

WESLEY FRANCIS COSTA COTA

SPREADING PHENOMENA ON COMPLEX NETWORKS AND SOCIAL SYSTEMS

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

Adviser: Silvio da Costa Ferreira Junior

Co-adviser: Jesús Gómez-Gardeñes

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

T

C843s Cota, Wesley Francis Costa, 1994-
2020 Spreading phenomena on complex networks and social
systems / Wesley Francis Costa Cota. – Viçosa, MG, 2020.
195 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Silvio da Costa Ferreira Junior.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.177-195.

1. Teoria das redes complexas (Física). 2. Fenômenos críticos. 3. Sociofísica. I. Universidade Federal de Viçosa. Departamento de Física. Programa de Pós-Graduação em Física. II. Título.

CDD 22. ed. 530.1

WESLEY FRANCIS COSTA COTA

SPREADING PHENOMENA ON COMPLEX NETWORKS AND SOCIAL SYSTEMS

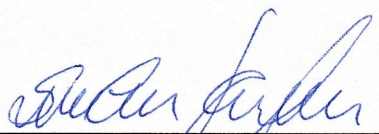
Thesis presented to the Universidade Federal de Viçosa, in partial fulfillment of the requirements of the Physics Graduate Program, for the degree of *Doctor Scientiae*.

APPROVED: September 29, 2020.

Assent:



Wesley Francis Costa Cota
Author



Silvio da Costa Ferreira Junior
Adviser

For my mother Lourdes and my father José,
who did everything to get me here.

Acknowledgements

First I thank my parents, Lourdes and José, for providing me all the education and support to get to this stage. Without your love and care, it would not be possible. You are incredible and irreplaceable.

Exceptional thanks to Thamires C. Soares, for all the love, care, and moments we spent together. Always supporting me in the good and bad moments of my life and, of course, while writing this thesis. You are the most awesome person in the world, and very soon it will be my time to see you shine!

I would like to thank enormously and to express my gratitude to my advisor Prof. Silvio C. Ferreira, who has been an extraordinary mentor during all my career. Your guidance, suggestions, and discussions allow me to learn how to be a research scientist. I hope that we can continue this collaboration and friendship that we have built throughout this time.

My most sincere thanks to all the staff and professors of Universidade Federal de Viçosa (UFV), an incredible and prestigious public university in Brazil and my academic home so far. You are very important for the formation of a whole generation of Brazilian scientists. In particular, my gratitude goes to the Department of Physics, and colleagues of the graduate program in Physics.

I also would like to thank my friends from GISC (Grupo de Investigação de Sistemas Complexos), especially Diogo H. Silva, José Carlos and Guilherme S. Costa. My special thanks go as well to former students of the group, especially my co-author Prof. Angélica S. Mata.

I would also like to thank immensely my co-advisor Prof. Jesús Gómez-Gardeñes for his hospitality and guidance during my internship at Universidad de Zaragoza (UZ), even in the middle of a real-time epidemic spreading process of COVID-19 alone in Spain. I hope that this collaboration has further increased your affection for Brazil.

I am deeply grateful for all friends made in the GOTHAM Lab in Zaragoza: David Soriano-Paños, Adriana Reyna-Lara, Sergio Faci-Lázaro, Francisco Bauzá, and Prof. Luis Mario Floría. You made me feel at home! In particular, David for the extensive discussions about metapopulation modeling, and both David and Adriana for conversations about epidemic models and, of course, about non-academic things in life.

I cannot fail to mention the fantastic hospitality of my co-author Prof. Géza Ódor during my

stay of a few days in Budapest. I also thank my co-authors Michele Starnini and Prof. Romualdo Pastor-Satorras for my stay in Barcelona that resulted in the work about echo chambers on Twitter. I hope we can all continue discussions and collaborations.

My thanks to all those who contributed in some way to the results of my thesis. I apologize in advance if I have forgotten to mention someone. In particular, I thank Diogo H. Silva and Prof. Gino Ceotto for the hashtags classification of Twitter.

I acknowledge the financial assistance received during these years for participation in events by the Brazilian agencies FAPEMIG, CNPq, and CAPES. I thank the graduate program of Physics for the financial help for many events, especially to attend the Lake Como School of Advanced Studies on complex networks of 2018 in Como, Italy. I also thank the Complex System Society for the registration fee and accommodation for the Winter Workshop on Complex Systems (WWCS 2019) in Poland, and Sunbelt 2020 conference, online; the ICTP for the travel grant to StatPhys 27 conference in Argentina and schools in São Paulo/SP; and the Instituto Nacional de Ciência e Tecnologia de Sistemas Complexos (INCT-SC) for the financial help for StatPhys 27.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. I am immensely grateful to CAPES for the one-year scholarship by the *Programa de Doutorado Sanduíche no Exterior* (PDSE), 2019–2020. The scholarships by CAPES received during my doctorate at UFV and internship at UZ were fundamental to the elaboration of this thesis.

*“I simply wish that, in a matter which so closely concerns the wellbeing
of the human race, no decision shall be made without all the
knowledge which a little analysis and calculation can provide.”*

— Daniel Bernoulli, 1760

Abstract

COTA, Wesley Francis Costa, D.Sc., Universidade Federal de Viçosa, September, 2020. **Spreading phenomena on complex networks and social systems.** Adviser: Silvio da Costa Ferreira Junior. Co-adviser: Jesús Gómez-Gardeñes.

Epidemic spreading has been one of the most prominent and widely investigated issues in the recent literature of complex networks. Despite several advances, the mathematical modeling of spreading processes remains the target of intensive investigations. In this thesis, we initially focused on the fundamental aspects of these dynamical processes on large and highly heterogeneous networks. We investigate some modifications of the susceptible-infected-susceptible (SIS) epidemic model to determine the robustness of epidemic thresholds by using extensive computer simulations and mean-field theories. Our results suggest that the metastable states of the standard SIS dynamics are not universal on networks that follow a power-law degree distribution $P(k) \sim k^{-\gamma}$, with $\gamma > 5/2$. We also investigated the effects of quenched disorder in the SIS dynamics on modular networks, whose results point to the existence of extended regions of criticality in networks with infinite dimension and lack of hierarchy, leading to Griffiths phases, which can be a mechanism to explain the criticality of complex systems such as the brain. We also focused on more specific problems of spreading processes involving social and human mobility patterns. First, we explored the influence of political lean in information spreading over a political communication network reconstructed using data collected from Twitter. In this case, we showed that although opposite sides of a discussion form the so-called echo chambers, sharing similar opinions with each other, some users were able to spread the information better than others, breaking these echo chambers. This spreading capacity was strongly related to a diversity of users reached. Finally, we also examined the epidemic spreading on top of metapopulations. We adapted an existing analytical framework for the SIS dynamics on populations with recurrent mobility to accommodate the heterogeneous nature of human social contacts. We derived an analytical expression that involves demography, mobility, and contact patterns, that were used in synthetic networks to understand how they can change the epidemic threshold. All these contributions take a step further to a better understanding of the relationship between the structure of networks and epidemic dynamics, which is extremely important to improve epidemic models to deal with real situations.

Keywords: Complex networks. Spreading processes. Critical phenomena. Sociophysics.

Resumo

COTA, Wesley Francis Costa, D.Sc., Universidade Federal de Viçosa, setembro de 2020. **Processos de espalhamento em redes complexas e sistemas sociais.** Orientador: Silvio da Costa Ferreira Junior. Coorientador: Jesús Gómez-Gardeñes.

A disseminação de epidemias tem sido um dos problemas mais proeminentes e amplamente investigados na literatura recente de redes complexas. Apesar de diversos avanços, a modelagem matemática de processos de espalhamento ainda permanece como alvo de investigações intensivas. Nesta tese focamos inicialmente em aspectos fundamentais desses processos dinâmicos em redes grandes e altamente heterogêneas. Investigamos algumas modificações do modelo suscetível-infectado-suscetível (SIS) para determinar a robustez dos limiares epidêmicos usando simulações computacionais intensivas e teorias de campo médio. Os resultados sugerem que os estados meta-estáveis do modelo SIS tradicional não são universais em redes que seguem distribuições em lei de potência $P(k) \sim k^{-\gamma}$, com $\gamma > 5/2$. Também investigamos os efeitos da desordem congelada na dinâmica SIS em redes modulares, cujos resultados apontam para a existência de regiões estendidas de criticalidade em redes com dimensão infinita e sem hierarquia, levando às fases de Griffiths, que podem ser um mecanismo para explicar a criticalidade de sistemas complexos como o cérebro. Focamos também em problemas mais específicos de processos de espalhamento envolvendo padrões sociais e de mobilidade humana. Primeiramente, exploramos a influência da inclinação política no espalhamento de informação em redes de comunicação política reconstruídas usando dados coletados do Twitter. Neste caso, mostramos que apesar de lados opostos de uma discussão formarem as chamadas câmaras de eco, compartilhando opiniões semelhantes entre si, alguns usuários foram capazes de espalhar a informação de forma melhor que outros, quebrando essas câmaras de eco. Essa capacidade de espalhamento está fortemente relacionada à diversidade dos usuários alcançados. Finalmente, também examinamos o espalhamento de epidemias em metapopulações. Adaptamos um formalismo teórico já existente para a dinâmica SIS em populações com mobilidade pendular para acomodar a natureza heterogênea dos contatos sociais humanos. Determinamos uma expressão analítica que envolve demografia, mobilidade e padrões de contato, que foram usados em redes sintéticas para entender como eles podem mudar o limiar epidêmico. Todas essas contribuições dão um passo adiante para o melhor entendimento da relação entre estrutura e dinâmica de epidemias, que é extremamente importante para aprimorar modelos epidêmicos para lidar com situações reais.

Palavras-chave: Redes complexas. Processos de espalhamento. Fenômenos críticos. Sociofísica.

List of Publications

1. [W. Cota](#), S. C. Ferreira, G. Ódor. *Griffiths effects of the susceptible-infected-susceptible epidemic model on random power-law networks*. [Physical Review E](#) **93**, 032322 (2016) [Chapter 5]
2. [W. Cota](#), S. C. Ferreira. *Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks*. [Computer Physics Communications](#) **219C** (2017) pp. 303-312 [Chapter 3]
3. [W. Cota](#), G. Ódor, S. C. Ferreira. *Griffiths phases in infinite-dimensional, non-hierarchical modular networks*. [Scientific Reports](#) **8**:9144 (2018) [Chapter 5]
4. [W. Cota](#), A. S. Mata, S. C. Ferreira. *Robustness and fragility of the susceptible-infected-susceptible epidemic models on complex networks*. [Physical Review E](#) **98**, 012310 (2018) [Chapter 4]
5. [W. Cota](#), S. C. Ferreira, R. Pastor-Satorras, M. Starnini. *Quantifying echo chamber effects in information spreading over political communication networks*. [EPJ Data Science](#) **8**, 35 (2019) [Chapter 6]
6. [W. Cota](#), D. Soriano-Paños, A. Arenas, S. C. Ferreira, J. Gómez-Gardeñes. *Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility*. [arXiv:2102.10614](#) (2021) [Chapter 7]

Other contributions

1. D. H. Silva, S. C. Ferreira, [W. Cota](#), R. Pastor-Satorras, C. Castellano. *Spectral properties and the accuracy of mean-field approaches for epidemics on correlated networks*. [Physical Review Research](#) **1**, 033024 (2019)
2. A. Arenas, [W. Cota](#), J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños, B. Steinegger. *A mathematical model for the spatiotemporal epidemic spreading of COVID-19*. [medRxiv](#) 2020.03.21.20040022 (2020)
3. [W. Cota](#). *Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level*. [SciELOPreprints](#):362 (2020)
4. G. S. Costa, [W. Cota](#), S. C. Ferreira. *Outbreak diversity in epidemic waves propagating through distinct geographical scales*. [Physical Review Research](#) **2**, 043306 (2020)
5. A. Arenas, [W. Cota](#), J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños, B. Steinegger. *Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions*. [Physical Review X](#) **10**, 041055 (2020)

List of Abbreviations

API	Application programming interface
APT	Absorbing-state phase transition
CP	Contact process
DP	Directed percolation
FSS	Finite-size scaling
GA	Gillespie algorithm
GP	Griffiths phase
HMF	Heterogeneous mean-field
LCC	Largest connected component
MC	Monte Carlo
ME	Master equation
MMN	Monodisperse modular network
NN	Nearest neighbor
OD	Origin-destination
OGA	Optimized Gillespie algorithm
PC	Political communication
PHMF	Pairwise heterogeneous mean-field
PL	Power-law
PMN	Polydisperse modular network
PQMF	Pairwise quenched mean-field
QMF	Quenched mean-field
QS	Quasi-stationary
RBC	Reflective bound condition
r.h.s.	Right hand side
RRN	Random regular network
RR	Rare-region
RT	Retweet
SCC	Strongly connected component
SF	Scale-free
SIR	Susceptible-infected-recovered
SIS	Susceptible-infected-susceptible
UCM	Uncorrelated configuration model

Contents

I Preliminaries	15
1 Introduction	16
2 Concepts of network theory	20
2.1 Characterization of networks	20
2.1.1 Directed and weighted networks	21
2.1.2 Connectivity and paths	22
2.1.3 Centrality measures	24
2.1.4 Communities in networks	25
2.2 Statistical characterization of complex networks	27
2.2.1 Degree distributions	27
2.2.2 Mixing and correlations	30
2.3 Synthetic networks	32
2.3.1 Simple networks	32
2.3.2 Configuration model	33
2.3.3 Annealed networks	33
2.4 Summary	34
3 Spreading processes on complex networks	35
3.1 Master equations for dynamical processes	35
3.1.1 Equilibrium and non-equilibrium systems	37
3.1.2 Poisson processes and the Gillespie algorithm	37
3.2 Epidemics in networks as an absorbing-state phase transition	39
3.2.1 Compartmental models for epidemics	39
3.2.2 SIS dynamics on a star graph	41
3.2.3 Critical exponents	42
3.2.4 Quasi-stationary state	44
3.3 Simulations and analysis of epidemic processes	47
3.3.1 Mean-field approximations	48

3.3.2	Building algorithms for generic epidemic models	49
3.3.3	Decay, spreading and quasi-stationary analysis	53
3.4	Summary	54
II	Criticality in the epidemic spreading on complex networks	55
4	Robustness and fragility of the susceptible-infected-susceptible model	56
4.1	Analytical approximations for the epidemic threshold	56
4.1.1	Modifications of the SIS model	57
4.1.2	Mean field theories and epidemic thresholds	59
4.1.3	Testing HMF theory using annealed networks	61
4.1.4	Simulations on quenched random networks	62
4.2	Activation mechanisms in networks	64
4.2.1	Approximated expressions for epidemic lifespan on star graphs	65
4.2.2	Activation mechanisms of the modified dynamics on quenched PL networks	67
4.2.3	Finite-size scaling of critical quantities	70
4.2.4	Accuracy of the QMF theory and modified dynamics on real networks . .	72
4.3	Summary and discussions	73
5	Griffiths phases on non-hierarchical modular random networks	75
5.1	Effects of quenched disorder on criticality of complex systems	75
5.1.1	Griffiths phases on ferromagnetic systems	76
5.1.2	Griffiths phases in the epidemic spreading	77
5.1.3	Conjectures about Griffiths phases on small-world networks	78
5.2	Independent power-law networks with cutoffs	79
5.2.1	Time-dependent analysis	80
5.2.2	Sample-to-sample fluctuations	82
5.2.3	Optimal fluctuation theory	83
5.2.4	Finite-size analysis	85
5.2.5	Quasi-stationary analysis	86
5.3	Griffiths phases on loosely coupled networks of modules	90
5.3.1	Infinite dimensional and non-hierarchical modular networks	91
5.3.2	Decay analysis	94

