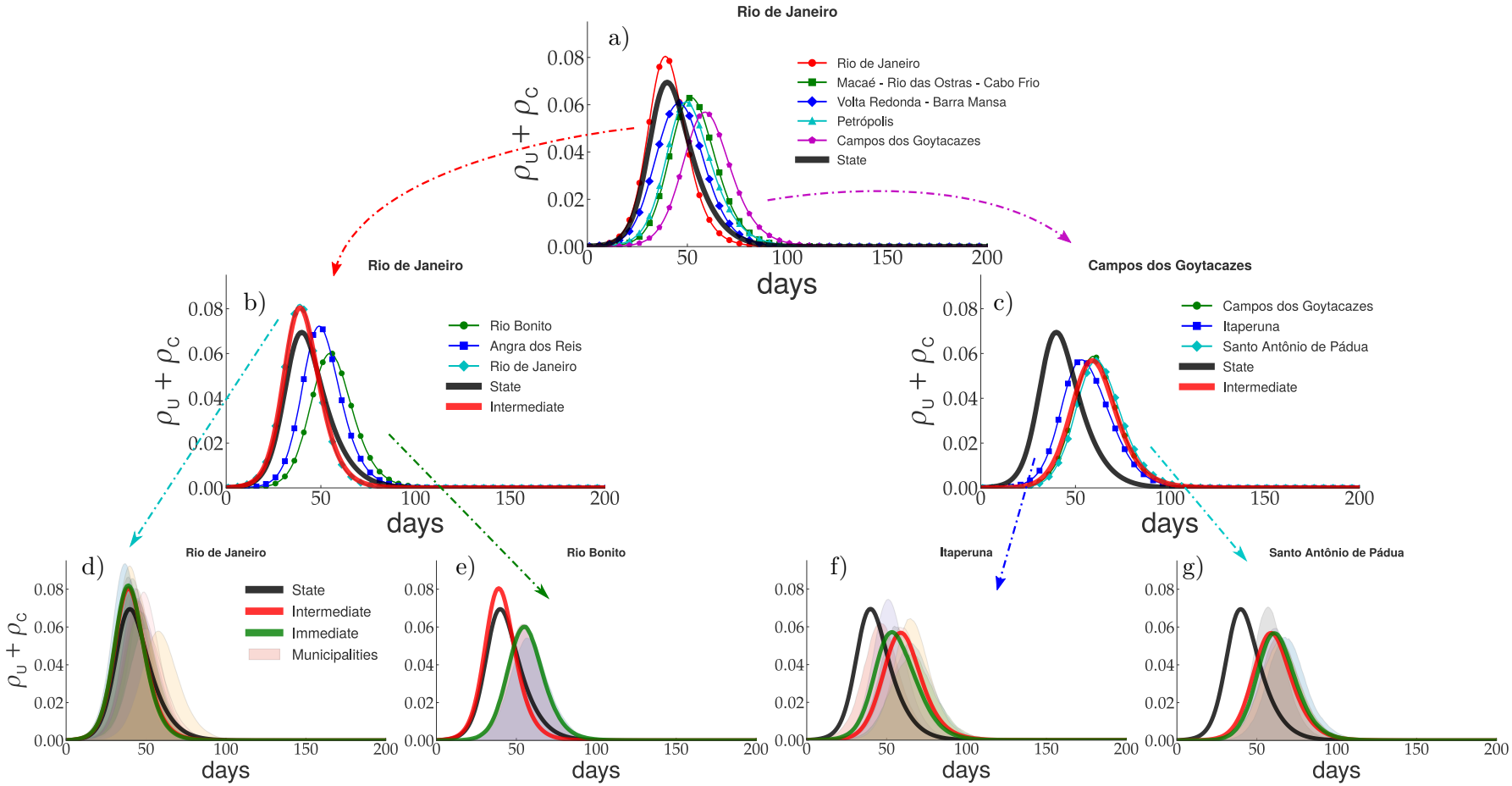


Figure B.4: Multi-scale analysis of the epidemic prevalence of symptomatic individuals for the RJ state at several scales of geographical organization considering a weak mitigation with parameters $(M, K) = (0.4, 0.3)$. Epidemic curves averaged with geographical resolution increasing from top to bottom are compared with lower resolution averages: a) intermediate regions, b,c) immediate regions, d-g) municipalities. The curves presenting the earliest and latest maxima are chosen as representative within each panel. Arrows indicate curves selected for zooming. Day 0 represents 31 March 2020.

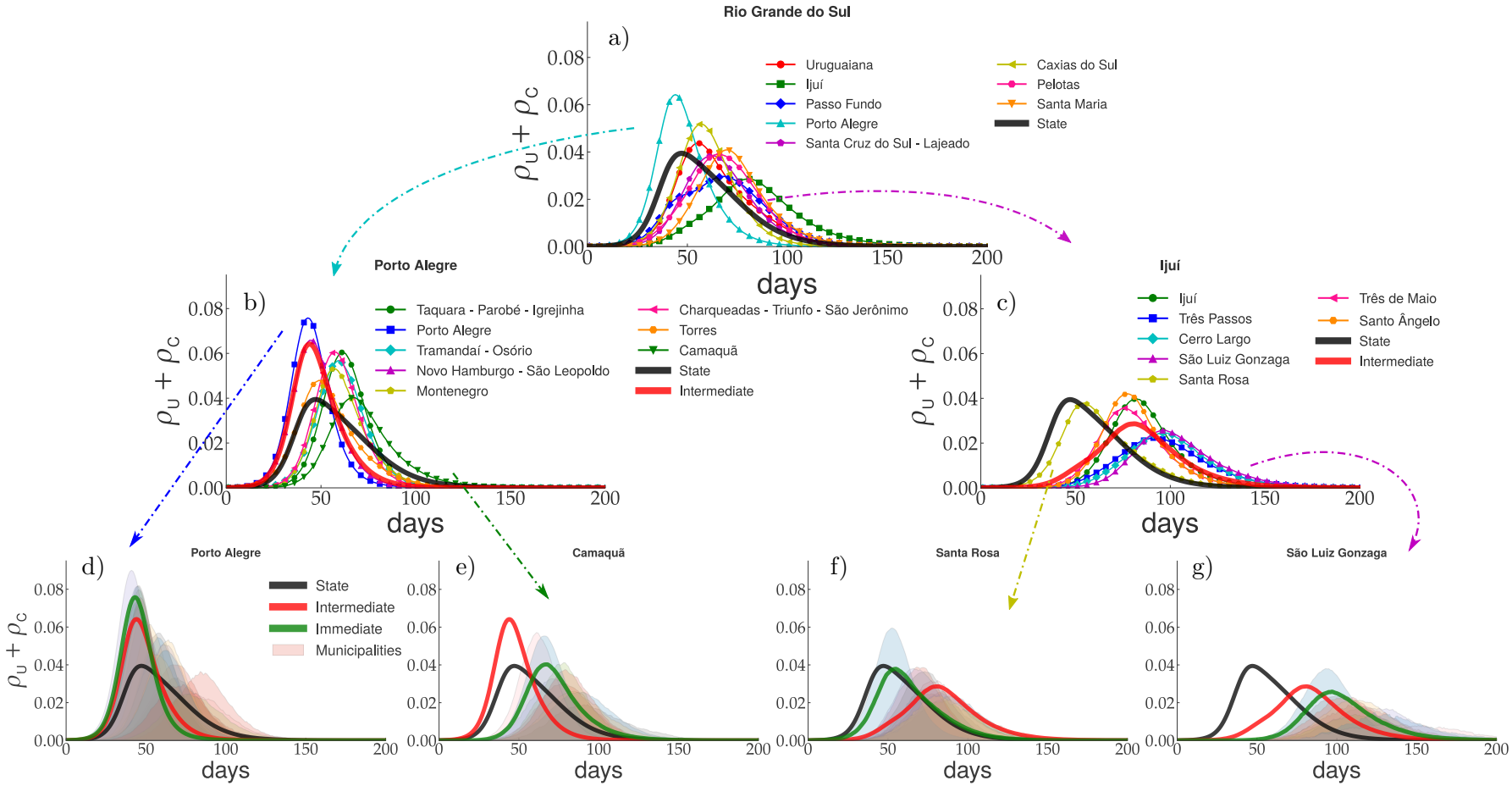


Figure B.5: Multi-scale analysis of the epidemic prevalence of symptomatic individuals for the RS state at several scales of geographical organization considering a weak mitigation with parameters $(M, K) = (0.4, 0.3)$. Epidemic curves averaged with geographical resolution increasing from top to bottom are compared with lower resolution averages: a) intermediate regions, b,c) immediate regions, d-g) municipalities. The curves presenting the earliest and latest maxima are chosen as representative within each panel. Arrows indicate curves selected for zooming. Day 0 represents 31 March 2020.

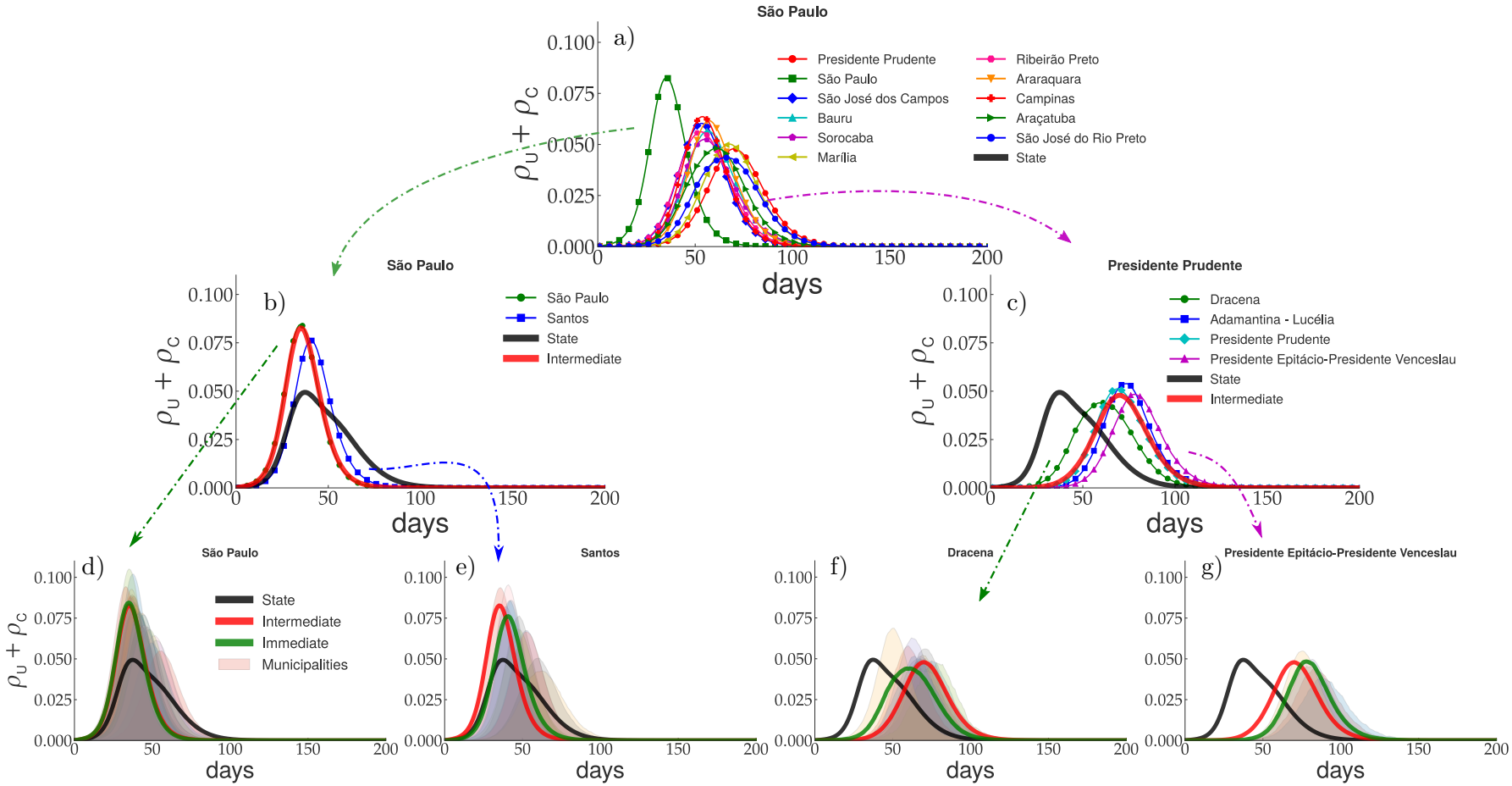


Figure B.6: Multi-scale analysis of the epidemic prevalence of symptomatic individuals for the SP state at several scales of geographical organization considering a weak mitigation with parameters $(M, K) = (0.4, 0.3)$. Epidemic curves averaged with geographical resolution increasing from top to bottom are compared with lower resolution averages: a) intermediate regions, b,c) immediate regions, d-g) municipalities. The curves presenting the earliest and latest maxima are chosen as representative within each panel. Arrows indicate curves selected for zooming. Day 0 represents 31 March 2020.

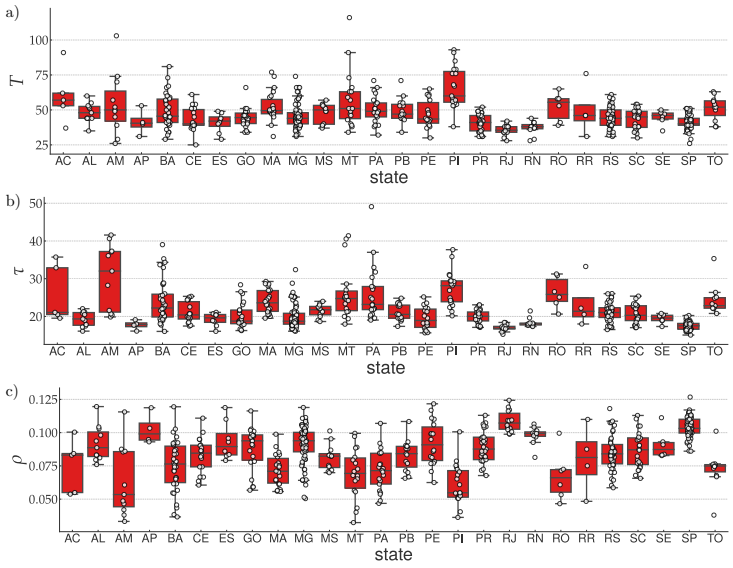


Figure B.7: Box plots for the peak a) time T, b) width τ, and c) prevalence ρ averaged over immediate regions for the 26 federation states without mitigation with parameters $(M, K) = (0, 0)$. Circles are data for different regions. As usual, boxes yield median, lower and upper quartiles while points outside whiskers are outliers.

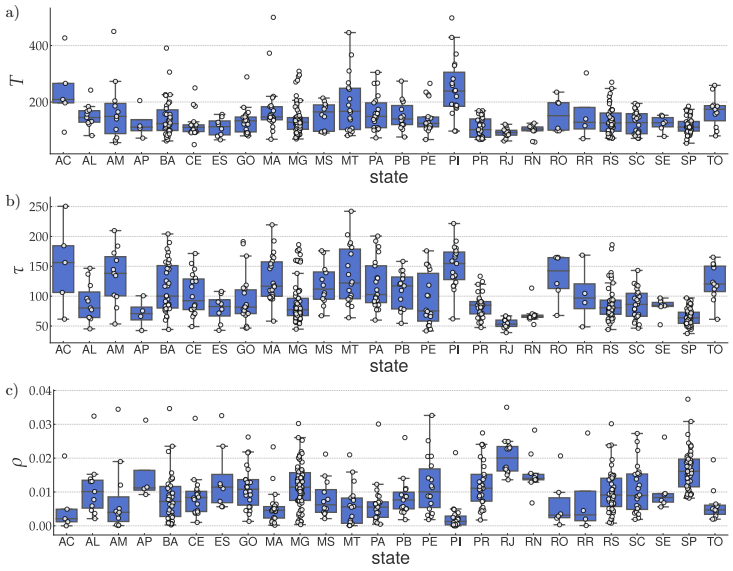


Figure B.8: Box plots for the peak a) time T, b) width τ, and c) prevalence ρ averaged over immediate regions for the 26 federation states with moderate mitigation using parameters $(M, K) = (0.8, 0.5)$. Circles are data for different regions. As usual, boxes yield median, lower and upper quartiles while points outside whiskers are outliers.