5.3.3	Spreading analysis	97
5.4	Summary and discussions	101
III	Models and applications on directed social and metapopulation networks	103
6	Quantifying echo chamber effects over political communication networks	104
6.1	Online social networks	104
6.1.1	Echo chambers in social networks	105
6.1.2	Obtaining data from Twitter	105
6.2	Reconstruction of political communication networks on Twitter	109
6.2.1	Classification of hashtags	109
6.2.2	Filtering process of the dataset to obtain the networks	110
6.2.3	Properties of the temporal and integrated networks	114
6.3	Defining the political position as a continuous variable	115
6.3.1	Content diversity and political position	117
6.3.2	Structural evidences of echo chambers	119
6.3.3	Spreading capacity as a function of political position	119
6.4	Summary and discussions	125
7	Metapopulation structures with heterogeneous subpopulations	127
7.1	Epidemics in metapopulations with recurrent mobility	127
7.2	Markov-chain approach in networks with recurrent mobility	128
7.2.1	Critical regime and linearization	131
7.2.2	Perturbation analysis for the homogeneous mixing case	131
7.2.3	Different critical scenarios and the detriment effect	132
7.3	Markovian equations for heterogeneous subpopulations	134
7.3.1	Synthetic structures with heterogeneous subpopulations	139
7.3.2	Agreement with Monte Carlo simulations	141
7.3.3	Analytical evaluation of the epidemic threshold	143
7.4	Summary and perspectives	148

IV	Concluding remarks	150
8	Summary and prospects	151
	Appendix	154
A	Computer implementations of SIS epidemic models	154
A.1	SIS- \mathcal{S}	154
A.2	SIS- \mathcal{A}	155
A.3	SIS- \mathcal{T}	155
A.4	Simulation on uncorrelated annealed networks	156
B	Supplementary tables and results for Chapter 6	157
B.1	List of the 443 hashtags with partial and full agreement	157
B.2	List of the 52 hashtags with divergent opinions	161
B.3	Effects of the hashtag classification on the reconstructed network	161
B.4	Results for SIS and SIR dynamics with different parameters	167
B.5	Results for the Watts threshold model	169
C	Supplementary calculations for Chapter 7	170
C.1	An upper bound limit of the leading eigenvalue of a matrix	170
C.2	Perturbation analysis of Eq. (7.40)	171
C.3	Exact evaluation of the epidemic threshold for a two-patches metapopulation	174
C.4	Exact evaluation of the epidemic threshold for a star-like metapopulation	175
	Bibliography	177

Part I

Preliminaries

Chapter 1

Introduction

Our daily life is permeated with networked systems [1–4] as the biochemical networks regulating cell processes inside our bodies, connections of neurons in our brain, transportation infrastructure through which we move, social media in which we get informed, which are just a few examples. Major advances have been made in the past few decades in the understanding of the structure and functioning of networks [1, 2], and network science has been marked by its interdisciplinary nature since its consolidation as a new branch [2, 5], especially in the investigation of dynamical processes on networked substrates [6]. This highly multidisciplinary discipline is dedicated to the understanding of how complex systems can be described by networks, which are mostly large and highly heterogeneous [2], and how dynamical processes evolve on top of them. There are many distinct fields of science such as mathematics, sociology, computer science, and physics involved in these studies. In particular, statistical physics contributes with its deep connection with graph theory and its characterizations of emergent collective phenomena by looking at the basic microscopic elements of a system [7].

Among the investigated problems, an important question is how pathogens spread in heterogeneous populations. Indeed, the epidemic spreading is one of the most prominent and widely investigated issues, usually through stochastic agent-based models [8]. Despite several advances in the understanding of epidemic models on networks [8–15], it remains target of recent intensive investigations [16–23] and there are still big challenges in the theoretical frameworks for dynamical processes in general evolving on top of these networked complex systems [2, 6].

Epidemic models on networks [8] assume that individuals of a population are represented by vertices while the infection can be transmitted through edges connecting them. Not only

approximated and exact theories have been developed, but also simulations of very large systems have played a key role in the understanding of central issues of epidemic spreading on highly heterogeneous networks. Most of these studies involve thresholds separating an absorbing, disease-free state and an active phase where epidemics can thrive and, thus, this problem can be suited in the framework of absorbing-state phase transitions (APTs) [8, 24, 25]. The location and the existence of a finite epidemic threshold [11, 15–17, 26–28] and the exponents describing the dynamics near the transition [29–33] have been a matter of intense discussions fomented by numerical simulations in networks with power-law degree distributions $P(k) \sim k^{-\gamma}$. Therefore, computer simulations have become fundamental tools to corroborate or to point out the limitations of the theories as well as to provide physical insights into the construction of new ones.

In this thesis, we investigate a series of spreading processes problems on networks. Initially, we briefly review some basic concepts of network theory and the analysis of epidemics on finite networks as an absorbing-state phase transition in Chapters 2 and 3, respectively. After discussing important properties of complex networks in the former, we proceed to describe dynamical processes on top of these structures in the latter. In particular, since accurate and efficient simulation methods have become imperative for the progress of this field, we define some optimized algorithms [34], which are simultaneously accurate and efficient, in Sec. 3.3.2 of Chapter 3.

One of the most basic but still not fully understood epidemic processes on networks is the susceptible-infected-susceptible (SIS) model [8], which consists of agents lying on the vertices of a network which can be infected or susceptible. Infected individuals become spontaneously healed (susceptible) with rate μ and transmit the disease to their susceptible contacts with rate λ . In principle, the SIS dynamics can exhibit a phase transition between a disease-free (absorbing) state and an active stationary phase, in which the epidemics persists in an endemic state. The transition occurs at an epidemic threshold λ_c . However, for uncorrelated random networks with a power-law degree distribution, it was rigorously proved [9] and later put in sound physical grounds [15] that the absorbing phase is unstable in the thermodynamic limit implying that λ_c is formally zero. We check the robustness and fragility of these results for the SIS model [35] in Chapter 4. We do this by looking at two modifications of the standard SIS model that preserve the central properties of spontaneous healing and infection capacity of a vertex increasing unlimitedly with its degree. For uncorrelated random networks with

$\gamma < 5/2$, the SIS dynamics is robust exhibiting essentially the same outcomes for all investigated models. However, we observe differences among them for $\gamma > 5/2$, implying different activation mechanisms.

Another interesting issue is the presence of critical states in complex systems, in which spatial and temporal correlations diverge [36, 37]. There are many questions about why and how these systems are tuned to criticality, and models [38, 39] and mechanisms [40, 41] on homogeneous substrates have been used to address these questions. Yet, the natural heterogeneity of networks can be relevant for the outcomes of models investigated on them, as discussed previously. In particular, the quasi-static (quenched) disorder, with timescales much longer than those of the dynamics, originated from the heterogeneous network topology affects the observed critical state. In Chapter 5 we discuss the so-called Griffiths phases (GPs) [42] that lead to dynamical criticality with long-term temporal correlations. Extensive simulations were made by averaging the activity density over many realizations of networks and we found extended regions of the control parameter space with Griffiths effects induced by these topological inhomogeneities [43]. We also found these extended control parameter regions with continuously changing dynamical exponents for single network realizations of heterogeneous random networks with a highly modular structure and non-hierarchically organized [44]. These results can have implications on the understanding of how systems with modular structures behave, such as in the brain [45], socio-technological [46, 47], and protein interaction networks [48].

Social interactions of individuals exchanging information and opinions about specific subjects on social networks can also be investigated as spreading processes. Indeed, social contagion phenomena have strong similarities with those described by epidemiological models [6]. In Chapter 6 we develop a method [49] able to quantify the effect of the so-called echo chambers [50] in social networks using spreading models such as the SIS model. These echo chambers are formed by users that tend to share information and opinions among others with similar beliefs, reinforcing their view about a subject. In particular, we look at discussions about the impeachment process of the former Brazilian president Dilma Rousseff [49], finding strong correlations between political lean and information spreading capacity on political communication networks. The proposed method can be applied for any context where opposing opinions are possible to observe echo chamber effects and their implications.

Finally, in Chapter 7 we investigate the effect of the interlacing of human mobility and social contacts with spreading processes. Besides social interactions described by networks, another

level such as infrastructure transportation has strong effects [51] on how people move and interact with others. Notably, this multiscale mobility [52] has changed drastically how diseases spread over the whole world and can have big impacts such as those that have happened with H1N1 [53, 54], Ebola [55, 56] and more recently, COVID-19 [57, 58]. A feasible manner to describe mathematically these mobility patterns is by using metapopulation networks [59–64], whose elements are places where individuals live and edges define the flows of individuals from one place to others. Here, we incorporate the heterogeneous nature of social contacts in a metapopulation SIS model and scrutinize their effects on how epidemics spread by generalizing an existent model [61] on metapopulations with recurrent mobility. We used synthetic networks to discuss the effects of heterogeneous connectivity on the epidemic threshold, showing that the epidemic threshold always increases when homogeneous mobility takes place in this model. We also extracted an analytical expression involving demographics, mobility, and social contact patterns that can be applied for any metapopulation structure when data are available. Certainly, the ideas shown here can be used to include the contact heterogeneity in more realistic models based on Ref. [61], such as the ones with our contributions to investigate the COVID-19 spreading in Brazil [63] and Spain [64], published in *Physical Review Research* and *Physical Review X*, respectively.

After discussing general aspects about networks and dynamical processes on them in Part I, all contributions presented in this thesis involve problems that relate the structure of complex networks with the dynamic properties of epidemics. Not only fundamental issues were investigated, as in Chapters 4 and 5 of Part II, but also specific problems of spreading processes on top of social and metapopulation networks in Chapters 6 and 7, respectively, of Part III.

Chapter 2

Concepts of network theory

The discovery of the structure of interactions of many complex systems and their description using networks has motivated several studies about the influence of structural properties and collective phenomena on top of them [65]. Using concepts of graph theory, the elements of a system can be represented by vertices and their interactions by edges or links between them. The connectivity properties of these elements are often universal, sharing aspects among different kinds of real networks, such as the small-world and scale-free phenomena that will be discussed in this chapter, in which we present a brief overview of fundamental concepts and definitions about networks. Concepts not used in any other part of this thesis were omitted and can be found in the references used as the basis for this chapter [1, 2, 6, 66]. Some specific definitions are postponed to the chapters where they are used.

2.1 Characterization of networks

Mathematically, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is composed by a set of vertices $\mathcal{V} = \{1, 2, \dots, N\}$, whose N elements are connected by E edges described by a set $\mathcal{E} = \{e_1, e_2, \dots, e_E\}$, where the ordered set $e_k = (i, j)$ represents an edge pointing from vertex i to j . The size of the network is referred as the total number of vertices N . Edges are also called links or connections. It is useful to define an adjacency matrix whose elements are

$$A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{E}, \\ 0 & \text{if } (i, j) \notin \mathcal{E}. \end{cases} \quad (2.1)$$

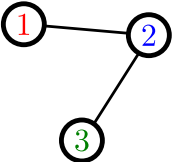
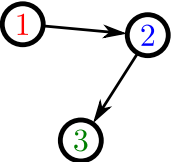
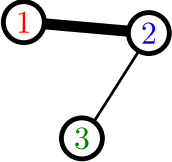
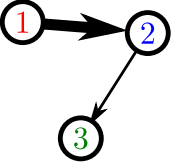
	Undirected	Directed																																
Unweighted	 <table><tr><th>A_{ij}</th><th>1</th><th>2</th><th>3</th></tr><tr><th>1</th><td>0</td><td>1</td><td>0</td></tr><tr><th>2</th><td>1</td><td>0</td><td>1</td></tr><tr><th>3</th><td>0</td><td>1</td><td>0</td></tr></table>	A_{ij}	1	2	3	1	0	1	0	2	1	0	1	3	0	1	0	 <table><tr><th>A_{ij}</th><th>1</th><th>2</th><th>3</th></tr><tr><th>1</th><td>0</td><td>1</td><td>0</td></tr><tr><th>2</th><td>0</td><td>0</td><td>1</td></tr><tr><th>3</th><td>0</td><td>0</td><td>0</td></tr></table>	A_{ij}	1	2	3	1	0	1	0	2	0	0	1	3	0	0	0
A_{ij}	1	2	3																															
1	0	1	0																															
2	1	0	1																															
3	0	1	0																															
A_{ij}	1	2	3																															
1	0	1	0																															
2	0	0	1																															
3	0	0	0																															
Weighted	 <table><tr><th>W_{ij}</th><th>1</th><th>2</th><th>3</th></tr><tr><th>1</th><td>0</td><td>4</td><td>0</td></tr><tr><th>2</th><td>4</td><td>0</td><td>1</td></tr><tr><th>3</th><td>0</td><td>1</td><td>0</td></tr></table>	W_{ij}	1	2	3	1	0	4	0	2	4	0	1	3	0	1	0	 <table><tr><th>W_{ij}</th><th>1</th><th>2</th><th>3</th></tr><tr><th>1</th><td>0</td><td>4</td><td>0</td></tr><tr><th>2</th><td>0</td><td>0</td><td>1</td></tr><tr><th>3</th><td>0</td><td>0</td><td>0</td></tr></table>	W_{ij}	1	2	3	1	0	4	0	2	0	0	1	3	0	0	0
W_{ij}	1	2	3																															
1	0	4	0																															
2	4	0	1																															
3	0	1	0																															
W_{ij}	1	2	3																															
1	0	4	0																															
2	0	0	1																															
3	0	0	0																															

Figure 2.1: Examples of different kinds of graphs regarding their edge directions and weights. The elements are represented by circles (1, 2, and 3), and edges by lines (undirected) or arrows (directed) connecting them. Edges can have different weights represented by the thickness of the lines. The adjacency matrix A_{ij} is shown for each case for undirected and directed networks. For weighted networks the weighted matrix W_{ij} is shown while their adjacency matrices are the same, depending only on the directionality of edges.

We will consider hereafter that self-connections $A_{ii} = 1$ are not allowed, except when stated otherwise, as in Chapter 7.

2.1.1 Directed and weighted networks

In undirected networks, connections between i and j are reciprocal such that $A_{ij} = 1$ implies $A_{ji} = 1$, while in directed networks $A_{ij} = 1$ does not necessarily mean that $A_{ji} = 1$. In the former case, A_{ij} is a symmetrical $N \times N$ matrix. Directionality has crucial consequences for the connectedness of a graph and information spreading, as will be discussed in Chapter 6. We can also assign a weight to the edges through a matrix W_{ij} whose elements specify the weight of the connection between i and j , and can assume any real value depending on the type of interaction is being described. By definition, $A_{ij} = 0$ implies $W_{ij} = 0$. For example, the weights can be seen as the number of messages sent by one user to another in a social network, in Chapter 6, or the flow of people from one city to other, as in Chapter 7. Figure. 2.1 shows graphical representations of vertices and edges, and classification of networks.

If there exists an edge between i and j , they are said to be connected or adjacent. In a directed network, if an edge points from i to j , i is called a predecessor neighbor of j and j is a successor neighbor of i . This concept is used in Sec. 6.3.2 of this thesis. For undirected networks, if i is

connected to j they are said to be neighbors.

The number of all possible pairs of vertices joined by edges in a network is $N(N - 1)/2$. A complete graph has all possible edges. We define the density of the network as $\mathcal{D} = \frac{E}{N(N - 1)}$, and the network is considered sparse if $\mathcal{D} \ll 1$. The sparseness of many networks has a big impact on computational efficiency, and it is useful to define adjacency lists with elements (i, j) , where $A_{ij} = 1$ and j belongs to $\mathcal{V}(i)$, the set of neighbors of i . Figure 2.2 shows an example.

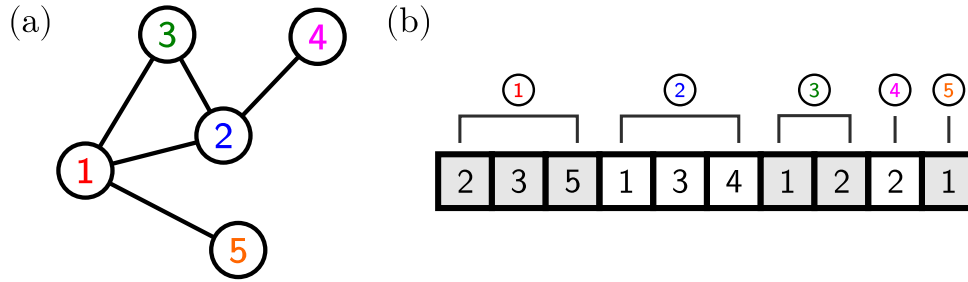


Figure 2.2: Example of (a) an undirected network and (b) its representation as an adjacency list. For each vertex i in (a), represented by circles, neighbors are shown as elements of the adjacency list in (b).

In Chapters 4 and 5 we will only deal with undirected and unweighted networks, while in Chapters 6 and 7 we will consider directed and weighted networks.

2.1.2 Connectivity and paths

A meaningful concept is the existence of paths in the network. A path is defined by the set of sequential edges that join a sequence of distinct vertices and its length is the number of edges in that path. If there exists a path from vertex i to j , we say that j is reachable from i , otherwise they are said to belong to different connected components. A connected component is defined by the set of vertices and edges in which there exists at least one path between all pairs of vertices in both directions. The largest connected component (LCC) contains the largest number of vertices.

Connected components

In Fig. 2.3(a) we show the components of an undirected graph and in (b) of a directed graph. Directionality plays an important role in the definition of a component. For example, while vertices $\{1, 2, 3, \dots, 8\}$ belong to the same component in Fig. 2.3(a), there are two different components in (b) for this set of vertices. This happens because even when a path connecting vertex 6 to 4 exists, 6 is not reachable from 4, for example. Another important fact is that

while in Fig. 2.3(a) the LCC is composed by edges containing vertices $\{9,10,\dots,18\}$, when we consider directionality, it is composed by vertices $\{1,2,\dots,5\}$. By this definition, in the case of directed networks, we also refer to the connected component as the strongly connected component (SCC). We can also say that Fig. 2.3(a) is the undirected network representation of Fig. 2.3(b), and the connected components of Fig. 2.3(a) are the weakly connected components of the directed network, Fig. 2.3(b).

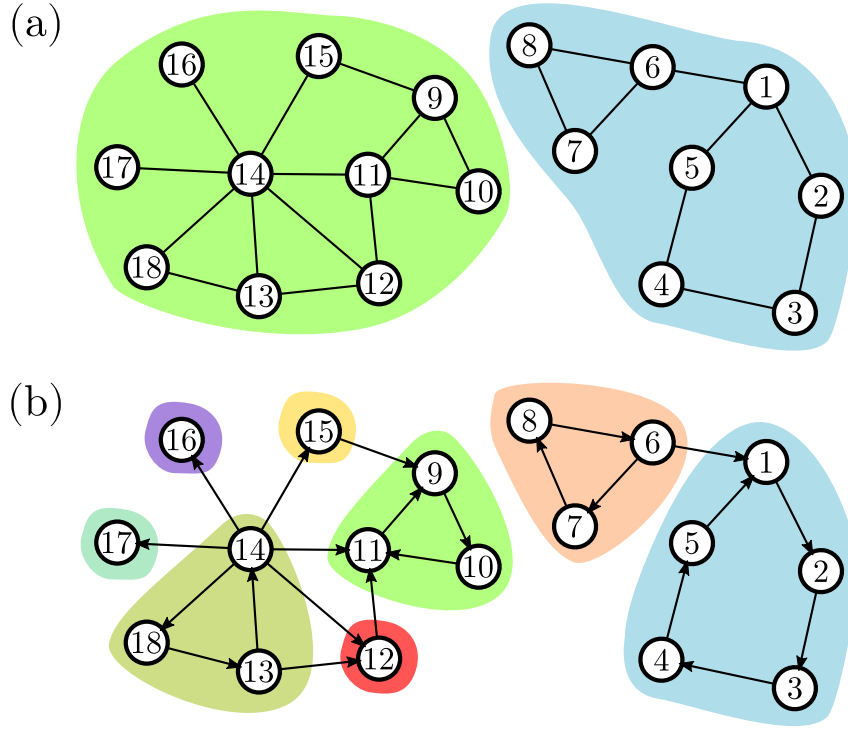


Figure 2.3: Example of connected components for (a) undirected and (b) directed networks. Shaded areas indicate vertices belonging to the same component. The graph in (a) is also called an undirected network representation of (b).

Distances and dimension of networks

Usually, there are many paths connecting vertices i and j . We define the distance ℓ_{ij} between i and j as the length of the shortest path between them, and in general it is not symmetrical in directed networks. If i and j belong to different connected components, we define $\ell_{ij} = \infty$. A measure of the size or dimension of the network is the average shortest path length, computed over all pairs of vertices $\{i,j\}$ belonging to the same connected component in the network,

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{\{i,j\}} \ell_{ij}. \quad (2.2)$$

In a complete graph, in which all vertices are connected to each other, we have $\langle \ell \rangle = 1$ while in d -dimensional hypercubic lattices it scales as $\langle \ell \rangle \sim N^{1/d}$. In most real networks, especially in random graphs, the average shortest path length scales logarithmically with the size N , $\langle \ell \rangle \sim \ln N$; see Sec. 2.2.1. This is a property of the so-called small-world phenomena and these networks are said to have an infinite dimension since an algebraic growth $N^{1/d}$ predicts an increase with N much faster than for any finite dimension d .

2.1.3 Centrality measures

Some characteristic properties of complex networks can be defined and many insights can be provided by specific vertex properties, characterizing their centrality in the network. Here we define some of the measures used in this thesis.

Degree centrality

In general, we define the in-degree $k_{\text{in},i}$ as the number of edges pointing to vertex i and the out-degree $k_{\text{out},i}$ as the number of edges departing from i . In terms of the adjacency matrix, we have

$$k_{\text{in},i} = \sum_j A_{ji} \quad \text{and} \quad k_{\text{out},i} = \sum_j A_{ij}, \quad (2.3)$$

respectively. In undirected networks $k_{\text{in},i}$ and $k_{\text{out},i}$ are equivalent and we only refer to those as the degree k_i of vertex i . This property tells how connected is a vertex in the network. The degree of a vertex has central implications in epidemic models, as will be discussed throughout this thesis.

k -core centrality

Another property related to the degree is the k -core centrality. It is given after a k -core decomposition [24, 67] that split the network in subgraphs $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, with $\mathcal{V}' \subset \mathcal{V}$ and $\mathcal{E}' \subset \mathcal{E}$, whose vertices have degree larger than an index k_s . This decomposition consists of a pruning process starting with the whole network. First, all vertices with degree $k_s = k_{\min}$ and their edges are removed. Then, we remove all other vertices that now have degree k_{\min} after the latter removal, until no more vertices of k_{\min} appear in the subgraph. Next, we repeat the procedure for all vertices of degree $k_s = k_{\min} + 1, k_{\min} + 2, \dots$, and so on until all vertices are removed. The maximal k -core corresponds to the subgraph with vertices and edges removed

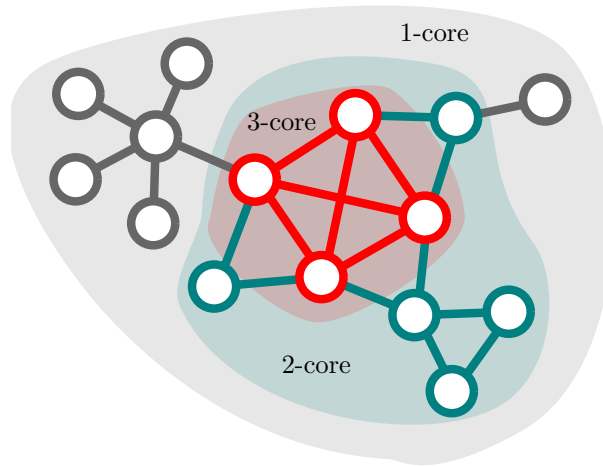


Figure 2.4: Example of a k -core decomposition. The 1-core consists of the whole network. Next, the 2-core is a subgraph of the 1-core and contains only vertices with degree $k \geq 2$. Finally, the 3-core only has vertices with degree $k \geq 3$, and it is a subgraph of the 2-core. It is impossible in this case to have a k -core with an index larger than 3. Note that vertices with smaller coreness are peripheral, while the ones in the maximum k -core are more central and densely connected.

in the last step of the decomposition, and its index k_s is called the coreness of the vertices belonging to this subgraph. In Figure 2.4 we show an example in an undirected network of $N = 15$ vertices and $k_{\min} = 1$, resulting in a maximum k -core with coreness 3.

Activity centrality

In weighted networks a significant measure is the vertex strength, defined in terms of weights of edges. The strength or activity centrality is defined analogously to the degree [66] as

$$a_i = \sum_j W_{ij}. \quad (2.4)$$

As an example, if the weight W_{ij} is the number of messages sent from an individual i to j in an online social network, a_i is the number of messages sent by i and

$$a_i^{\text{IN}} = \sum_j W_{ji} \quad (2.5)$$

is the number of messages received by i . One can expect that individuals with large activity are more central in information spreading. This will be discussed in Chapter 6.

2.1.4 Communities in networks

Even in sparse networks, we can find complete subgraphs in which all vertices are connected to each other. These subgraphs are called cliques of the network. On another level,

some local structures are the building blocks of larger subgraphs where vertices are cohesively interconnected but badly connected to others outside them, defining communities.

Clustering coefficient

Analyzing the local neighborhoods in a graph, many real networks present vertices that tend to form cliques among themselves. In a social network, for example, if an individual i is connected to j , and j is connected to l , there is a high frequency that i is also connected to l . This connection can represent a friendship, for example. We can measure this property by the clustering coefficient, defined [68], in an undirected network, as the ratio between the number of connected pairs among neighbors of i , denoted by e_i , and the maximum number of possible pairs between them. Then,

$$C_i = \frac{2e_i}{k_i(k_i - 1)}, \quad \text{if } k_i > 1. \quad (2.6)$$

For $k_i = 1$, $C_i \equiv 0$. Figure 2.5 shows an example of the clustering coefficient calculation. In general the average clustering coefficient

$$\langle C \rangle = \frac{1}{N} \sum_i C_i$$

is used to measure the cohesiveness of the network. Real networks often have a high average clustering coefficient.

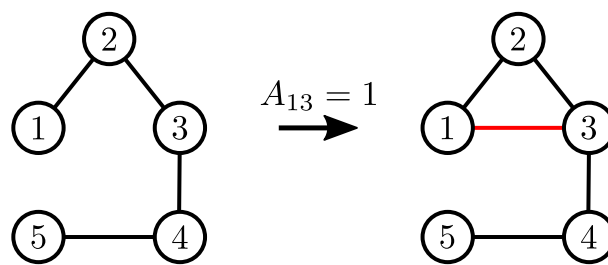


Figure 2.5: Left: clustering coefficients of all vertices are zero. Right: when vertex 1 connects to 3, by making $A_{13} = 1$, a triangle is formed between 1, 2 and 3. Since all neighbors of 1 and 2 are connected to each other, $C_1 = C_2 = 1$. For vertex 3, its neighbor 4 does not have any connection neither to vertex 1 nor 2, while 1 and 2 are connected, resulting in $C_3 = 1/3$. Vertices 4 and 5 have $C_4 = C_5 = 0$.

Modularity

Some systematic local density fluctuations [2] can appear in networks, defining what we interpret as communities, which are locally denser connected subgraphs that do not appear

in randomly wired networks. The community structure is encoded in the adjacency matrix, partitioning the set of N vertices into M communities or modules g . We denote the group to which the vertex i belongs as g_i . The modularity [1] is defined as

$$Q_{\text{mod}} = \frac{1}{N\langle k \rangle} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{N\langle k \rangle} \right) \delta_{g_i g_j}, \quad (2.7)$$

being δ_{ij} the Kronecker delta and $\langle k \rangle$ the average degree defined in the next section. Q_{mod} is always smaller than 1, and the higher is its value, the better partitioned is the set $\mathcal{M} = \{g_1, g_2, \dots, g_N\}$. If one can find the optimal community structure, the maximum value of Q_{mod} is obtained. The computational complexity for the evaluation of this optimal structure is too high and there are many different algorithms to this aim [2]. We use the Louvain algorithm [69] in Chapters 5 and 6 to determine modular structures.

2.2 Statistical characterization of complex networks

With the contemporary possibility of collecting and analyzing big datasets to build several large-scale networks, many properties cannot be observed only with local measures, and statistical analyses are needed. Here, we comment on some of these statistical characterizations and their relations with the way vertices connect to each other.

2.2.1 Degree distributions

Important insights are obtained by analyzing the degree distribution $P(k)$ defined as the probability that a randomly selected vertex has degree k . If N_k is the number of vertices with degree k in a network of N vertices, then

$$P(k) = \frac{N_k}{N}. \quad (2.8)$$

It can be defined in directed networks for k_{in} and k_{out} in analogous way. The moments of the degree distributions are essential to understand critical regimes in dynamical processes such as those described in Chapter 4 and 7. The n -th moment of the degree distribution is defined as

$$\langle k^n \rangle = \sum_k k^n P(k). \quad (2.9)$$

In the case of sparse networks, $\langle k \rangle \ll N$. In particular, the second moment $\langle k^2 \rangle$ is associated to the variance $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, that measures how broad is the distribution. Homogeneous distributions, shown in Fig. 2.6(a), have small variance (comparable with $\langle k \rangle$), while heterogeneous distributions, such as represented in Fig. 2.6(b), have large variance with heavy tails. That means that heterogeneous networks have *hubs*, that are vertices whose degrees are much bigger than the average. Then, although the probability to find a hub of degree $k \gg \langle k \rangle$ is small, its contribution is not negligible.

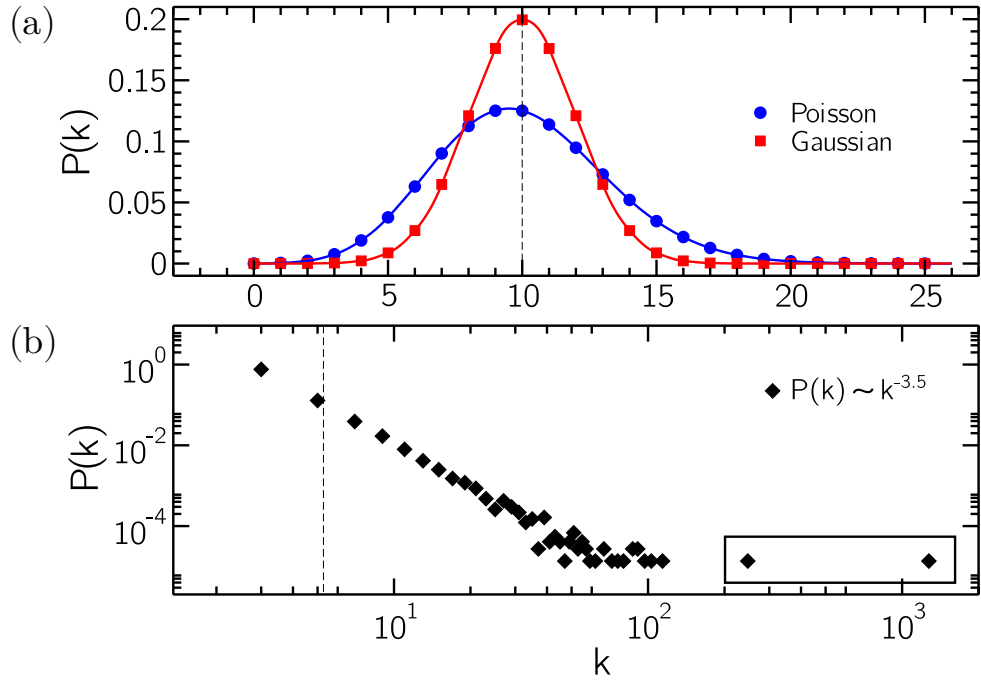


Figure 2.6: (a) Examples of homogeneous Poisson and Gaussian distributions, and (b) heterogeneous power-law degree distribution with $\gamma = 3.5$, with outliers indicated inside the box. The dashed lines represent the averages $\langle k \rangle$ of the distributions, and hubs, for which $k \gg \langle k \rangle$, only appear in (b). Note that axes are in (a) linear and (b) logarithmic scales.