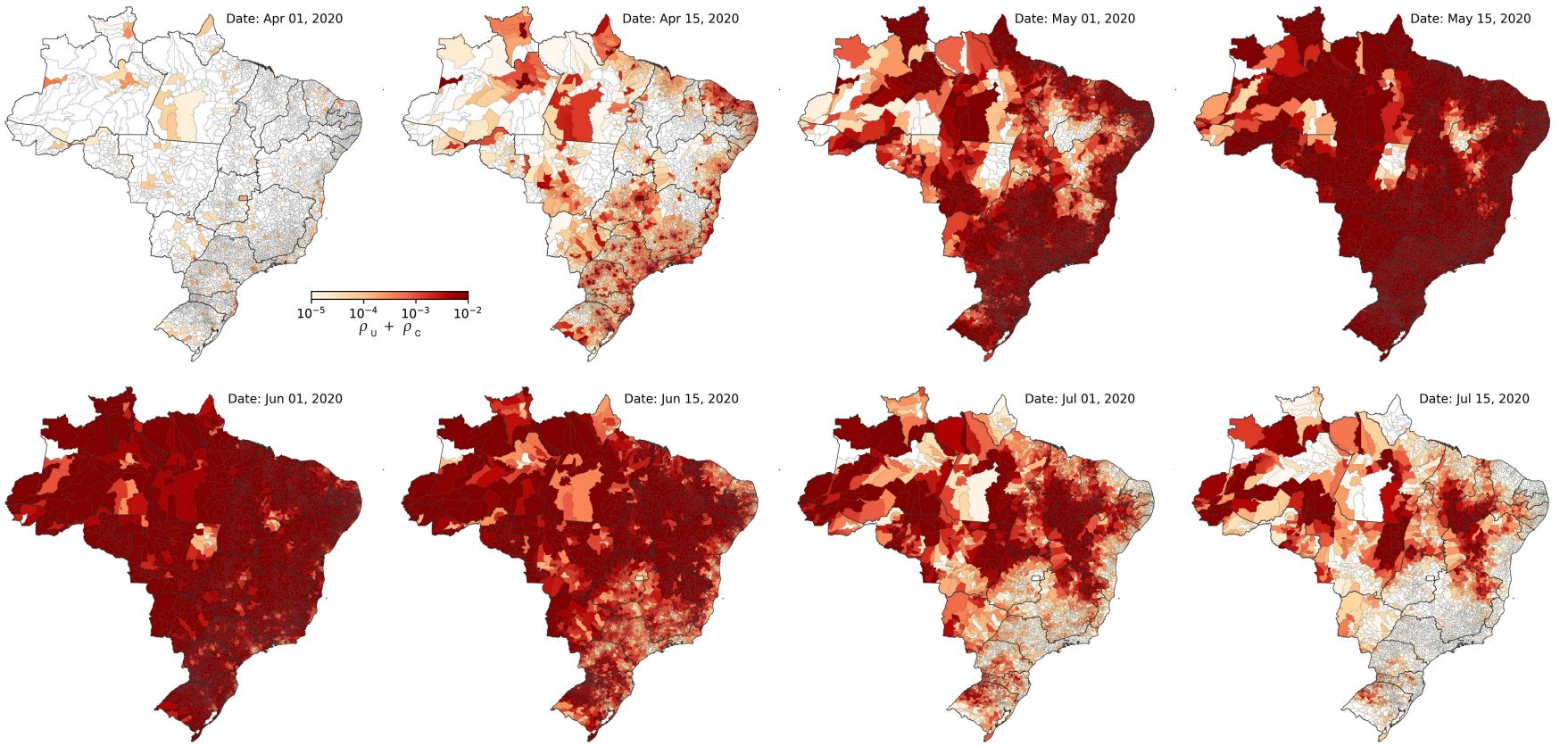


Figure B.9: Color maps with the epidemic prevalence of symptomatic cases (U and C) on Brazil for simulations without mitigation with parameters $(M, K) = (0, 0)$. Dates of the simulations are shown in the upper right corner for all frames. The darker colors represent higher prevalences in a logarithm scale.

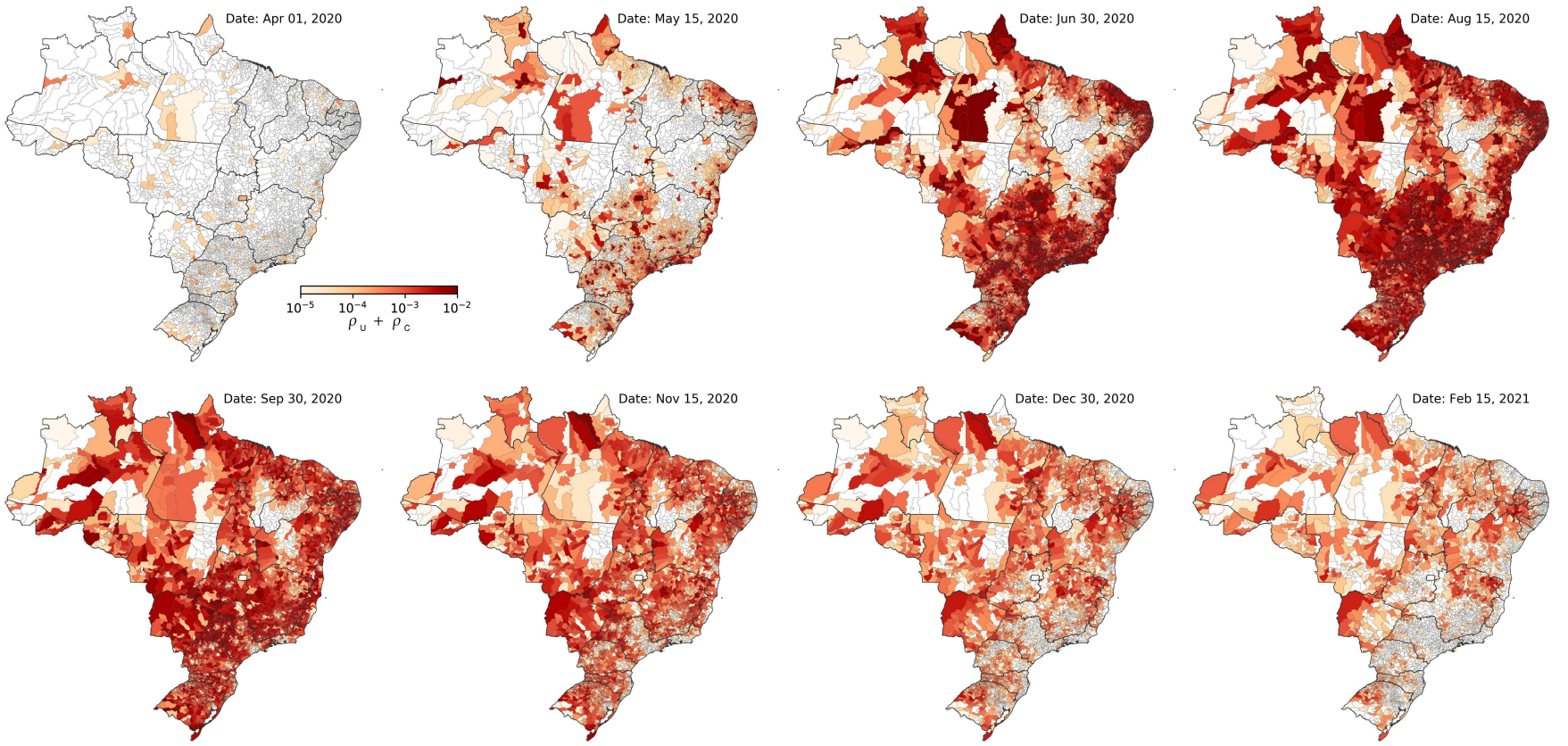


Figure B.10: Color maps with prevalence of symptomatic cases (U and C) on Brazil for a moderate mitigation simulation with parameters $(M, K) = (0.8, 0.5)$. Dates of the simulations are shown in the upper right corner for all frames. The darker colors represent higher prevalences in a logarithm scale.