A power-law (PL) distribution follows $P(k) \sim k^{-\gamma}$, with $k \in [k_{\min}, k_{\max}]$, and has moments that strongly depend on the exponent γ . The minimum degree of the network is denoted by k_{\min} and the maximum is denoted by k_{\max} . Cutoffs can be fixed for the distribution, defining the minimum (k_0) and maximum (k_c) degrees allowed when sampling from $P(k)$. For $\gamma < 2$, the average $\langle k \rangle$ diverges as $k_c \rightarrow \infty$. Making a continuous approximation, we can evaluate the moments $\langle k^n \rangle$ replacing summations by integrals in the form

$$\langle k^n \rangle \simeq \int_{k=k_0}^{k_c} A k^n k^{-\gamma} dk, \quad (2.10)$$

where A is a normalization constant given by $\int A k^{-\gamma} dk = 1$. Considering a natural cutoff of

the size of the network, $k_c = N$ and $N \rightarrow \infty$, the average degree is well defined and given by

$$\langle k \rangle \approx k_0 \frac{\gamma - 1}{\gamma - 2} \quad (2.11)$$

for $\gamma > 2$, while the second moment $\langle k^2 \rangle$ diverges with $N \rightarrow \infty$ for $2 < \gamma < 3$. For $\gamma > 3$, the second moment is finite and given by

$$\langle k^2 \rangle \approx k_0^2 \frac{\gamma - 1}{\gamma - 3}. \quad (2.12)$$

However, even with a finite variance, the probability to find an outlier still exists. Indeed, if we consider a PL distribution with $k_c = N$, using extreme value theory we can show that the mean value of the largest degree k_{\max} scales as [70]

$$\langle k_{\max} \rangle \sim N^{\frac{1}{\gamma - 1}}, \quad (2.13)$$

with large fluctuations [71]. Outliers are hubs that besides having degree $k \gg \langle k \rangle$, their values are not common among hubs. Figure 2.6(b) shows a PL degree distribution with outliers. PL distributions for other cutoffs can be seen in Fig. 5.3, in Chapter 5. Networks presenting PL degree distributions are said to be scale-free, and many real networks follow this distribution for different values of γ , such as those in Fig. 2.7. These hubs and heavy tails can have strong effects on spreading processes on networks [8, 72], and their roles are discussed in all chapters of this thesis.

With the definition of average degree, we can perform a simple calculation [2] to find $\langle \ell \rangle$. Consider an undirected network whose N vertices are connected randomly and have average degree $\langle k \rangle$. We expect to find $\langle k \rangle$ vertices at distance $\ell = 1$, $\langle k \rangle^2$ vertices at distance $\ell = 2$, and $\langle k \rangle^\ell$ vertices at distance ℓ . Then, the expected number of vertices $n(\ell)$ from a distance ℓ of the starting vertex is

$$n(\ell) \approx \sum_{i=0}^{\ell} \langle k \rangle^i = \frac{\langle k \rangle^{\ell+1} - 1}{\langle k \rangle - 1}, \quad (2.14)$$

and $n(\ell)$ cannot be larger than the total number of vertices N . Assuming that $\langle k \rangle \gg 1$, the maximum distance ℓ_{\max} will obey the condition

$$n(\ell_{\max}) \approx N, \quad (2.15)$$

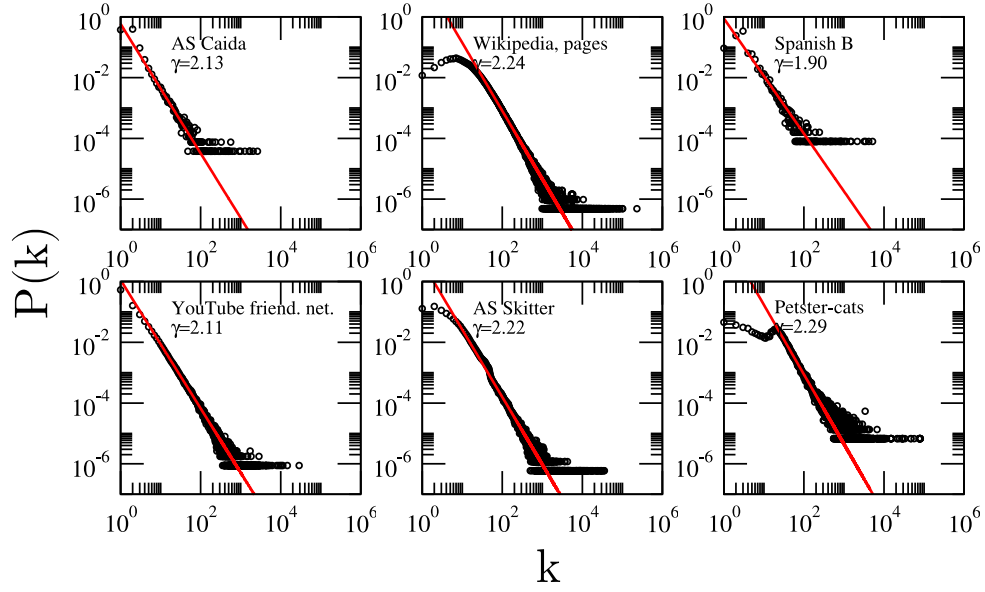


Figure 2.7: Degree distributions for some real networks, PL fits, and their respective exponents γ . The complete list and references for these networks are found in Ref. [73].

implying that

$$\ell_{\max} \approx \frac{\ln N}{\ln \langle k \rangle}. \quad (2.16)$$

Since ℓ_{\max} is often dominated by few large paths, the latter expression gives an approximation for the average distance between any pair of vertices, implying in $\langle \ell \rangle \sim \ln N / \ln \langle k \rangle$. It means that in a network of $N = 10^{10}$ vertices with $\langle k \rangle = 10$, we can connect any pair of vertices, in average, by a path of length $\langle \ell \rangle \approx 10$.

2.2.2 Mixing and correlations

The way vertices are connected can be correlated, in the sense they tend to connect to other vertices with specific properties. This correlation can be with the degree, which can be measured by the probability $P(k'|k)$ that a vertex of degree k is connected to another of degree k' . This probability is associated with another important measure, the average degree of the neighbors of i , given by

$$k_{\text{nn},i} = \frac{1}{k_i} \sum_j k_j A_{ij}.$$

Statistically, it is more informative to take the average over all vertices with the same degree k ,

$$k_{\text{nn}}(k) = \frac{1}{N_k} \sum_{i|k_i=k} k_{\text{nn},i},$$

where N_k is the number of vertices with degree k and the summation is performed over all vertices i of degree $k_i = k$. By means of the probability $P(k'|k)$, we can write

$$k_{nn}(k) = \sum_{k'} k' P(k'|k). \quad (2.17)$$

If there are no correlations, this probability is given by the probability of finding an edge emanating from a vertex of degree k' , that is the ratio between the number of edges from vertices with degree k' over the total number of edges. So, $P(k'|k)$ for the uncorrelated case is given by

$$P(k'|k)^{[unc]} = \frac{k' N_{k'}}{\sum_{k''} k'' N_{k''}} = \frac{k' P(k')}{\langle k \rangle}, \quad (2.18)$$

independent of k , implying that

$$k_{nn}^{[unc]}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}, \quad (2.19)$$

which is constant and only depends on the first and second moments of the degree distribution. However, networks can have other types of correlations: assortative and disassortative. In the former case, vertices with a low (high) degree connect more often to others with a low (high) degree, while in the latter the ones with a low (high) degree connect more often with others with a high (low) degree.

With a similar reasoning, a statistical measure can be used in term of the clustering coefficient as a function of the degree, defined as

$$C(k) = \frac{1}{N_k} \sum_{i|k_i=k} C_i.$$

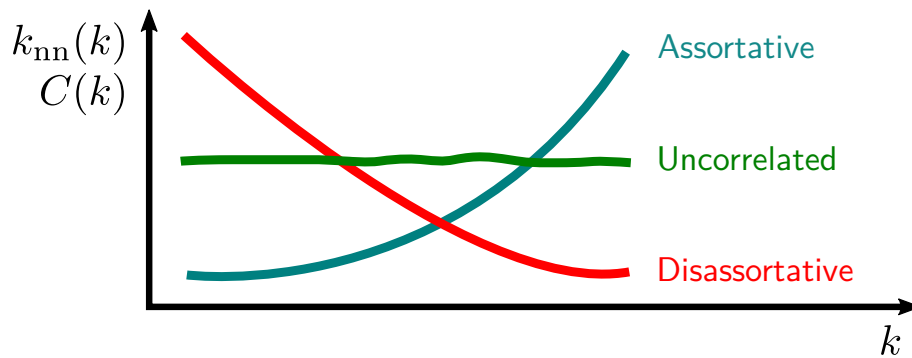


Figure 2.8: Schematic representation of three different types of correlations in networks. If the curves are increasing (decreasing), the correlation is said to be assortative (disassortative). If the value is practically uniform, the network is uncorrelated.

Correlations in this expression can be related to the presence of structures as three-vertex correlation pattern and can measure the existence of a hierarchy in the network [74, 75], expressed by $C(k) \sim k^{-1}$. See Chapter 5 for more information about the role of hierarchy in criticality.

For both k_{nn} and C as a function of k , the three basic patterns of correlations are schematically drawn in Fig. 2.8.

2.3 Synthetic networks

Throughout this thesis, we use synthetic networks to explain concepts and to perform analysis of physical phenomena. Some of these networks are simple to define, as the complete graph. Others are built by following some rules and a complex network emerges from them.

2.3.1 Simple networks

In Fig. 2.9 we present some simple networks used in this thesis. The first is a complete graph, already defined before, in which all vertices are connected to each other. The second is a star graph, that consists of a vertex with k leaves around with degree 1, mimicking the hubs of a network. When in a star graph the leaves are also connected, it can be called a wheel graph, the third case. Finally, a random regular network (RRN) is composed of all vertices having the same degree but randomly connected, in Fig. 2.9(d).

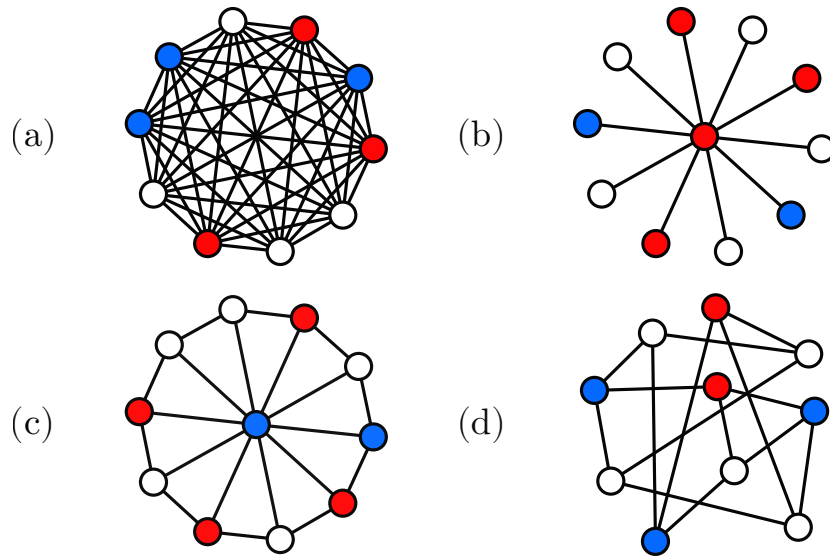


Figure 2.9: Examples of simple networks: (a) complete graph of size $N = 10$, (b) star graph and (c) wheel graph with $k = 10$ leaves, and (d) RRN network with $k = 3$ and $N = 10$.

2.3.2 Configuration model

There are many models to build complex networks, such as the model of random networks [76], better known as Erdős–Rényi [77], the Watts–Strogatz [68], and Barabási–Albert [78] models. They are famous to explain in simple grounds the emergence of the small-world property, the high clustering coefficient, and PL degree distributions, respectively, in many real networks. It is worth to mention that the term “complex” in complex networks comes from the fact that these macroscopic properties emerge from single local rules of connectivity.

However, in this thesis we will use the configuration model [79–82], that generates synthetic undirected and unweighted networks from a predetermined degree distribution $P(k)$ with cutoffs k_0 and k_c . First, each one of the N vertices receives a degree k by sampling $P(k)$. We refer to edges not connected as stubs, in a way that a vertex i has k_i stubs at the beginning of the process. These stubs are randomly selected in pairs and connected if the vertices in each end are not the same and have not been connected before, avoiding self-connections and multiple connections, respectively; see Fig. 2.10. For $P(k) \sim k^{-\gamma}$, the large value of $\langle k^2 \rangle$ can introduce correlations in the network [70, 82]. In order to avoid them, it is necessary to assign a stronger cutoff k_c to the degree distribution. The uncorrelated configuration model (UCM) is defined by using the configuration model in a PL degree distribution $P(k) \sim k^{-\gamma}$ with a structural cutoff $k_c = \sqrt{N}$, for which it guarantees that k_{nn} is approximately constant for $\gamma > 2$ [82].

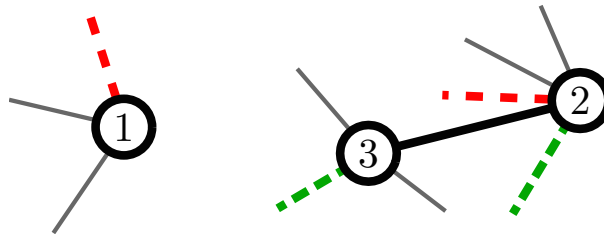


Figure 2.10: Example of trials in the configuration model. These three of N vertices have degrees $k_1 = 3$, $k_2 = 5$ and $k_3 = 4$. When two stubs in red are selected, the connection between them is allowed since 1 is not connected to 2. However, if the stubs in green are selected, they cannot be connected since 2 is already connected to 3. An edge connects 2 and 3, and there are other stubs (in grey) that can be connected to other $N - 3$ vertices of the network.

2.3.3 Annealed networks

So far we have assumed that networks do not evolve with time, which we call quenched networks. On the opposite extreme lay the annealed networks, in which the connections are rewired at a rate much larger than the rates of the processes taking place on them [24]. We

can define the adjacency matrix for this kind of network as the probability that a vertex i is connected to j in a given time [24]. If vertex i has k_i stubs, in an uncorrelated model a stub can be momentarily connected with any other stub with equal chance [71]. In this case, the elements of the adjacency matrix are given by

$$A_{ij}^{\text{ann}} = \frac{k_i k_j}{\sum_{l=1}^N k_l} = \frac{k_i k_j}{N \langle k \rangle}. \quad (2.20)$$

The computer implementation of uncorrelated annealed networks consists in keeping a list with $\sum_{i=1}^N k_i$ elements containing k_i copies of each vertex i . The sampling of a neighbor of a vertex is done by choosing randomly one element of this list.

2.4 Summary

In this chapter, we defined useful properties and measures in network structures. There are many other concepts and properties not covered here and can be read in textbooks about network theory such as those in Refs. [1, 2, 6, 66].

Chapter 3

Spreading processes on complex networks

Related publication:

Optimized Gillespie algorithms for the simulation of Markovian epidemic processes on large and heterogeneous networks [34]

Wesley Cota, Silvio C. Ferreira

Computer Physics Communications **219C** (2017) pp. 303-312

Statistical physics has been dedicated to studying how interactions between different elements at a microscopic level emerge to macroscopic phenomena. Many systems can go through phase transitions, that are featured by singular macroscopic changes in the order parameter, such as in the transition from liquid to the gas state by increasing the temperature of the water. Models of dynamical processes are developed to describe the time evolution of systems by using, for example, their network of interactions as a substrate for these processes. In a disease spreading process, local rules of contact between individuals can lead to epidemics, in which a macroscopic fraction of the system is infected as a non-linear result of these interactions. In this chapter, we provide a basic introduction to the theory and simulation procedures of dynamical processes discussed in this thesis.

3.1 Master equations for dynamical processes

The time evolution of the macroscopic state ω of a system can be described by means of microscopic states ω_i for each element i , such as the vertices of a network. Without loss of generality, we assume that there are Ω possible states for each vertex, $\omega_i = 1, 2, \dots, \Omega$. The state variables of all vertices of a network define, therefore, the microstate of the system, described by the set $\omega = \{\omega_1, \omega_2, \dots, \omega_N\}$, in which N is the size of the network. The temporal evolution

of the system will be given by transitions $\omega \rightarrow \omega'$, with $\omega \neq \omega'$. It can be impossible to follow completely the microscopic dynamics of big systems, with large complexity in relation to the number of variables and stochasticity. A possibility is to introduce master equations (MEs), evaluating the probability $P(\omega, t)$ of a system to be in a state ω at time t . Then, the temporal evolution can be described by

$$\frac{dP(\omega, t)}{dt} = \sum_{\omega'} [P(\omega', t)W_{\omega' \rightarrow \omega} - P(\omega, t)W_{\omega \rightarrow \omega'}], \quad (3.1)$$

with a normalization condition $\sum_{\omega} P(\omega, t) = 1$. The summation is over all possible configurations and terms $W_{\omega' \rightarrow \omega}$ represent transition rates from a configuration ω' to ω . The first term in the right hand side (r.h.s.) of Eq. (3.1) is related to transitions from states ω' to ω , while the second represents transitions from state ω to ω' . In principle, the solution of the ME allows the computation of average values at time t of a quantity $A(\omega)$, given by

$$\langle A(t) \rangle = \sum_{\omega} A(\omega) P(\omega, t). \quad (3.2)$$

In some cases, the global rate $W_{\omega' \rightarrow \omega}$ can also be written in a simpler way. Considering that the change of state of a vertex i only depends on local interactions of vertices connected to it, we can write [6]

$$W_{\omega' \rightarrow \omega} = \prod_i w_{\omega'_i \rightarrow \omega_i | \omega_j},$$

where the terms in the r.h.s. are transition rates from state ω'_i to ω_i conditioned to the state ω_j , where $j \in \mathcal{V}(i)$ is a neighbor of i . In this case we consider that the parameters for the local dynamics are the same for all vertices. Here we see explicitly the dependence of dynamical processes with the structure of the network mediating the interactions between the elements of a system.

It is not always possible to solve MEs, and analytical approximations have to be compared with statistical exact results. Among the methods, there are the mean-field approximation, that consists of using average values of the microscopic variables under some simplifying hypothesis, and computer simulations, in which random numbers are used associating probabilities to transition rates and Monte Carlo (MC) steps [83]. In this context, the Gillespie algorithm [84, 85] for Markovian processes appears. Hereafter we will only consider Markovian processes, whose transitions depend exclusively on the state reached in the previous event at time t , without

dependence on past events.

3.1.1 Equilibrium and non-equilibrium systems

There are processes in which the MEs can result in a stationary probability $P(\omega, t \rightarrow \infty) = P_{\text{eq}}(\omega)$. These processes describe equilibrium systems, in which the detailed balance condition

$$P_{\text{eq}}(\omega)W(\omega \rightarrow \omega') = P_{\text{eq}}(\omega')W(\omega' \rightarrow \omega) \quad (3.3)$$

is obeyed. In this case, the r.h.s. of Eq. (3.1) is zero, describing necessarily an equilibrium state. However, there are non-equilibrium systems that can be out of the detailed balance condition and some subset of configurations may be preferred. The detailed balance condition is a sufficient but not necessary condition for equilibrium. Indeed, even when the condition in Eq. (3.3) does not hold, a stationary state can exist. We will denote this asymptotic limit as $P^\infty(\omega) = \lim_{t \rightarrow \infty} P(\omega, t)$, where ω is a stationary state of the system. When the system has a finite number of states, the Markovian nature of the processes guarantees the unicity of the stationary state [86].

Differently of equilibrium phenomena, in which one can evaluate the partition function of the system, most dynamical processes in nature are far from equilibrium and we cannot use the same formulation of the equilibrium statistical mechanics. However, analogous concepts can be used, such as those of phase transitions and criticality. Before proceeding, it is useful to define Poisson processes, that form the basis for the modeling of Markovian processes.

3.1.2 Poisson processes and the Gillespie algorithm

In a scenario where we need to count the occurrences of certain events we can define a rate ν such that the probability of an event to happen in a time interval dt is νdt . This type of event is described by a Poisson process. Let $P_0(t)$ be the probability that no event occurred until time t , and the complementary probability $P_1(t) = 1 - P_0(t)$ that it did. The ME of this process is given by

$$\begin{aligned} \frac{dP_0}{dt} &= -\nu P_0, \\ \frac{dP_1}{dt} &= \nu P_0. \end{aligned} \quad (3.4)$$

Considering an initial condition $P_j(0) = \delta_{j0}$, such that no event has happened until time $t = 0$, the solution of Eq. (3.4) is

$$P_0(t) = e^{-vt}, \quad (3.5a)$$

$$P_1(t) = 1 - e^{-vt}. \quad (3.5b)$$

Now let us consider that there are Z independent Poisson processes $p = 1, 2, \dots, Z$, each happening at a rate v_p . The probability that the next event will be p and that it happens within the time interval $[t + \tau, t + \tau + dt]$ is given by [34]

$$\begin{aligned} W(\tau, p)dt &= \overbrace{\left(\prod_{q=1}^Z e^{-v_q \tau} \right)}^{\text{no event has occurred}} \cdot \underbrace{v_p dt}_{p \text{ occurs}} \\ &= \left(\frac{v_p}{R} \right) (R e^{-R\tau}) dt \end{aligned} \quad (3.6)$$

with $R = \sum_p v_p$ being the total rate of transitions. Then, it is possible to split the problem in two parts: the next event will occur after a time τ given by a distribution

$$P_R(\tau) = R e^{-R\tau}, \quad (3.7)$$

where $\langle \tau \rangle = R^{-1}$, and the event will be p with probability v_p/R . Thus, a set of different Poisson processes can be described by only one exponential distribution with rate R .

Based on these ideas the GA algorithm is proposed as follows:

1. Build a list with all processes and their respective rates;
2. Select the time step size τ from the exponential distribution $P_R(\tau)$ as $\tau = -\ln(u)/R$, where u is a pseudo random number uniformly distributed in the interval $(0,1]$;
3. Choose the process p to take place with probability v_p/R , implement it and update the state of the system;
4. Increment time as $t \rightarrow t + \tau$;
5. Return to step 1.

In step 2, the value $u = 0$ must be strictly forbidden since it leads to an infinity time step. With

this procedure, it is possible to evaluate averages $\langle A(\omega) \rangle$ with the same statistical properties as those given by the original MEs.

3.2 Epidemics in networks as an absorbing-state phase transition

In epidemic spreading, some individuals can be infected and transmit the disease to others following some rates and interactions among vertices in a network. In closed systems, states where the disease is eradicated are called absorbing since once one of them is visited, the dynamics remains frozen in such a state forever, breaking the detailed balance condition. Therefore, a state without infected individuals is an absorbing one, since infection can only be produced through interactions involving infected and susceptible pairs and defines a non-equilibrium process.

3.2.1 Compartmental models for epidemics

Different types of contagion can be described by epidemic processes on complex networks, such as [2] the propagation of pathogens by physical contact between individuals, like in flu, Ebola, COVID-19, and other infectious diseases; computer viruses in the internet, with infinite reproductive capacity and infection patterns similar to biological ones [87]; and spreading of information, behaviors, rumors, and news on the internet. Information spreading dynamics is investigated in Chapter 6. All these processes involve agents that spread something by using a network as substrate and share dynamical properties [6]. In general, we can assume that the states of each vertex can be given by a finite set of compartments [88–90], describing individuals that are in the same stage of a disease; see Fig. 3.1. Here we define the most used compartments:

Susceptible (S): vertices that have not been in contact with the pathogen and can be infected;

Infected (I): infectious vertices, that can transmit the disease to its contacts; and

Recovered (R): vertices that were infected before, but are neither susceptible nor infectious.

Rates and rules of interactions that can lead individuals from one compartment to another depend on the epidemic process considered. Other compartments can be used, improving epidemic models [90]. We will denote susceptible vertices by S and $\omega_i = 0$, infected by I and $\omega_i = 1$ and recovered by R and $\omega_i = 2$.

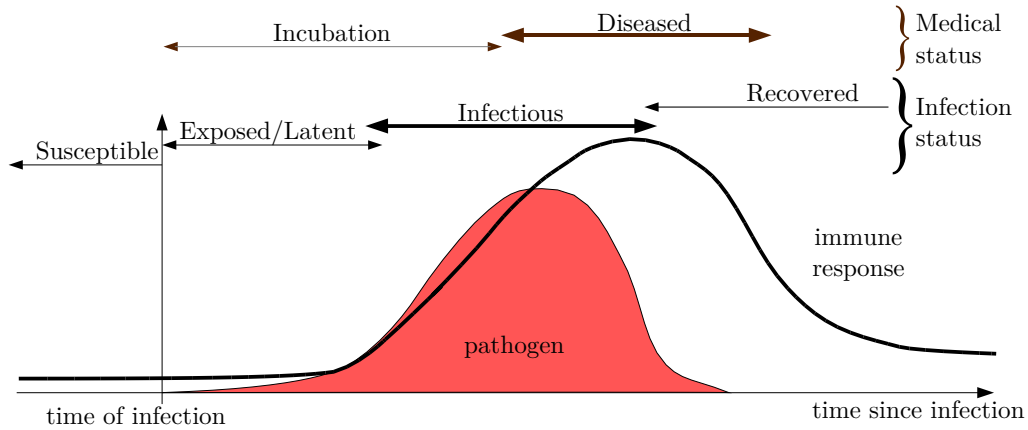


Figure 3.1: Timeline of an infection, showing the pathogen concentration (shaded area) and the immunological response (black curve) in different stages by the medical or epidemiological states of a host. Even when symptoms are shown and the individual is sick, it can be already recovered by the epidemiological point of view. Figure adapted from Ref. [90].

In this thesis, we investigated processes that have a transition to an absorbing-state, from which the dynamics do not evolve. A paradigmatic epidemic model exhibiting an absorbing-state phase transition is the susceptible-infected-susceptible [91] (SIS) model. Its rules are the following. Vertices can be infected or susceptible. The infected individuals spontaneously heal with a rate μ . An infected vertex i can transmit the disease to a susceptible neighbor j with a rate $\lambda_{ij} = \lambda A_{ij}$, which means that an infected vertex infects each one of its susceptible neighbors with rate λ irrespective of how many it has. This infection rule has deep impacts on the behavior of the SIS model. A noticeable one is the absence of a finite epidemic threshold for a PL degree distribution as the network size goes to infinity [11, 15]. However, thresholds of the SIS model can still be numerically determined for finite sizes [16, 26]; see Sec. 3.2.4. Investigations about criticality in the SIS model are made in both chapters of Part II.

It is important to mention that the SIS model is similar to the contact process (CP) in regular d -dimensional hypercubic lattices. The CP was initially proposed as an interacting particle system involving self-annihilation and catalytic creation of particles. A vertex i can be occupied ($\omega_i = 1$) or vacant ($\omega_i = 0$). Particles are destroyed spontaneously with rate μ , and a particle in i can replicate in a vacant neighbor j with rate $\lambda_{ij} = \lambda A_{ij}/k_i$. Notice that in CP the average number of infections by unit of time is bounded by λ , while in SIS by λk_{\max} , where k_{\max} diverges on PL networks, as discussed in Chapter 2. Due to this simple difference, CP and SIS are very different dynamical processes when investigated in heterogeneous networks.

3.2.2 SIS dynamics on a star graph

Let us exemplify the GA with the important problem of the SIS dynamics on a star graph, which is composed by $j = 1, \dots, k$ vertices of degree 1, called leaves, connected to a vertex of degree k , the center, as defined in the previous chapter. The state is determined by the number of infected leaves and the state of the center. Let $P(n, \omega)$ be the probability that there are $n = 0, 1, \dots, k$ infected leaves and the center is in either the states $\omega = 0$ (susceptible) or 1 (infected). The transitions and the respective rates are

$$\begin{aligned}
 (n, 0) & \xrightarrow{\lambda n} (n, 1) \\
 (n, 0) & \xrightarrow{\mu n} (n-1, 0) \\
 (n, 1) & \xrightarrow{\lambda(k-n)} (n+1, 1) \\
 (n, 1) & \xrightarrow{\mu} (n, 0) \\
 (n, 1) & \xrightarrow{\mu n} (n-1, 1).
 \end{aligned} \tag{3.8}$$

The ME for this process is a set of $2k$ equations that we numerically integrated using fourth-order Runge-Kutta method [92]. This stochastic dynamics can be decomposed into four independent spontaneous events: a leaf is healed with rate $v_1 = \mu n$; the center is healed with rate $v_2 = \mu \omega$; the center is infected with rate $v_3 = \lambda n(1 - \omega)$; and a leaf is infected with rate $v_4 = \lambda(k - n)\omega$. In Fig. 3.2, we compare the probability distribution $P(n, \omega)$ obtained using GA

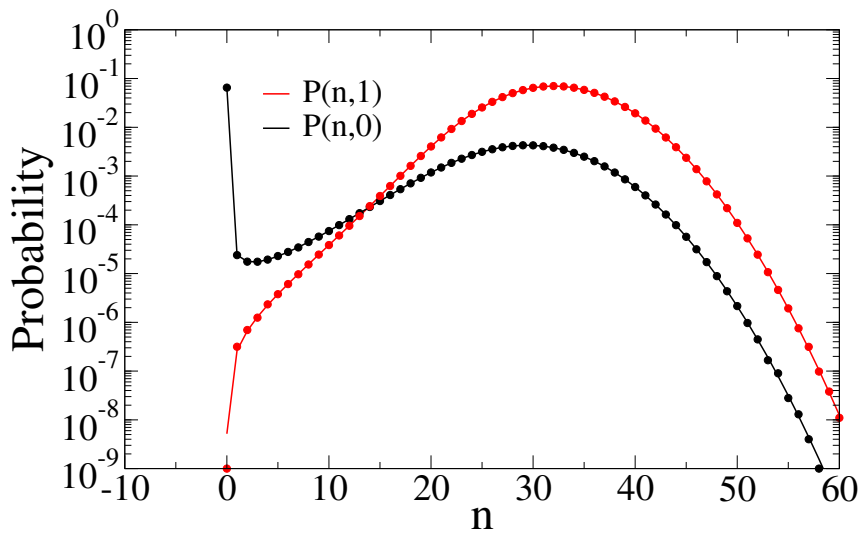


Figure 3.2: Probability distribution for SIS dynamics on a star graph with $k = 100$ leaves at $t = 10$ for an infection rate $\lambda = 0.5$ and $\mu = 1$. The initial condition is the center infected and all leaves susceptible. Symbols are GA simulations (10^7 samples) and lines are numerical integrations of the ME.

simulations with the numerical integration of the ME, showing that the results are completely equivalent, as expected.