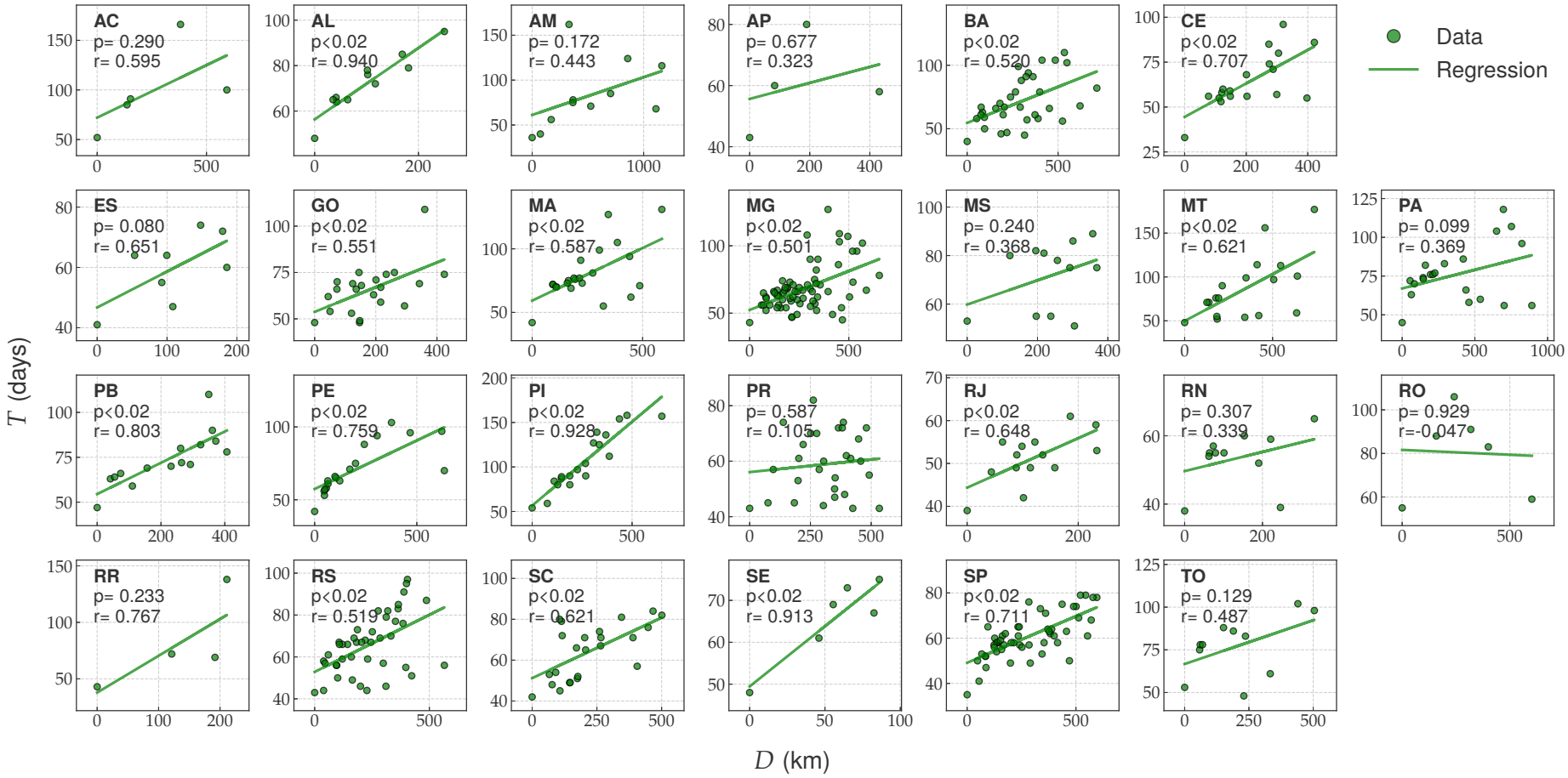


Figure B.11: Scatter plots of days for epidemic peak versus distance from capital city for each federative state of Brazil considering a weak mitigation scenario with parameters $(M, K) = (0.4, 0.3)$. Full lines represent linear regressions o the data. Both p-values and Pearson coefficients obtained in the statistical analyses of linear correlations are shown in each panel.

B.3 Chapter 7

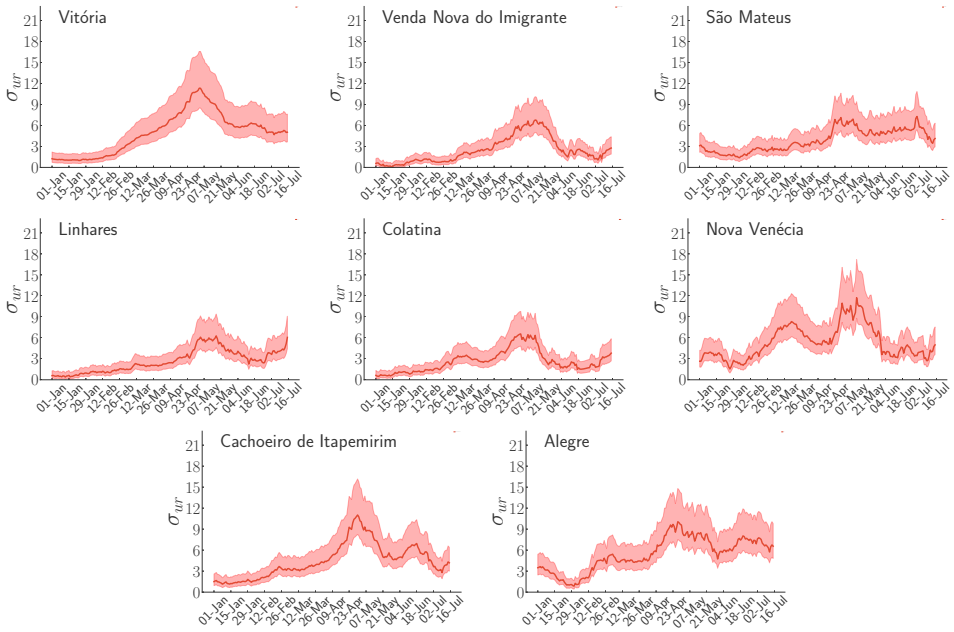


Figure B.12: Evolution of the under-reporting coefficient for immediate regions of ES state 3-week time windows. The regions are indicated in the upper-left corners for each panel.

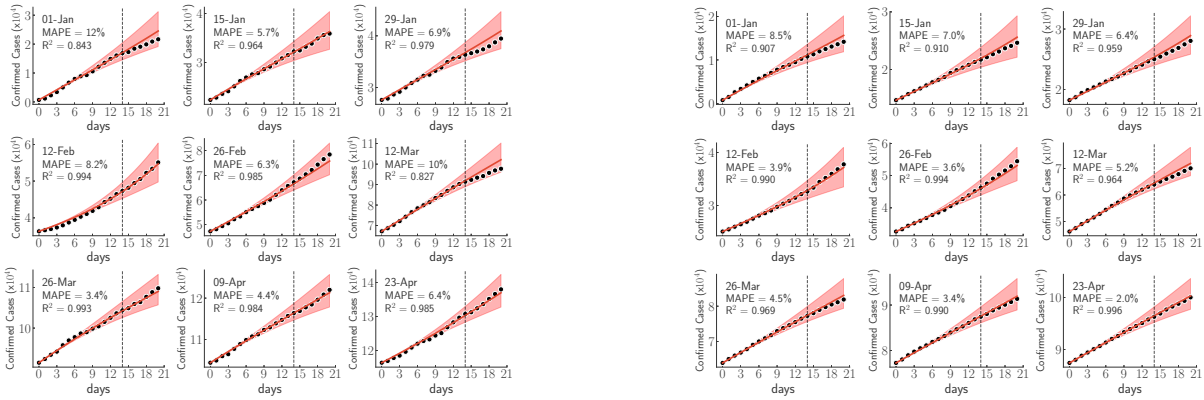


Figure B.13: Calibration curves for the immediate regions of (left) Curitiba and (right) Vitória in different time windows of 14 days. The initial day is indicated in the top of each panel. Symbols are the cumulative diagnosed cases while lines with shaded regions represent the calibrated curves and the corresponding confidence interval.

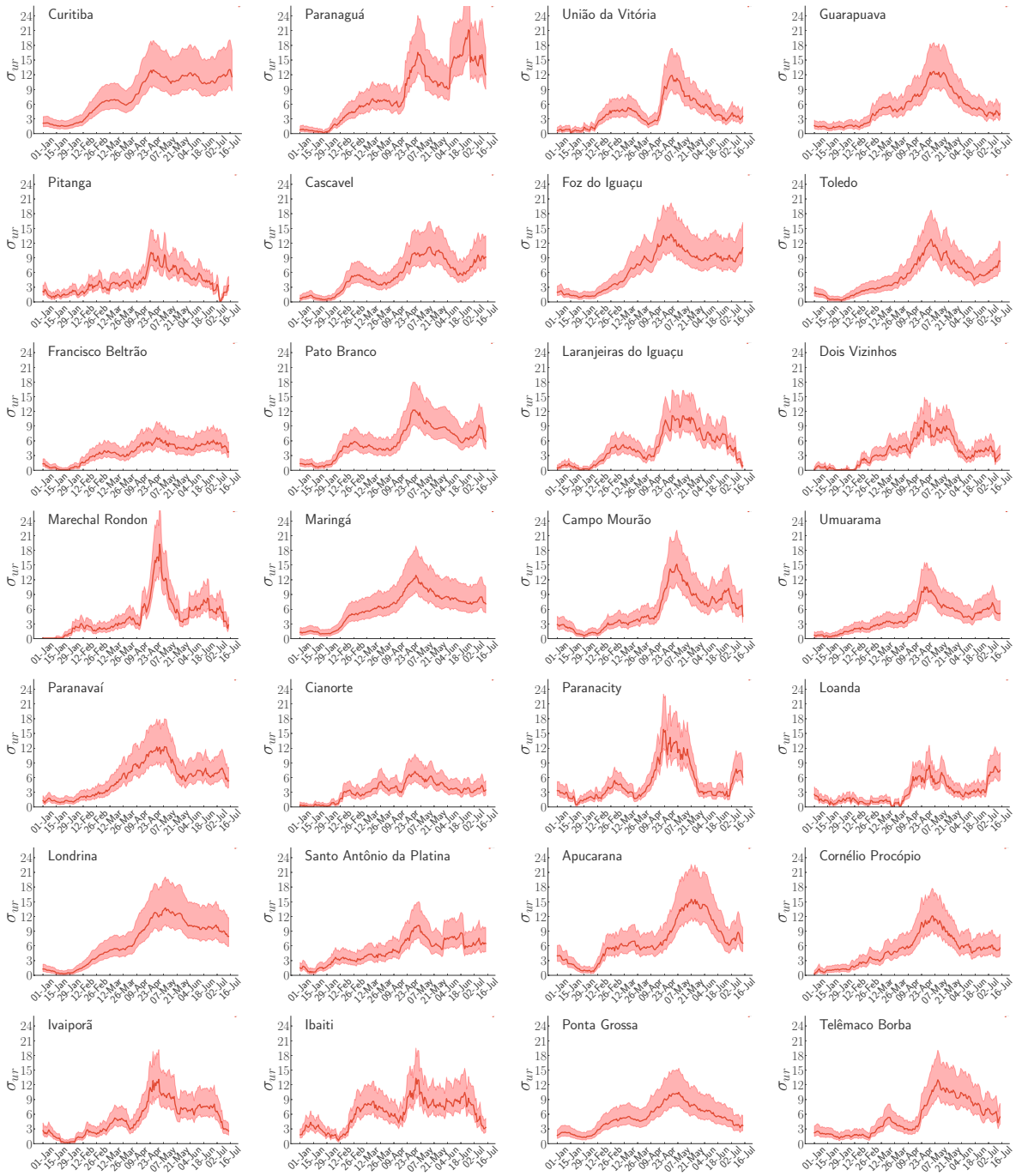


Figure B.14: Evolution of the under-reporting coefficient for immediate regions of PR state using 3-week time windows. The regions are indicated in the upper-left corners for each panel.

Bibliography

[1] I. Gribkovskaia, Ø. Halskau Sr, and G. Laporte, "The bridges of königsberg—a historical perspective," *Networks: An International Journal*, vol. 49, no. 3, pp. 199–203, 2007.

[2] A.-L. Barabási and M. Pósfai, *Network science*. Cambridge, UK: Cambridge University Press, 2016.

[3] H. Maia, S. Ferreira, and M. Martins, "Adaptive network approach for emergence of societal bubbles," *Physica A: Statistical Mechanics and its Applications*, vol. 572, p. 125588, 2021.

[4] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, "Quantifying echo chamber effects in information spreading over political communication networks," *EPJ Data Sci.*, vol. 8, p. 35, 2019.

[5] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, "Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results," *Phys. Rev. E*, vol. 86, p. 041125, Oct 2012.

[6] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proc. Natl. Acad. Sci.*, vol. 106, pp. 21484–21489, 2009.

[7] R. Pastor-Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Phys. Rev. Lett.*, vol. 86, pp. 3200–3203, 2001.

[8] G. Chechik, E. Oh, O. Rando, J. Weissman, A. Regev, and D. Koller, "Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network," *Nature Biotechnology*, vol. 26, pp. 1251–1259, Nov 2008.

[9] M. A. Nowak and R. M. May, "Evolutionary games and spatial chaos," *Nature*, vol. 359, pp. 826–829, 1992.

[10] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Phys. Rev. A*, vol. 38, pp. 364–374, 1988.

[11] N. Goldenfeld and L. P. Kadanoff, "Simple lessons from complexity," *Science*, vol. 284, pp. 87–89, 1999.

[12] A. L. Barabási, "Taming complexity," *Nat. Phys.*, vol. 1, pp. 68–70, nov 2005.

[13] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, p. 016128, 2002.

[14] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys.*, vol. 87, pp. 925–979, 2015.

[15] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics," *Proc. Natl. Acad. Sci.*, vol. 103, pp. 2015–2020, 2006.

[16] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, 2021.

[17] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge,UK: Cambridge University Press, 1st ed., 2008.

[18] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani, "Seasonal transmission potential and activity peaks of the new influenza a(h1n1): a monte carlo likelihood analysis based on human mobility," *BMC Medicine*, vol. 7, p. 45, Sep 2009.

[19] M. F. C. Gomes, A. Pastore Y Piontti, L. Rossi, D. Chao, I. Longini, M. E. Halloran, and A. Vespignani, "Assessing the international spreading risk associated with the 2014 west african ebola outbreak," *PLoS currents*, vol. 6, Sep 2014.

[20] D. A. M. Villela, L. S. Bastos, L. M. de Carvalho, O. G. Cruz, M. F. C. Gomes, B. Durovni, M. C. Lemos, V. Saraceni, F. C. Coelho, C. T. Codeço, and et al., "Zika in rio de janeiro: Assessment of basic reproduction number and comparison with dengue outbreaks," *Epidemiology and Infection*, vol. 145, no. 8, p. 1649–1657, 2017.

[21] S. T. R. Pinho, C. P. Ferreira, L. Esteva, F. R. Barreto, V. C. Morato e Silva, and M. G. L. Teixeira, "Modelling the dynamics of dengue real epidemics," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1933, pp. 5679–5693, 2010.

[22] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track covid-19 in real time," *The Lancet Infectious Diseases*, vol. 20, pp. 533–534, May 2020.

[23] W. Cota, "Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level," *SciELOPreprints:362*, May 2020.

[24] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia," *N. Engl. J. Med.*, vol. 382, pp. 1199–1207, 2020.

[25] J. M. Read, J. R. E. Bridgen, D. A. T. Cummings, A. Ho, and C. P. Jewell, "Novel coronavirus 2019-nCoV (COVID-19): early estimation of epidemiological parameters and epidemic size estimates," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 376, p. 20200265, jul 2021.

[26] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, pp. 395–400, 2020.

[27] Z. Du, L. Wang, S. Cauchemez, X. Xu, X. Wang, B. J. Cowling, and L. A. Meyers, "Risk for Transportation of Coronavirus Disease from Wuhan to Other Cities in China," *Emerg. Infect. Dis.*, vol. 26, no. 5, pp. 1049–1052, 2020.

[28] L. Danon, E. Brooks-Pollock, M. Bailey, and M. Keeling, "A spatial model of covid-19 transmission in england and wales: early spread, peak timing and the impact of seasonality," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, no. 1829, p. 20200272, 2021.

[29] A. Arenas, W. Cota, J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños, and B. Steinegger, "Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions," *Phys. Rev. X*, vol. 10, p. 041055, Dec 2020.

[30] A. Aleta and Y. Moreno, "Evaluation of the potential incidence of COVID-19 and effectiveness of containment measures in spain: a data-driven approach," *BMC Med.*, vol. 18, p. 157, dec 2020.

[31] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*. Cambridge,UK: Cambridge University Press, 1999.

[32] M. Henkel, M. Pleimling, H. Hinrichsen, and S. Lübeck, *Non-Equilibrium Phase Transitions*, vol. 2. Dordrecht: Springer Netherlands, 2008.

[33] M. Bellac, L. Bellac/Mortessagne/Bat, M. Bellac, F. Mortessagne, G. Batrouni, G. Batrouni, G. George, and C. U. Press, *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge University Press, 2004.

[34] M. M. de Oliveira and R. Dickman, "How to simulate the quasistationary state," *Phys. Rev. E*, vol. 71, p. 016129, 2005.

[35] H. K. Lee, P.-S. Shim, and J. D. Noh, "Epidemic threshold of the susceptible-infected-susceptible model on complex networks," *Phys. Rev. E*, vol. 87, p. 062812, 2013.

[36] G. Pruessner, "Equivalence of conditional and external field ensembles in absorbing-state phase transitions," *Phys. Rev. E*, vol. 76, p. 061103, 2007.

[37] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control*. Oxford: Oxford Univ. Press, 1991. OCLC: 844937291.

[38] A. S. Mata and S. C. Ferreira, "Multiple transitions of the susceptible-infected-susceptible epidemic model on complex networks," *Phys. Rev. E*, vol. 91, p. 012816, 2015.

[39] W. Cota, S. C. Ferreira, and G. Ódor, "Griffiths effects of the susceptible-infected-susceptible epidemic model on random power-law networks," *Phys. Rev. E*, vol. 93, p. 032322, 2016.

[40] W. Cota, G. Ódor, and S. C. Ferreira, "Griffiths phases in infinite-dimensional, non-hierarchical modular networks," *Sci. Rep.*, vol. 8, p. 9144, 2018.

[41] S. Chatterjee and R. Durrett, "Contact processes on random graphs with power law degree distributions have critical value 0," *Ann. Probab.*, vol. 37, pp. 2332–2356, 2009.