3.2.3 Critical exponents

In order to illustrate the epidemic spreading as an absorbing-state phase transition (APT), let us start with the SIS on an arbitrary graph of size N and adjacency matrix elements A_{ij} . The total infection rate of the network is given by

$$L = \lambda \sum_{i,j=1}^N A_{ij} \omega_i (1 - \omega_j), \quad (3.9)$$

while the total healing rate is

$$M = \mu \sum_{i=1}^N \omega_i = \mu N_I, \quad (3.10)$$

where N_I is the number of infected vertices. Now, let us consider the simple case of an annealed network with a homogeneous degree distribution $P(s) = \delta_{sk}$, that leads to the annealed adjacency matrix $A_{ij}^{\text{ann}} = k/N$, which is introduced in Eq. (3.9) to obtain a total infection rate

$$L = \lambda k N_I \left(1 - \frac{N_I}{N}\right). \quad (3.11)$$

Therefore, both total rate of infection and healing of this model are functions of the number of infected vertices and does not depend on the specific state.

Let $P_n(t)$ be the probability that there are n infected individuals at time t and W_{nm} the transition rate from a state with n to another with m infected vertices. The time evolution of $P_n(t)$ is given by the ME

$$\frac{dP_n}{dt} = \sum_{m=0}^N [P_m W_{mn} - P_n W_{nm}], \quad (3.12)$$

with rates

$$W_{n,n+1} = g(n) = \lambda k n \left(1 - \frac{n}{N}\right) \quad (3.13)$$

and

$$W_{n,n-1} = r(n) = \mu n. \quad (3.14)$$

The transition rates are zero otherwise.

Plugging Eqs. (3.13) and (3.14) in (3.12) leads to

$$\frac{d\langle n \rangle}{dt} = \langle g(n) \rangle - \langle r(n) \rangle, \quad (3.15)$$

where $\langle f(n) \rangle = \sum_n f(n)P_n$. Neglecting fluctuations by assuming that $\langle g(n) \rangle \approx g(\langle n \rangle)$, the density of infected vertices defined by $\rho = \langle n \rangle / N$ evolves as

$$\frac{d\rho}{dt} = (\lambda k - \mu)\rho - \lambda k \rho^2, \quad (3.16)$$

which is easily solved providing the stationary solution

$$\rho_s = \begin{cases} \frac{\lambda k - \mu}{\lambda k} \sim (\lambda - \lambda_c)^\beta & , \quad \lambda \geq \mu/k \\ 0 & , \quad \lambda < \mu/k \end{cases}, \quad (3.17)$$

defining an epidemic threshold $\lambda_c = \mu/k$ and an exponent $\beta = 1$. The r.h.s. of Eq. (3.17) defines a critical exponent β associated with the vanishing of the order parameter. The typical scenario of an APT is illustrated in Fig. 3.3 in terms of the fraction of infected individuals.

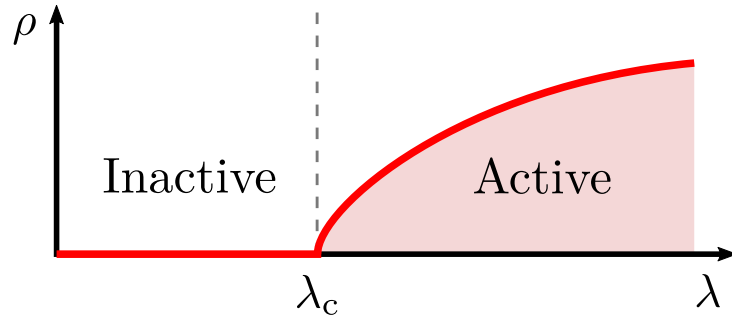


Figure 3.3: Schematic representation of a standard APT where the absorbing or inactive phase is characterized by a null fraction of infected vertices while the active has a non-zero value.

For $\lambda \neq \lambda_c$, the average density approaches its stationary value as [25]

$$\langle \rho \rangle - \rho^* \sim \exp\left(-\frac{t}{\tau_{\text{rlx}}}\right) \quad (3.18)$$

where

$$\tau_{\text{rlx}} \sim |\lambda - \lambda_c|^{-\nu_{\parallel}} \quad (3.19)$$

is the characteristic relaxation time, where the exponent ν_{\parallel} gives the divergence of the relaxation time near to the critical point. At $\lambda = \lambda_c$, the density of infected individuals asymptotically

scales as [25]

$$\langle \rho \rangle \sim t^{-\alpha}, \quad (3.20)$$

where α is another critical exponent [93]. Other exponents also appear in the temporal evolution of critical spreading processes, with relations for the survival probability $P_s(t) \sim t^{-\delta}$ and the number of active or infected vertices $N_a(t) \sim t^\eta$, for example, which are discussed in Sec. 5.3.3.

The existence of critical exponents describing phase transitions in these systems is not exclusive to APTs. In fact, there are distinct processes that, despite their dynamical and structural differences, share the same critical behaviors. When different processes have the same set of critical exponents, they are said to belong to the same *universality class*. A very important one is the directed percolation (DP) [25] universality class, of which many processes with APTs, such as the CP, belong to. There are other critical exponents that will appear throughout this thesis and scaling relations among them can be derived exploring symmetry properties and scale invariance [94]. For example, the survival probability $P_s(t)$ follows a similar behavior of Eq. (3.17) for $t \rightarrow \infty$,

$$P_s \sim |\lambda - \lambda_c|^{\beta'}, \quad (3.21)$$

with critical exponent β' . The scaling relation $\beta = \beta'$ holds under rapidity-reversal symmetry [93]. Other exponents related to the fluctuations of the order parameter ρ will be introduced with the aid of the quasi-stationary state. See further discussion of the SIS critical behavior in Part II.

3.2.4 Quasi-stationary state

Despite of the approximated solution neglecting fluctuations predicted an APT, one can easily verify that a stationary (normalized) solution of the master equation (3.12) with rates given by Eqs. (3.13) and (3.14) is $P_n = \delta_{n0}$, irrespective of λ and μ . For a finite system, the stationary solution is unique if W_{nm} forms an irreducible matrix [86], which is the present case. So, for finite sizes, this epidemic always visits the absorbing-state at a sufficiently long time since the unique true stationary state is the absorbing one. This is a universal feature of closed systems with absorbing-states [25]. Close to the epidemic threshold dynamical fluctuations become too high and these effects are computationally impossible to be ignored. To overcome these difficulties, we can use quasi-stationary values: averages are evaluated only using samples that have not visited the absorbing-state up to a time t .

To introduce the QS state, consider a Markovian process X_t^* described by MEs. The symbol $*$ is used to indicate that absorbing-states are present in these processes. Representing the set of absorbing-states by \otimes , we have $W_{\omega \rightarrow \omega'} = 0$ for $\omega \in \otimes$. The survival probability $P_s(t)$ that the system does not visit the absorbing-state up to a time t is given by

$$P_s(t) = \sum_{\omega \notin \otimes} P(\omega, t) = 1 - \sum_{\omega \in \otimes} P(\omega, t), \quad (3.22)$$

and the QS distribution for this process is given by

$$\bar{P}(\omega) = \lim_{t \rightarrow \infty} \frac{P(\omega, t)}{P_s(t)}, \quad \omega \notin \otimes. \quad (3.23)$$

It is important to notice that the QS distribution is already normalized. Summing Eq. (3.1) over $\omega \in \otimes$, for which $W_{\omega \rightarrow \omega'} = 0$, and using Eq. (3.22), we have

$$\frac{dP_s}{dt} = - \sum_{\omega \in \otimes} \sum_{\omega' \notin \otimes} P(\omega', t) W_{\omega' \rightarrow \omega} \equiv -W_*. \quad (3.24)$$

Assuming that the QS distribution exists and is unique, and writing $P(\omega, t) = P_s(t) \bar{P}(\omega)$ in Eq. (3.24), we get

$$\frac{1}{P_s} \frac{dP_s}{dt} = - \sum_{\omega \in \otimes} \sum_{\omega' \notin \otimes} \bar{P}(\omega') W_{\omega' \rightarrow \omega} \equiv -\tilde{W}_*, \quad (3.25)$$

whose solution is a surviving probability decaying exponentially as $P_s(t) \sim e^{-t/\tau}$, with a lifespan

$$\tau = \frac{1}{\tilde{W}_*}. \quad (3.26)$$

Now consider a stochastic process X_t with identical transitions of X_t^* , except those involving absorbing-states. So, the ME is modified replacing the summation on $\omega' \in \otimes$ by $Q(\omega, t) \tilde{W}_*(t)$:

$$\frac{dQ(\omega, t)}{dt} = \sum_{\omega' \notin \otimes} [Q(\omega', t) W_{\omega' \rightarrow \omega} - Q(\omega, t) W_{\omega \rightarrow \omega'}] + Q(\omega, t) \tilde{W}_*(t), \quad (3.27)$$

in which

$$\tilde{W}_*(t) = \sum_{\omega \in \otimes} \sum_{\omega' \notin \otimes} Q(\omega', t) W_{\omega' \rightarrow \omega} \quad (3.28)$$

accounts all transitions to absorbing-states, and $Q(\omega, t)$ is the new probability function. The last term in the r.h.s. of Eq. (3.27) means that all transitions to the absorbing-state in the original

dynamics are redirected to an active configuration proportionally to the probability that the configuration occurs. A QS solution of the ME (3.1) is a stationary solution of (3.27) as well. This can be checked by replacing $P(\omega, t) = P_s(t)\bar{P}(\omega)$, $\omega \notin \otimes$, in Eqs. (3.1) and (3.24) to find a self-consistent equation

$$\bar{P}(\omega) = \frac{\sum_{\omega' \notin \otimes} \bar{P}(\omega') W_{\omega' \rightarrow \omega}}{W_\omega^* - \bar{W}_*}, \quad \omega \notin \otimes, \quad (3.29)$$

with $W_\omega^* = \sum_{\omega'} W_{\omega \rightarrow \omega'}$ and \bar{W}_* given by Eq. (3.25). Making the r.h.s. of Eq. (3.27) equal to zero, corresponding to a stationary configuration of the modified dynamics, for $t \rightarrow \infty$, we find

$$\bar{Q}(\omega) = \frac{\sum_{\omega' \notin \otimes} \bar{Q}(\omega') W_{\omega' \rightarrow \omega}}{W_\omega - \bar{W}_*}, \quad (3.30)$$

with $W_\omega = \sum_{\omega' \notin \otimes} W_{\omega \rightarrow \omega'}$, \bar{W}_* given by Eq. (3.28) as $t \rightarrow \infty$, and $\bar{Q}(\omega) \equiv Q(\omega, t \rightarrow \infty)$. Equation (3.30) has the same form of Eq. (3.29), showing that the QS solution of the original ME is a stationary solution of Eq. (3.27) in the modified dynamics. Note that Eq. (3.27) cannot be written in the form of Eq. (3.1), and is non-linear, therefore it is not a ME by definition. Other modified dynamics can be implemented, implying in different types of QS analyses [95].

The average QS density of infected vertices is defined as

$$\langle \rho_s \rangle = \frac{1}{N} \sum_{n=1}^N n \bar{P}_n, \quad (3.31)$$

where \bar{P}_n is the QS probability that the system has n infected vertices, computed after a relaxation time t_{rlx} during an averaging time t_{avg} ; see Sec. 3.3.3. We also define the dynamical susceptibility [26] as

$$\chi = N \frac{\langle \rho_s^2 \rangle - \langle \rho_s \rangle^2}{\langle \rho_s \rangle}, \quad (3.32)$$

where the averages $\langle \rho \rangle$ and $\langle \rho^2 \rangle$ are evaluated in the QS state. In some cases we can omit the subscript “s” and the brackets in Eq. (3.31) when referring to QS quantities. The susceptibility provides a pronounced peak at the epidemic threshold for networks without outliers [17], as shown in Fig. 3.4. This is possible because the dynamical susceptibility around the critical point scales as

$$\chi \sim |\lambda - \lambda_c|^{-\zeta}, \quad (3.33)$$

where ζ is the associated critical exponent, and can be used as a criterion to locate epidemic

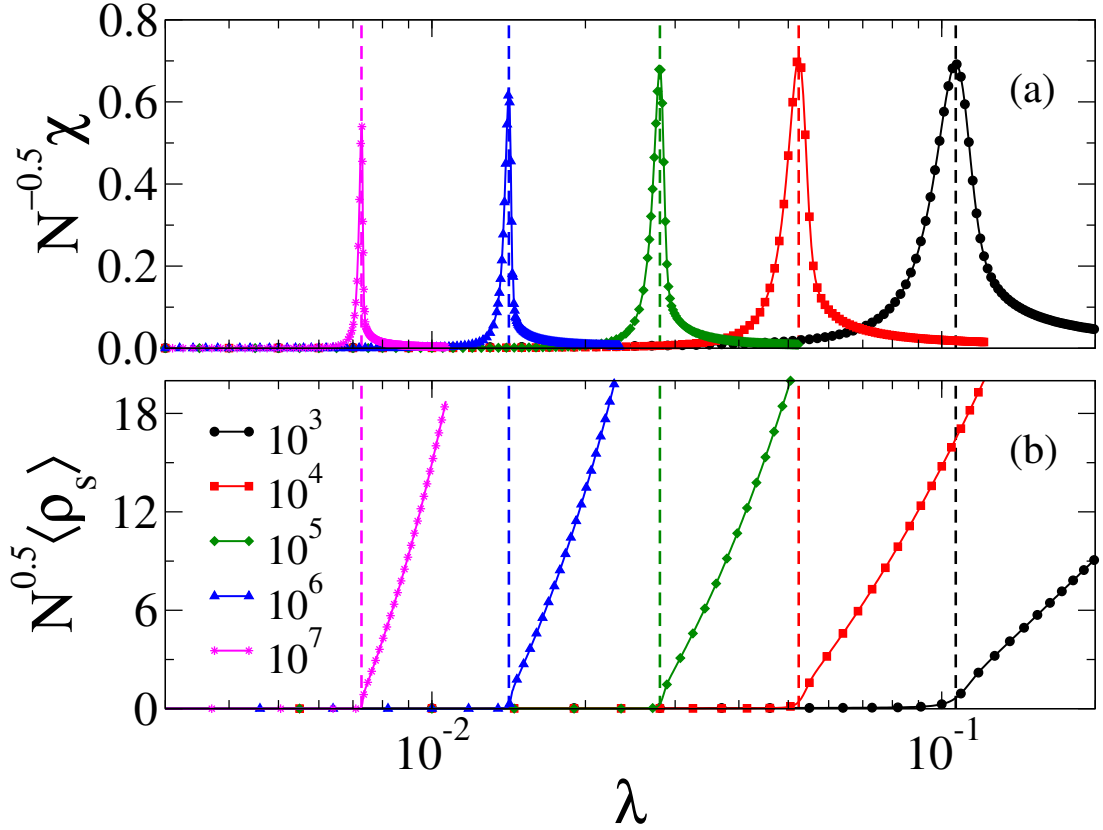


Figure 3.4: Quasi-stationary (a) susceptibility and (b) density divided and multiplied, respectively, by \sqrt{N} (to improve visibility) against infection rate λ for $\mu = 1$ and different network sizes indicated in the legends. The network has a PL degree distribution with $\gamma = 2.3$ and cutoff $k_c \sim N^{1/\gamma}$. Note that the susceptibility peak coincides with the point where density becomes appreciable, indicated by dashed lines.

thresholds [26, 33]. However, in some cases, this is not possible due to, for example, a smearing [96] of the epidemic transition, in which the peaks are not so pronounced; see Chapter 5 for more details. In Fig. 3.4 we used a simple QS method that just prevents the system from falling into the absorbing-state using a reflecting boundary at $n = 0$, in which the dynamics returns to the state that it was immediately before visiting the absorbing-state. The comparison of different QS methods is performed in Ref. [95].