[42] T. E. Harris, "Contact Interactions on a Lattice," *The Annals of Probability*, vol. 2, no. 6, pp. 969 – 988, 1974.

[43] A. G. Moreira and R. Dickman, "Critical dynamics of the contact process with quenched disorder," *Phys. Rev. E*, vol. 54, pp. R3090–R3093, Oct 1996.

[44] M. M. de Oliveira, S. G. Alves, S. C. Ferreira, and R. Dickman, "Contact process on a voronoi triangulation," *Phys. Rev. E*, vol. 78, p. 031133, Sep 2008.

[45] T. Vojta and M. Y. Lee, "Nonequilibrium phase transition on a randomly diluted lattice," *Phys. Rev. Lett.*, vol. 96, p. 035701, Jan 2006.

[46] S. C. Ferreira, R. S. Ferreira, C. Castellano, and R. Pastor-Satorras, "Quasistationary simulations of the contact process on quenched networks," *Phys. Rev. E*, vol. 84, p. 066102, Dec 2011.

[47] S. C. Ferreira, R. S. Ferreira, and R. Pastor-Satorras, "Quasistationary analysis of the contact process on annealed scale-free networks," *Phys. Rev. E*, vol. 83, p. 066113, Jun 2011.

[48] D.-J. Liu, X. Guo, and J. W. Evans, "Quadratic contact process: Phase separation with interface-orientation-dependent equistability," *Phys. Rev. Lett.*, vol. 98, p. 050601, Feb 2007.

[49] D. Centola and M. Macy, "Complex contagions and the weakness of long ties," *American Journal of Sociology*, vol. 113, no. 3, pp. 702–734, 2007.

[50] S. Iwata, K. Kobayashi, S. Higa, J. Yoshimura, and K. ichi Tainaka, "A simple population theory for mutualism by the use of lattice gas model," *Ecological Modelling*, vol. 222, no. 13, pp. 2042 – 2048, 2011.

[51] U. Dobramysl and U. C. Täuber, "Environmental versus demographic variability in two-species predator-prey models," *Phys. Rev. Lett.*, vol. 110, p. 048105, Jan 2013.

[52] M. M. de Oliveira, R. V. Dos Santos, and R. Dickman, "Symbiotic two-species contact process," *Phys. Rev. E*, vol. 86, p. 011121, Jul 2012.

[53] M. M. de Oliveira and R. Dickman, "Phase diagram of the symbiotic two-species contact process," *Phys. Rev. E*, vol. 90, p. 032120, Sep 2014.

[54] M. M. de Oliveira, S. G. Alves, and S. C. Ferreira, "Dynamical correlations and pairwise theory for the symbiotic contact process on networks," *Phys. Rev. E*, vol. 100, p. 052302, Nov 2019.

[55] C. I. N. Sampaio Filho, T. B. dos Santos, N. A. M. Araújo, H. A. Carmona, A. A. Moreira, and J. S. Andrade, "Symbiotic contact process: Phase transitions, hysteresis cycles, and bistability," *Phys. Rev. E*, vol. 98, p. 062108, Dec 2018.

[56] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in wuhan, china: a modelling study," *Lancet*, vol. 395, pp. 689–697, feb 2020.

[57] G. Pullano, F. Pinotti, E. Valdano, P.-Y. Boelle, C. Poletto, and V. Colizza, "Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020," *Eurosurveillance*, vol. 25, 2020.

[58] M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, P.-Y. Boelle, E. D'Ortenzio, Y. Yazdanpanah, S. P. Eholie, M. Altmann, B. Gutierrez, M. U. G. Kraemer, and V. Colizza, "Preparedness and vulnerability of african countries against importations of COVID-19: a modelling study," *Lancet*, vol. 395, pp. 871–877, mar 2020.

[59] C. Zhan, C. Tse, Y. Fu, Z. Lai, and H. Zhang, "Modelling and Prediction of the 2019 Coronavirus Disease Spreading in China Incorporating Human Migration Data," *SSRN Electron. J.*, pp. 1–13, 2020.

[60] M. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[61] J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas, "Critical regimes driven by recurrent mobility patterns of reaction-diffusion processes in networks," *Nat. Phys.*, vol. 14, no. 4, pp. 391–395, 2018.

[62] Instituto Brasileiro de Geografia e Estatística (IBGE), "Censo demográfico 2010: resultados gerais da amostra," 2012. [Online; accessed 04-May-2020].

[63] Instituto Brasileiro de Geografia e Estatística (IBGE), "Estimativas da população residente para os municípios e para as unidades da federação brasileiros com data de referência em 1º de julho de 2019," tech. rep., IBGE Rio de Janeiro, 2019. [Online; accessed 04-May-2020].

[64] Agência Nacional de Aviação Civil (ANAC), "Dados estatísticos da agência nacional de aviação civil," 2019. [Online; accessed 04-May-2020].

[65] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)," *Science*, vol. 368, pp. 489–493, may 2020.

[66] J. C. Emery, T. W. Russell, Y. Liu, J. Hellewell, C. A. Pearson, K. E. Atkins, P. Klepac, A. Endo, C. I. Jarvis, N. G. Davies, E. M. Rees, S. R. Meakin, A. Rosello, K. van Zandvoort, J. D. Munday, W. J. Edmunds, T. Jombart, M. Auzenbergs, E. S. Nightingale, M. Jit, S. Abbott, D. Simons, N. I. Bosse, Q. J. Leclerc, S. R. Procter, C. J. Villabona-Arenas, D. C. Tully, A. K. Deol, F. Y.

Sun, S. Hué, A. M. Foss, K. Prem, G. Medley, A. Gimma, R. Lowe, S. Clifford, M. Quaife, C. Diamond, H. P. Gibbs, B. J. Quilty, K. OReilly, G. M. Knight, R. M. Eggo, A. J. Kucharski, S. Funk, S. Flasche, and R. M. Houben, "The contribution of asymptomatic SARS-CoV-2 infections to transmission on the diamond princess cruise ship," *eLife*, vol. 9, pp. 1–68, aug 2020.

[67] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020," *Eurosurveillance*, vol. 25, mar 2020.

[68] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M.-L. McLaws, and P. Glasziou, "Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis," *Off. J. Assoc. Med. Microbiol. Infect. Dis. Canada*, vol. 5, pp. 223–234, dec 2020.

[69] D. Buitrago-Garcia, D. Egli-Gany, M. J. Counotte, S. Hossmann, H. Imeri, A. M. Ipekci, G. Salanti, and N. Low, "Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis," *PLoS Med.*, vol. 17, pp. 1–25, sep 2020.

[70] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthélemy, "Elementary processes governing the evolution of road networks," *Scientific Reports*, vol. 2, p. 296, Mar 2012.

[71] M. J. Herrgård, N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Blüthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novère, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasić, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kirdar, M. Penttilä, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen, and D. B. Kell, "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology," *Nature biotechnology*, vol. 26, pp. 1155–1160, Oct 2008.

[72] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proceedings of the Conference on Applications, Technologies,*

Architectures, and Protocols for Computer Communication, SIGCOMM '99, (New York, NY, USA), p. 251–262, Association for Computing Machinery, 1999.

[73] M. Newman, *Networks: An Introduction*. OUP Oxford, 2010.

[74] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hungary. Acad. Sci.*, vol. 5, pp. 17–61, 1960.

[75] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun 1998.

[76] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[77] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1, pp. 1–101, 2011.

[78] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014. The structure and dynamics of multilayer networks.

[79] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M. M. Murray, T. P. Peixoto, F. Vaccarino, and G. Petri, "The physics of higher-order interactions in complex systems," *Nature Physics*, vol. 17, pp. 1093–1098, Oct 2021.

[80] E. Ravasz and A.-L. Barabási, "Hierarchical organization in complex networks," *Phys. Rev. E*, vol. 67, p. 026112, Feb 2003.

[81] S. Milgram *Psychology Today*, no. 1, pp. 61–67.

[82] L. B. J. Ugander, B. Karrer and C. Marlow, "The anatomy of the facebook social graph," *arXiv:1111.4503 [cs.SI]*, 2011.

[83] P. B. M. R. J. U. L. Backstrom and S. Vigna, "Four degrees of separation," *Proceedings of the 4th Annual ACM Web Science Conference*, 2012.

[84] "Erdos-bacon number." Acessado em Novembro de 2021.

[85] F. Radicchi and C. Castellano, "Breaking of the site-bond percolation universality in networks," *Nat. Commun.*, vol. 6, p. 10196, 2015.

[86] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "*k*-core organization of complex networks," *Phys. Rev. Lett.*, vol. 96, p. 040601, Feb 2006.

[87] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Critical phenomena in complex networks," *Rev. Mod. Phys.*, vol. 80, pp. 1275–1335, Oct 2008.

[88] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras, "Collective versus hub activation of epidemic phases on networks," *Phys. Rev. E*, vol. 93, p. 032314, Mar 2016.

[89] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nat. Phys.*, vol. 6, pp. 888–893, 2010.

[90] T. P. Peixoto, "The graph-tool python library," *figshare*, 2014.

[91] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, pp. 509–512, oct 1999.

[92] G. Ghoshal, L. Chi, and A.-L. Barabási, "Uncovering the role of elementary processes in network evolution," *Scientific Reports*, vol. 3, p. 2920, Oct 2013.

[93] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, "Cut-offs and finite size effects in scale-free networks," *The European Physical Journal B*, vol. 38, pp. 205–209, Mar 2004.

[94] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, "Generation of uncorrelated random scale-free networks," *Phys. Rev. E*, vol. 71, p. 027103, 2005.

[95] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures and Algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[96] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Phys. Rep.*, vol. 544, pp. 1–122, 2014.

[97] G. Bianconi, *Multilayer Networks*. Oxford: Oxford University Press, 2018.

[98] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, pp. 203–271, 07 2014.

[99] G. F. de Arruda, E. Cozzo, T. P. Peixoto, F. A. Rodrigues, and Y. Moreno, "Disease Localization in Multilayer Networks," *Phys. Rev. X*, vol. 7, p. 011014, 2017.

[100] N. Van Kampen, *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library, Elsevier Science, 2011.

[101] T. Tome and M. De Oliveira, *Dinamica Estocastica E Irreversibilidade*. EDUSP.

[102] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, 1976.

[103] R. M. Lana, M. F. d. C. Gomes, T. F. M. d. Lima, N. A. Honório, and C. T. Codeço, "The introduction of dengue follows transportation infrastructure changes in the state of acre, brazil: A network-based analysis," *PLOS Neglected Tropical Diseases*, vol. 11, pp. 1–27, 11 2017.

[104] G. S. Costa, W. Cota, and S. C. Ferreira, "Outbreak diversity in epidemic waves propagating through distinct geographical scales," *Phys. Rev. Res.*, vol. 2, p. 043306, dec 2020.

[105] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "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*, vol. 115, no. 772, pp. 700–721, 1927.

[106] D. J. D. Earn, J. Dushoff, and S. A. Levin, "Ecology and evolution of the flu," Tech. Rep. 7, 2002.

[107] S. Strogatz, *Nonlinear Dynamics And Chaos*. Studies in nonlinearity, Levant Books, 2007.

[108] M. Boguñá and R. Pastor-Satorras, "Epidemic spreading in correlated complex networks," *Phys. Rev. E*, vol. 66, p. 047104, Oct 2002.

[109] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Trans. Inf. Syst. Secur.*, vol. 10, jan 2008.

[110] A. S. Mata and S. C. Ferreira, "Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks," *EPL (Europhysics Letters)*, vol. 103, no. 4, p. 48003, 2013.

[111] F. Chung, L. Lu, and V. Vu, "Spectra of random graphs with given expected degrees," *Proceedings of the National Academy of Sciences*, vol. 100, no. 11, pp. 6313–6318, 2003.

[112] C. Castellano and R. Pastor-Satorras, "Thresholds for epidemic spreading in networks," *Phys. Rev. Lett.*, vol. 105, p. 218701, Nov 2010.

[113] H. Sayama, *Introduction to the Modeling and Analysis of Complex Systems*. Open Textbook Library, Open SUNY Textbooks, Milne Library, State University of New York at Geneseo, 2015.

[114] W. Cota and S. C. Ferreira, "Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks," *Comput. Phys. Commun.*, vol. 219, pp. 303–312, 2017.

[115] M. Kardar, G. Parisi, and Y.-C. Zhang, "Dynamic scaling of growing interfaces," *Phys. Rev. Lett.*, vol. 56, pp. 889–892, Mar 1986.

[116] I. Corwin, "The kardar–parisi–zhang equation and universality class," *Random matrices: Theory and applications*, vol. 1, no. 01, p. 1130001, 2012.

[117] R. Dickman and R. Vidigal, "Quasi-stationary distributions for stochastic processes with an absorbing state," *Journal of Physics A: Mathematical and General*, vol. 35, pp. 1147–1166, jan 2002.

[118] R. S. Sander, G. S. Costa, and S. C. Ferreira, "Sampling methods for the quasistationary regime of epidemic processes on regular and complex networks," *Phys. Rev. E*, vol. 94, p. 042308, 2016.

[119] J. Blanchet, P. Glynn, and S. Zheng, "Analysis of a stochastic approximation algorithm for computing quasi-stationary distributions," *Adv. Appl. Probab.*, vol. 48, pp. 792–811, 2016.

[120] M. Boguñá, C. Castellano, and R. Pastor-Satorras, "Nature of the Epidemic Threshold for the Susceptible-Infected-Susceptible Dynamics in Networks," *Phys. Rev. Lett.*, vol. 111, p. 068701, 2013.

[121] P. Moretti and M. A. Muñoz, "Griffiths phases and the stretching of criticality in brain networks," *Nat. Commun.*, vol. 4, p. 2521, 2013.

[122] Z. W. Wei and B. H. Wang, "Susceptible-infected-susceptible model on networks with eigenvector localization," *Phys. Rev. E*, vol. 101, p. 042310, 2020.

[123] D. H. Silva and S. C. Ferreira, "Dissecting localization phenomena of dynamical processes on networks," *J. Phys. Complex.*, vol. 1, no. 2, pp. 1–13, 2021.

[124] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras, "Collective versus hub activation of epidemic phases on networks," *Phys. Rev. E*, vol. 93, p. 032314, mar 2016.

[125] C. Castellano and R. Pastor-Satorras, "Competing activation mechanisms in epidemics on networks," *Sci. Rep.*, vol. 2, p. 371, 2012.

[126] R. Pastor-Satorras and C. Castellano, "Eigenvector Localization in Real Networks and Its Implications for Epidemic Spreading," *J. Stat. Phys.*, vol. 173, pp. 1110–1123, 2018.

[127] R. S. Ferreira and S. C. Ferreira, "Critical behavior of the contact process on small-world networks," *Eur. Phys. J. B*, vol. 86, p. 462, 2013.

[128] R. S. Ferreira, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes, "Metastable localization of diseases in complex networks," *Phys. Rev. E*, vol. 94, p. 062305, 2016.

[129] Codes are freely available at https://github.com/ghscosta/qsmethods.

[130] K.-I. Goh, D.-S. Lee, B. Kahng, and D. Kim, "Sandpile on Scale-Free Networks," *Phys. Rev. Lett.*, vol. 91, p. 148701, 2003.

[131] R. S. Sander, S. C. Ferreira, and R. Pastor-Satorras, "Phase transitions with infinitely many absorbing states in complex networks," *Phys. Rev. E*, vol. 87, p. 022820, 2013.

[132] J. Depersin and M. Barthelemy, "From global scaling to the dynamics of individual cities," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, pp. 2317–2322, 2018.

[133] Q. Zhang, K. Sun, M. Chinazzi, A. Pastore y Piontti, N. E. Dean, D. P. Rojas, S. Merler, D. Mistry, P. Poletti, L. Rossi, M. Bray, M. E. Halloran, I. M. Longini, and A. Vespignani, "Spread of Zika virus in the Americas," *Proc. Natl. Acad. Sci.*, vol. 114, pp. E4334–E4343, 2017.

[134] S. Merler, M. Ajelli, L. Fumanelli, S. Parlamento, A. Pastore y Piontti, N. E. Dean, G. Putoto, D. Carraro, I. M. Longini, M. E. Halloran, and A. Vespignani, "Containing Ebola at the Source with Ring Vaccination," *PLoS Negl. Trop. Dis.*, vol. 10, p. e0005093, 2016.

[135] S. S. Manna, "Two-state model of self-organized criticality," *J. Phys. A. Math. Gen.*, vol. 24, pp. L363–L369, 1991.

[136] R. Dickman, M. Alava, M. A. Muñoz, J. Peltola, A. Vespignani, and S. Zapperi, "Critical behavior of a one-dimensional fixed-energy stochastic sandpile," *Phys. Rev. E*, vol. 64, p. 056104, 2001.

[137] S. Lubeck, R. D. Willmann, S. L beck, and R. D. Willmann, "Universal scaling behaviour of directed percolation and the pair contact process in an external field," *J. Phys. A. Math. Gen.*, vol. 35, pp. 10205–10217, 2002.

[138] M. M. de Oliveira, M. G. E. da Luz, and C. E. Fiore, "Generic finite size scaling for discontinuous nonequilibrium phase transitions into absorbing states," *Phys. Rev. E*, vol. 92, 2015.