3.3 Simulations and analysis of epidemic processes

Here we present some details about usual methods used to investigate epidemic processes on top of networks employing mean-field approaches and computational simulations. The results from these approximations are compared to the ones in stochastic simulations in Chapter 4.

3.3.1 Mean-field approximations

Many real and synthetic networks on top of which epidemic spreading happens share the small-world property, in which vertices of the network are not far from each other. Therefore, vertices are expected to be subjected to a similar local density of infected neighbors [8, 24]. This property resembles the mean-field assumption where the local field is replaced by the one averaged over the whole system. For low-dimensional regular lattices, this hypothesis is strongly violated but becomes precise in higher dimensions [94]. In the other limit, in an unweighted complete graph the mean-field hypothesis is exact.

We also use the term mean-field when assuming that the states of distinct vertices are statistically independent [6, 8]. Consequently, pairwise a probability factor as $[A_i B_j]$, that is the probability that a vertex i is in state A and j in state B , is assumed to be in the form $[A_i B_j] = [A_i][B_j]$, where $[A_i]$ is the probability that a vertex i is in state A , neglecting dynamical correlations between i and j . These correlations are relevant in quenched networks, since, for example, a susceptible vertex with a larger number of infected neighbors is more probable to be infected at a later time. On the other hand, if the connections are rewired at a time scale much shorter than state transitions, this hypothesis becomes more accurate and is exact for annealed networks in the infinite size limit.

Consider the SIS model on a heterogeneous graph with adjacency matrix A_{ij} . The probability that a vertex i is infected evolves as

$$\frac{d[1_i]}{dt} = -\mu[1_i] + \sum_{j=1}^N \lambda A_{ij} [0_i 1_j]. \quad (3.34)$$

This equation is exact [25, 97] but not solvable for general networks since it depends on $[0_i 1_j]$, and a dynamical equation for this term depends on triplets and so on. Disregarding dynamical correlations, it becomes

$$\frac{d\rho_i}{dt} = -\mu\rho_i + \lambda(1 - \rho_i) \sum_{j=1}^N A_{ij} \rho_j, \quad (3.35)$$

where $\rho_i = [1_i]$ is the probability that i is infected. For a quenched network, it corresponds to the quenched mean-field (QMF) theory [11] where the full structure of the network is included, but dynamical correlations are not. If we now consider the annealed network limit, in which

Eq. (2.20) holds, the previous equation becomes

$$\frac{d\rho_i}{dt} = -\mu\rho_i + \frac{\lambda k_i}{\langle k \rangle}(1 - \rho_i)\Phi, \quad (3.36)$$

where $\Phi = \sum_{j=1}^N k_j \frac{\rho_j}{N}$ is independent of i . In the stationary limit,

$$\rho_i = \frac{\frac{\lambda k_i}{\langle k \rangle} \Phi}{\mu + \frac{\lambda k_i}{\langle k \rangle} \Phi}, \quad (3.37)$$

meaning that the probability that a vertex i is infected only depends on its degree k_i in annealed networks. Indeed, there is a heterogeneous mean-field (HMF) theory [87] where vertices are compartmentalized according to their degree. Both HMF and QMF theories are discussed in Chapter 4, with results for both quenched and annealed networks.

Pair quenched [97] and heterogeneous [33] mean-field approximations can also be defined to make the previous approximations more accurate by breaking triplets as

$$[A_i B_j C_l] \approx \frac{[A_i B_j][B_j C_l]}{[B_j]}. \quad (3.38)$$

See Refs. [33, 97] for more details.

3.3.2 Building algorithms for generic epidemic models

Due to constant inconsistencies among simulations of the same epidemic model on networks that are frequently reported, we describe a generic method to perform statistically exact simulations that can be extended to arbitrary epidemic processes detailed in Ref. [34]. To build algorithms for generic epidemic models, let us consider an unweighted network with an adjacency matrix A_{ij} where an infected vertex i becomes spontaneously recovered at rate μ_i , transmits the infection to vertex j with rate λ_{ij} and, if recovered, turns again to a susceptible state (waning immunity [88]) with rate α_i . Here we will consider the general case in which the three compartments S, I, and R are available.

The algorithms require dynamical lists which are represented by capital calligraphic letters. In both the standard and optimized GAs for the generic epidemic process, we build and constantly update two lists $\mathcal{V}^{(I)}$ and \mathcal{M} with the positions (labels of the vertices) $\mathcal{V}_p^{(I)}$ and recovering rates \mathcal{M}_p of the $p = 1, \dots, N_I$ infected vertices. The list updates are simple: the entries of a

new infected vertex are added in their ends. When an infected vertex is chosen using $\mathcal{V}^{(I)}$ and becomes recovered (or susceptible if $\alpha_i = \infty$), the last entries of the lists are moved to the index p of the selected vertex, and the list sizes are shortened by 1. Similarly, we build and keep updated the lists $\mathcal{V}^{(R)}$ and \mathcal{A} , with the positions $\mathcal{V}_p^{(R)}$ and rates \mathcal{A}_p of the $p = 1, \dots, N_R$ recovered vertices. We also keep updated the total rate that infected vertices are recovered and that recovered ones become susceptible, which are given by

$$M = \sum_{i=1}^N \mu_i \delta_{\omega_i 1} = \sum_{p=1}^{N_I} \mathcal{M}_p \quad (3.39)$$

and

$$A = \sum_{i=1}^N \alpha_i \delta_{\omega_i 2} = \sum_{p=1}^{N_R} \mathcal{A}_p, \quad (3.40)$$

respectively, where δ_{ij} is the Kronecker delta. The update of M or A is done by adding (subtracting) \mathcal{M}_p or \mathcal{A}_p when a new element is added (removed) in $\mathcal{V}^{(I)}$ or $\mathcal{V}^{(R)}$, respectively.

Implementation of the Gillespie Algorithm

To implement the GA, we need the lists $\mathcal{V}^{(IS)}$ and \mathcal{L} , with the positions $\mathcal{V}_p^{(IS)}$ of the susceptible vertices and infection rates \mathcal{L}_p involving the $p = 1, \dots, N_{IS}$ edges connecting infected and susceptible vertices. It worths to remark that a same susceptible vertex i will appear n_i times in the list $\mathcal{V}^{(IS)}$, where n_i is its number of infected neighbors. Due to the multiplicity of edges to be added and deleted from the list every time a change of state occurs, both $\mathcal{V}^{(IS)}$ and \mathcal{L} are rebuilt and the total infection

$$L = \sum_{i,j=1}^N \lambda_{ij} \delta_{\omega_i 1} \delta_{\omega_j 0} = \sum_{p=1}^{N_{IS}} \mathcal{L}_p \quad (3.41)$$

is computed after each event visiting only the infected vertices and their neighbors with the aid of $\mathcal{V}^{(I)}$.

With these three lists, the steps of GA (*cf.* section 3.1.2) can be implemented as follows. The total rate of spontaneous processes is $R = M + A + L$. With probabilities $m = M/R$, $a = A/R$, and $l = L/R$ we choose the class of event $I \rightarrow R$, $R \rightarrow S$, or $IS \rightarrow II$, respectively. If the event is a recovering $I \rightarrow R$, one element p of $\mathcal{V}^{(I)}$ is chosen with probability proportional to \mathcal{M}_p and the respective infected vertex is recovered. If a waning of immunity $R \rightarrow S$ was selected, one element p of $\mathcal{V}^{(R)}$ is chosen with probability proportional to \mathcal{A}_p and the recovered vertex

becomes susceptible. Finally, if an infection event $IS \rightarrow II$ was selected, one element p of $\mathcal{V}^{(IS)}$ is selected with probability proportional to \mathcal{L}_p and the susceptible vertex is infected. The time is incremented by τ drawn from the distribution $P_R(\tau)$ given by Eq. (3.7).

The choice of the events proportionally to \mathcal{M}_p , \mathcal{A}_p , and \mathcal{L}_p can be implemented using the *rejection method*, in which an event p is selected with equal chance and accepted with probability ν_p/ν_{\max} where ν_{\max} is the largest rate in the corresponding kind of event. The rejection is iteratively repeated until a choice is accepted. It is simpler and usually more efficient to adopt $\mu_{\max} = \max_i \{\mu_i\}$, $\alpha_{\max} = \max_i \{\alpha_i\}$, and $\lambda_{\max} = \max_{ij} \{\lambda_{ij}\}$ along the whole network instead of to update this value after every step.

The implementation of the infection events in this GA follows an edge-based while the healing and waning of immunity use a vertex-based update scheme. Alternatively, all events could be performed with vertex-based schemes [98] using a rate $\nu_i = \mu_i \delta_{\omega_i 1} + \alpha_i \delta_{\omega_i 2} + \delta_{\omega_i 0} \sum_j \lambda_{ji}$. This approach, which is not considered in this work, is computationally slightly simpler to use since no list is necessary. However, its simplest version demands to visit the whole network to compute the rates after every time step, which is computationally prohibitive for very large networks.

Optimized Gillespie algorithm (OGA)

Most of the computational load in the original GA holds in building the list $\mathcal{V}^{(IS)}$. Here, we describe a strategy that optimizes this step by introducing *phantom processes*, that do not change the state of the system but do contribute to time counting. The phantom processes here consist of infected vertices i trying to infect other infected or recovered vertex j with the same rate λ_{ij} that they would infect j if they were susceptible, resulting therefore in no change of state; see Fig. 3.5. We refer to this algorithm as the optimized Gillespie algorithm (OGA) [34]. The method is exact by construction because it includes processes that are implemented according to the GA rules but do not change states neither interfere in the processes that actually do.

We write the total infection rate given by Eq. (3.41) as $L = W - P$ where

$$W = \sum_{i,j=1}^N \lambda_{ij} \delta_{\omega_i 1} \quad (3.42)$$

is the total infection rate emanating from infected vertices, including the phantom processes,

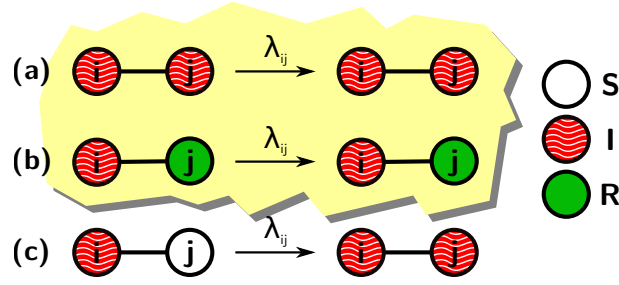


Figure 3.5: Examples of (a)-(b) phantom and (c) real infection processes for a pair of vertices (i, j) .

and

$$P = \sum_{i,j=1}^N \lambda_{ij} \delta_{\omega_i 1} [\delta_{\omega_j 1} + \delta_{\omega_j 2}] \quad (3.43)$$

is the total infection attempts due to the phantom processes. The total rate of processes is $R = M + A + W$, which is larger than in GA since $W \geq L$.

We need now to calculate once at the beginning of and to keep stored along the simulation the maximal infection that can be produced by each vertex, given by

$$w_i = \sum_{j=1}^N \lambda_{ij}. \quad (3.44)$$

The general epidemic dynamics can be simulated as follows. We build and keep updated the lists $\mathcal{V}^{(I)}$, \mathcal{M} , $\mathcal{V}^{(R)}$, and \mathcal{A} defined previously. We also need a list \mathcal{W} with the infection \mathcal{W}_p produced by the $p = 1, \dots, N_I$ infected vertices of the list $\mathcal{V}^{(I)}$, and the total infection rate

$$W = \sum_{p=1}^{N_I} \mathcal{W}_p, \quad (3.45)$$

produced by these vertices. The update of \mathcal{W} and W follows the same steps of \mathcal{M} , M , \mathcal{A} , and A described previously instead of the heavier task of building the list $\mathcal{V}^{(IS)}$ and calculating L in the original GA. This point is essential for the algorithm efficiency as discussed in Ref. [34].

With probabilities $m = M/R$ and $a = A/R$ we perform, respectively, a recovering or waning of immunity as described for the original GA. With probability $w = W/R$ an infected vertex is randomly selected as an element p of $\mathcal{V}^{(I)}$ with probability proportional to \mathcal{W}_p and let $i = \mathcal{V}_p^{(I)}$. Next, one neighbor j of i is chosen with probability proportional to λ_{ij} using the rejection probability $\lambda_{ij}/\lambda_i^{(\max)}$, where $\lambda_i^{(\max)} = \max_j \{\lambda_{ij}\}$. If the selected neighbor is susceptible, it is infected. If the selected neighbor is infected or recovered, *i.e.* a phantom process, no change of

state is implemented, the time is incremented as in the original GA using $R = M + A + W$, and the simulation runs to the next time step. Note that the frustrated infection attempts reckon exactly the total rate of phantom processes P given by Eq. (3.43). The values of $\lambda_i^{(\max)}$ need to be computed once at the beginning of the simulation. Depending on the model, further simplifications and improvements can be adopted.

The computational time required for the OGA is an order of magnitude smaller than those to the original GA. With that, we can easily simulate networks with, for example, $N = 10^8$ vertices. More details are available in Ref. [34], while the particular application of the OGA for the SIS model is presented in Appendix A.1.

3.3.3 Decay, spreading and quasi-stationary analysis

Through this thesis, we will refer to some kinds of analyses performed. In some, we start by considering all individuals infected and observe the temporal evolution of the fraction of infected individuals decaying in time. In others, we can see how a single individual, or a small fraction of the network, can spread the disease to the whole network, defining a spreading dynamics. Finally, we can also perform a QS analysis, without analyzing the temporal evolution explicitly, looking only at the quasi-stationary behavior.

Decay analysis

In a decay analysis the initial condition is the whole network infected: $\omega_i = 1 \forall i \in [1, N]$. The dynamics evolves by using the SIS model and averages over independent stochastic simulations samples are computed. The usual quantities are the following:

- **Fraction of infected individuals over time:** the order parameter ρ is observed as a function of time, measuring the number of infected individuals for each time step of the dynamics;
- **Survival probability over time:** it is defined as the number of dynamical samples for which $N_{\text{inf}} > 0$ at a time t , denoted by $P_s(t)$. In an active state, it approaches a finite value for $t \rightarrow \infty$ and decays exponentially in the absorbing phase.

Spreading analysis

In this analysis, we start with only one or a few vertices infected, while others are susceptible. Different independent stochastic dynamical samples are run for each vertex of the network. The measures are similar to the previous ones, and usually the number of infected individuals $N(t)$ is computed instead of the fraction $\rho(t)$. In the absorbing phase, we have exponential decays,

while in the active we can see exponential growths. At the epidemic threshold, we observe a PL behavior with time, that also happens for curves of survival probability $P_s(t)$. Another interesting measure is of the avalanche size distribution $P_{\text{ava}}(s)$. The size s is defined as the number of vertices that were infected at least once during a spreading dynamics. For more details about how the structure of networks can affect these regimes, see Chapter 5.