[139] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[140] M. Boguñá, C. Castellano, and R. Pastor-Satorras, "Langevin approach for the dynamics of the contact process on annealed scale-free networks," *Phys. Rev. E*, vol. 79, p. 036110, 2009.

[141] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, "Heterogeneous pair-approximation for the contact process on complex networks," *New J. Phys.*, vol. 16, p. 053006, 2014.

[142] M. E. J. Newman and C. R. Ferrario, "Interacting Epidemics and Coinfection on Contact Networks," *PLoS One*, vol. 8, p. e71321, 2013.

[143] J. Sanz, C.-Y. Xia, S. Meloni, and Y. Moreno, "Dynamics of Interacting Diseases," *Phys. Rev. X*, vol. 4, p. 041005, 2014.

[144] S. Iwata, K. Kobayashi, S. Higa, J. Yoshimura, and K.-i. Tainaka, "A simple population theory for mutualism by the use of lattice gas model," *Ecol. Modell.*, vol. 222, pp. 2042–2048, 2011.

[145] Instituto Brasileiro de Geografia e Estatística (IBGE), *Divisão Regional do Brasil em Regiões Geográficas Imediatas e Regiões Geográficas Intermediárias*. Instituto Brasileiro de Geografia e Estatística Rio de Janeiro, 2017.

[146] A. Aktay, S. Bavadekar, G. Cossoul, J. Davis, D. Desfontaines, A. Fabrikant, E. Gabrilovich, K. Gadepalli, B. Gipson, M. Guevara, *et al.*, "Google COVID-19 community mobility reports: Anonymization process description (version 1.0)," *arXiv preprint arXiv:2004.04145*, 2020.

[147] H. Hu, K. Nigmatulina, and P. Eckhoff, "The scaling of contact rates with population density for the infectious disease models," *Math. Biosci.*, vol. 244, pp. 125–134, 2013.

[148] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley, 2000.

[149] P. Kareiva, "Habitat fragmentation and the stability of predator–prey interactions," *Nature*, vol. 326, pp. 388–390, Mar 1987.

[150] I. Hanski, "Metapopulation dynamics: Does it help to have more of the same?," *Trends Ecol. Evol.*, vol. 4, pp. 113–114, 1989.

[151] L. Danon, T. House, and M. J. Keeling, "The role of routine versus random movements on the spread of disease in Great Britain," *Epidemics*, vol. 1, pp. 250–258, 2009.

[152] V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction-diffusion processes and metapopulation models in heterogeneous networks," *Nat. Phys.*, vol. 3, pp. 276–282, 2007.

[153] W. Cota, D. Soriano-Paños, A. Arenas, S. C. Ferreira, and J. Gómez-Gardeñes, "Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility," *New Journal of Physics*, vol. 23, p. 073019, jul 2021.

[154] P. Hallal, F. Hartwig, B. Horta, G. D. Victora, M. Silveira, C. Struchiner, L. P. Vidaletti, N. Neumann, L. C. Pellanda, O. A. Dellagostin, M. N. Burattini, A. M. Menezes, F. C. Barros, A. J. Barros, and C. G. Victora, "Remarkable variability in SARS-CoV-2 antibodies across Brazilian regions: nationwide serological household survey in 27 states," *medRxiv*, p. 2020.05.30.20117531, 2020.

[155] V. E. Pitzer, G. M. Leung, and M. Lipsitch, "Estimating variability in the transmission of severe acute respiratory syndrome to household contacts in Hong Kong, China," *American Journal of Epidemiology*, vol. 166, no. 3, pp. 355–363, 2007.

[156] W. Cota, "Monitoring the number of COVID-19 cases and deaths in brazil at municipal and federative units level," *SciELOPreprints.362*, may 2020.

[157] R. Subramanian, Q. He, and M. Pascual, "Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity," *Proc. Natl. Acad. Sci.*, vol. 118, p. e2019716118, mar 2021.

[158] N. J. Irons and A. E. Raftery, "Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 31, 2021.

[159] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, "Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, pp. 1199–1207, mar 2020.

[160] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, "Serial interval of novel coronavirus (COVID-19) infections," *Int. J. Infect. Dis.*, vol. 93, pp. 284–286, apr 2020.

[161] D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, and G. Favre, "Real estimates of mortality following COVID-19 infection," *Lancet Infect. Dis.*, vol. 20, p. 773, jul 2020.

[162] X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, and G. M. Leung, "Temporal dynamics in viral shedding and transmissibility of COVID-19," *Nat. Med.*, vol. 26, pp. 672–675, may 2020.

[163] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, and R. M. Eggo, "Age-dependent effects in the transmission and control of COVID-19 epidemics," *Nat. Med.*, vol. 26, pp. 1205–1211, aug 2020.

[164] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application," *Ann. Intern. Med.*, vol. 172, pp. 577–582, may 2020.

[165] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, "Impact of lockdown on COVID-19 epidemic in île-de-france and possible exit strategies," *BMC Med.*, vol. 18, p. 240, dec 2020.

[166] S. Moore, E. M. Hill, M. J. Tildesley, L. Dyson, and M. J. Keeling, "Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study," *Lancet Infect. Dis.*, vol. 21, pp. 793–802, mar 2021.

[167] F. Ahmed, N. Ahmed, C. Pissarides, and J. Stiglitz, "Why inequality could spread COVID-19," *Lancet Public Heal.*, vol. 5, p. e240, may 2020.

[168] K. Sun, W. Wang, L. Gao, Y. Wang, K. Luo, L. Ren, Z. Zhan, X. Chen, S. Zhao, Y. Huang, Q. Sun, Z. Liu, M. Litvinova, A. Vespignani, M. Ajelli, C. Viboud, and H. Yu, "Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2," *Science*, vol. 371, p. eabe2424, jan 2021.

[169] V. N. Petrova and C. A. Russell, "The evolution of seasonal influenza viruses," *Nature Reviews Microbiology*, vol. 16, pp. 47–60, Jan 2018.

[170] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, and N. M. Ferguson, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *Lancet Infect. Dis.*, vol. 20, pp. 669–677, 2020.

[171] G. do Estado do Paraná, 2021. Boletim - Informe Epidemiológico Coronavírus (COVID-19), https://www.saude.pr.gov.br/Pagina/Coronavirus-COVID-19 [Online; accessed 08-Sep-2021].

[172] G. do Estado do Espírito Santo, 2021. COVID-19 - Painel COVID-19 - Estado do Espírito Santo, https://coronavirus.es.gov.br/painel-covid-19-es [Online; accessed 08-Sep-2021].

[173] "Codes and datasets are freely available at https://github.com/ghscosta/covid19-cal.."

[174] M. C. Castro, S. Kim, L. Barberia, A. F. Ribeiro, S. Gurzenda, K. B. Ribeiro, E. Abbott, J. Blossom, B. Rache, and B. H. Singer, "Spatiotemporal pattern of COVID-19 spread in brazil," *Science*, vol. 372, pp. 821–826, apr 2021.

[175] L. F. Buss, C. A. Prete, C. M. M. Abrahim, A. Mendrone, T. Salomon, C. de Almeida-Neto, R. F. O. França, M. C. Belotti, M. P. S. S. Carvalho, A. G. Costa, M. A. E. Crispim, S. C. Ferreira, N. A. Fraiji, S. Gurzenda, C. Whittaker, L. T. Kamaura, P. L. Takecian, P. da Silva Peixoto, M. K. Oikawa, A. S. Nishiya, V. Rocha, N. A. Salles, A. A. de Souza Santos, M. A. da Silva, B. Custer, K. V. Parag, M. Barral-Netto, M. U. G. Kraemer, R. H. M. Pereira, O. G. Pybus, M. P. Busch, M. C. Castro, C. Dye, V. H. Nascimento, N. R. Faria, and E. C. Sabino, "Three-quarters attack rate of SARS-CoV-2 in the brazilian amazon during a largely unmitigated epidemic," *Science*, vol. 371, pp. 288–292, dec 2020.

[176] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecast.*, vol. 32, pp. 669–679, jul 2016.

[177] "Supplementary material."

[178] K. Prem, A. R. Cook, and M. Jit, "Projecting social contact matrices in 152 countries using contact surveys and demographic data," *PLOS Comput. Biol.*, vol. 13, p. e1005697, 2017.

[179] R. B. Griffiths, "Nonanalytic behavior above the critical point in a random ising ferromagnet," *Phys. Rev. Lett.*, vol. 23, pp. 17–19, Jul 1969.

[180] A. F. Rozenfeld, R. Cohen, D. ben Avraham, and S. Havlin, "Scale-free networks on lattices," *Phys. Rev. Lett.*, vol. 89, p. 218701, Nov 2002.

[181] D. Zwillinger, *Table of Integrals, Series, and Products*. Elsevier, 2015.