QS analysis

In the two previous analyses, we are interested in the time evolution of ρ to check if the curve tends to grow or to decay for a given value of the infection rate λ . On the other hand, in the QS analysis we force the dynamics to avoid the absorbing-state. There are different methods [95, 99] to do this. One simple way is to use the reflective bound condition (RBC) method described in Ref. [95] by adding the following rule to the SIS algorithm: if there is only one infected vertex and a healing process is going to happen, time evolves but the healing is not implemented. The measured quantities are the dynamical susceptibility χ and the average fraction of infected individuals in the QS state, $\langle \rho_s \rangle$. In both cases, averages of ρ and ρ^2 are performed after a relaxation time t_{rlx} during an average time t_{avg} , as shown in Fig. 3.6. These methods converge to the real stationary state in the thermodynamic limit $N \rightarrow \infty$ for both active and inactive states, since in the former there is no visit to the absorbing-state, and in the latter ρ converges to zero.

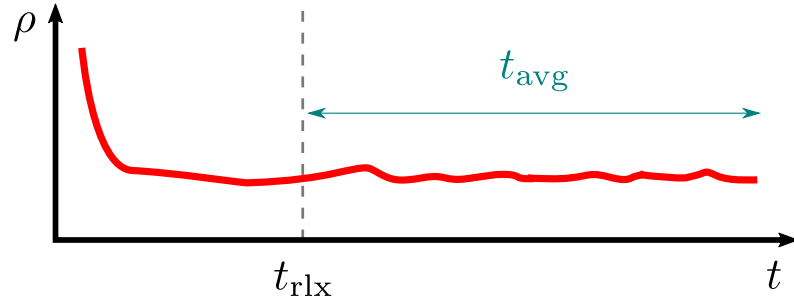


Figure 3.6: The average of ρ in a QS analysis is computed after a relaxation time t_{rlx} during an average time t_{avg} to evaluate the quasi-stationary values of the order parameter.

3.4 Summary

There are many aspects of APTs not covered in this chapter that can be found in textbooks [25, 93, 94]. For a more complete description of algorithms for computation simulations of epidemic processes, see Ref. [34]. We will use concepts defined in this chapter in the rest of this thesis, while other types of analyses specific to some problems are defined in their respective chapters.

Part II

Criticality in the epidemic spreading on complex networks

Chapter 4

Robustness and fragility of the susceptible-infected-susceptible model

Related publication:

Robustness and fragility of the susceptible-infected-susceptible epidemic models on complex networks [35]

Wesley Cota, Angélica S. Mata, Silvio C. Ferreira

Physical Review E **98**, 012310 (2018)

We propose two types of modifications in the standard SIS model described in Chapter 3, preserving the central properties of spontaneous healing and infection capacity growing unlimitedly with the degree. All models have the same mean-field epidemic thresholds, but stochastic simulations describe different scenarios that depend on the structural properties of the networks: the SIS threshold can change drastically, or remain robust with the proposed changes. This study is important since the SIS behavior is not completely understood in complex networks [8], even though it is a basic model for epidemics. If small changes in the model lead to conflicting results, which would be the actual mechanisms used in models that correspond to real epidemics and which would be the best approaches to analytically investigate real epidemic processes? Our results suggest that the hub activation mechanism, intensively investigated recently [9, 11, 15–17, 19, 21, 27, 100], seems to be more a peculiarity than a rule in epidemic spreading.

4.1 Analytical approximations for the epidemic threshold

The SIS dynamics presents a transition between an inactive absorbing-state to an active stationary phase, in an endemic state. Since real and synthetic networks are finite, a systematic

study of the threshold dependence on the network size is a fundamental problem. Analytically, this can be done through mean-field methods [12], such as the heterogeneous mean-field (HMF) [2, 6] and quenched mean-field (QMF) [11] theories. For power-law uncorrelated networks with exponent γ , these theories predict conflicting results for $\gamma > 3$ [11], and different asymptotic behaviors for $5/2 < \gamma < 3$.

For $2 < \gamma < 5/2$, both HMF and QMF theories agree that the epidemic threshold is null in the thermodynamic limit ($N \rightarrow \infty$), while for $\gamma > 3$ the HMF theory predicts a finite while QMF a null threshold. The existence of a null threshold predicted by QMF theory is in accordance with exact results [9, 10] and stochastic simulations [15, 26, 97]. In the interval $5/2 < \gamma < 3$, both theories predict a null threshold, but with different finite-size scaling. It is important to stress that even the addition of dynamical correlations using pair approximations [12] does not change this scenario [20, 33, 97].

A criterion based on the mutual reinfection of hubs was developed recently for the SIS model [15] to understand the nature of the epidemic thresholds in dynamical processes in complex networks with power-law degree distributions [16]. In Ref. [16], according to this criterion, when waning immunity [88] is added to the SIS model, a finite threshold is found for $\gamma > 3$, agreeing with numerical simulations and in contrast with the QMF prediction. In this section, we investigate the robustness of this mechanism by modifying the SIS model.

4.1.1 Modifications of the SIS model

In all modified models and the standard one, infected individuals are spontaneously healed with rate μ and have the infection capacity increasing with their degrees. In the standard SIS, denoted by SIS- \mathcal{S} , an infected vertex infects each susceptible neighbor with rate λ . In the modified SIS- \mathcal{T} model, infection happens with rate λ when a certain number of neighbors of a susceptible vertex is infected. Here we consider a threshold of at least one infected neighbor. The second modified model is the SIS- \mathcal{A} , a modification of the contact process in which all susceptible neighbors of an infected vertex are infected simultaneously with rate λ . Symbols \mathcal{S} , \mathcal{T} and \mathcal{A} refer to *standard*, *threshold*, and *all*, respectively. A summary of the rules for each model is presented in Table 4.1. It is important to mention that the SIS- \mathcal{T} model is an asynchronous version of models investigated in seminal papers [72, 87] about epidemic processes in complex networks.

All models can be mapped in dynamical processes on top of ordered regular networks with

Table 4.1: Definition of the epidemic models and some computational implementation details of the Gillespie algorithm presented in the main text. Symbols: N_{inf} is the number of infected vertices; N_{SI} is the number of susceptible vertices with at least one infected neighbor; N_e is the number of stubs pointing from infected vertices; and u is a random number uniformly distributed in the interval $(0,1]$.

SIS- \mathcal{T} (<i>threshold</i>) [101, 102]	SIS- \mathcal{A} (<i>all</i>) [103, 104]	SIS- \mathcal{S} (<i>standard</i>) [8]
Infected vertices are spontaneously healed with rate μ	Infected vertices are spontaneously healed with rate μ	Infected vertices are spontaneously healed with rate μ
Susceptible vertices become infected with rate λ if they have at least one infected neighbor	Infected vertices infect at once all susceptible neighbors with rate λ	Infected vertices independently infect each susceptible neighbor with rate λ
GA infection probability $q = \frac{\lambda N_{\text{SI}}}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}$	GA infection probability $q = \frac{\lambda}{\mu + \lambda}$	GA infection probability $q = \frac{\lambda N_e}{\mu N_{\text{inf}} + \lambda N_e}$
GA time step $\tau = \frac{-\ln(u)}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}$	GA time step $\tau = \frac{-\ln(u)}{(\mu + \lambda) N_{\text{inf}}}$	GA time step $\tau = \frac{-\ln(u)}{\mu N_{\text{inf}} + \lambda N_e}$

coordination number k : SIS- \mathcal{S} is mapped in the contact process [25, 105], in which the infected vertices transmit with rate $\lambda_{\text{CP}} = k\lambda_{\text{SIS}}$. SIS- \mathcal{A} was studied in Refs. [103, 104], while \mathcal{T} was investigated in Refs. [101, 102]. All models belong to the DP universality class [25] in these networks. However, in complex networks a finite-size scaling analysis presented in Sec. 4.2.3 show non-universal exponents for these models.

To understand the microscopic difference between these three models, consider an infinitesimal time step Δt and an infected vertex (center) with k susceptible neighbors (leaves); see Fig. 4.1(a). The probability that s leaves are infected by the center in both SIS- \mathcal{S} and SIS- \mathcal{T} is

$$P_{\text{leaf}}^{(\mathcal{S}, \mathcal{T})}(s) = \binom{k}{s} (\lambda \Delta t)^s (1 - \lambda \Delta t)^{k-s}, \quad (4.1)$$

while in SIS- \mathcal{A}

$$P_{\text{leaf}}^{(\mathcal{A})}(s) = \lambda \Delta t \delta_{sk}, \quad (4.2)$$

where δ_{sk} is the Kronecker delta. Note that both expressions produce the same average number of leaves infected $\langle s \rangle = \lambda k \Delta t$. The probability that a susceptible center surrounded by $s > 0$

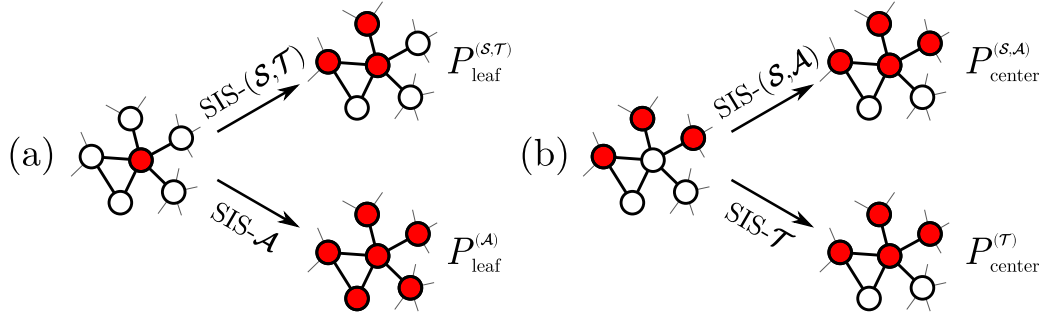


Figure 4.1: Some infection processes in the SIS models. (a) An infected vertex (center) with many susceptible neighbors (leaves). (b) A susceptible center with infected leaves. Transition probabilities are defined in Eqs. (4.1)–(4.4).

infected leaves, Fig. 4.1(b), becomes infected is given by

$$P_{\text{center}}^{(S,A)}(s) = 1 - (1 - \lambda \Delta t)^s \approx \lambda s \Delta t \quad (4.3)$$

for $SIS-S$ and $SIS-A$, while for $SIS-T$ it is

$$P_{\text{center}}^{(T)} = \lambda \Delta t. \quad (4.4)$$

Consequently, while the infection of leaves by the center in the $SIS-S$ is equivalent to the $SIS-T$ model, the infection of the center by leaves in the $SIS-S$ model is equivalent to $SIS-A$.

4.1.2 Mean field theories and epidemic thresholds

In the HMF theory, dynamical equations describe the probability $\rho_k(t)$ that a vertex of degree k is infected at a time t . The interaction between a vertex of degree k with a neighbor of degree k' is calculated by considering the probability $P(k'|k)$ that a vertex of degree k is connected to another of degree k' . Thus, the probability that a neighbor is infected is [87]

$$\Theta_k = \sum_{k'} P(k'|k) \rho_{k'},$$

taking into account the probability of finding an infected individual over all degrees k' . So, $\rho_k(t)$ evolves as

$$\frac{d\rho_k}{dt} = -\mu \rho_k + \lambda(1 - \rho_k) \Psi_k(\Theta_k), \quad (4.5)$$

where $\Psi_k(\Theta_k) = k\Theta_k$ for $SIS-S$ and $SIS-A$, and $\Psi_k(\Theta_k) = 1 - (1 - \Theta_k)^k$ for $SIS-T$. In the first two cases, the infection can arrive from k neighbors, multiplied by the probability that they are

infected. In the latter case, this probability depends on the existence of at least one infected neighbor, regardless of the others. In all models, the term $-\mu\rho_k$ represents the spontaneous healing at rate μ .

The QMF theory computes the individual probability $\rho_i(t)$ that a vertex $i = 1, 2, \dots, N$ is infected at a time t considering explicitly the network structure by means of the adjacency matrix A_{ij} . In this case, $\rho_i(t)$ evolves as

$$\frac{d\rho_i}{dt} = -\mu\rho_i + \lambda(1 - \rho_i)\Psi_i, \quad (4.6)$$

where $\Psi_i = \sum_j A_{ij}\rho_j$ for SIS- \mathcal{S} and SIS- \mathcal{A} models, and

$$\Psi_i = 1 - \prod_j (1 - \rho_j A_{ij})$$

for SIS- \mathcal{T} , with the same arguments for the HMF case.

Through stability analysis and linearization of Eqs. (4.5) and (4.6) around the fixed points $\rho_k = 0$ and $\rho_i = 0$, respectively, we have the same linearized equations for the three models in both theories:

$$\frac{d\rho_k}{dt} = -\mu\rho_k + \lambda \sum_{k'} C_{kk'} \rho_{k'} \quad (4.7)$$

and

$$\frac{d\rho_i}{dt} = -\mu\rho_i + \lambda \sum_j A_{ij} \rho_j, \quad (4.8)$$

where $C_{kk'} = kP(k'|k)$ is the connectivity matrix. The HMF and QMF epidemic thresholds are obtained when the leading eigenvalues of their respective Jacobians $J_{kk'}^{\text{HMF}} = -\mu\delta_{kk'} + \lambda C_{kk'}$ and $J_{ij}^{\text{QMF}} = -\mu\delta_{ij} + \lambda A_{ij}$ are zero. Note that both $C_{kk'}$ and A_{ij} are positive semidefinite and irreducible matrices, and, by the Perron-Frobenius (PF) theorem [1, p. 346], a real square matrix with non-negative elements has a positive eigenvalue Λ_{\max} such that $|\Lambda| \leq \Lambda_{\max}$ for any other eigenvalue Λ of this matrix. Besides, it guarantees that the associated eigenvector is non-degenerated and has all elements non-negative. In the limit $t \rightarrow \infty$, the dominant term of $\rho_i(t)$ will be given by the leading eigenvalue Λ_{\max} of $C_{kk'}$ and $A_{kk'}$, for Eqs. (4.7) and (4.8), respectively [106]. When the eigenvalues of the Jacobians are zero, the absorbing-state becomes unstable and defines a transition to an active state.

The HMF epidemic threshold is given by [107]

$$\lambda_c^{\text{HMF}} = \frac{\mu}{\Lambda_{\max}}, \quad (4.9)$$

where Λ_{\max} is the leading eigenvalue of $C_{kk'}$ given by

$$\sum_{k'} C_{kk'} v_{k'} = \Lambda_{\max} v_k.$$

For uncorrelated networks, $C_{kk'} = kP(k'|k) = kk'P(k')/\langle k \rangle$ [108]. An eigenvector is $v_k = k$ [6], which is non-negative, resulting in

$$\sum_{k'} k \frac{k'P(k')}{\langle k \rangle} k' = k \frac{\langle k^2 \rangle}{\langle k \rangle} = \Lambda_{\max} k$$

implying that $\Lambda_{\max} = \langle k^2 \rangle / \langle k \rangle$. Therefore,

$$\lambda_c^{\text{HMF}} = \mu \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (4.10)$$

Similarly, for the QMF theory, we have [11]

$$\lambda_c^{\text{QMF}} = \frac{\mu}{\Lambda_{\max}}, \quad (4.11)$$

where Λ_{\max} is the leading eigenvalue of the adjacency matrix A_{ij} , regardless of whether or not there are correlations in the network. In many cases we will use $\mu = 1$, without loss of generality.

4.1.3 Testing HMF theory using annealed networks

In annealed networks, the edges are randomly rewired at each time step following the probability $P(k'|k)$. In this regime the HMF is exact, predicting the epidemic threshold and density of infected vertices in the thermodynamic limit [24, 71]. Simulations with absorbing-states near the transition require special techniques [34], like the standard quasi-stationary (QS) method described in Ref. [95], in which the averaging is constrained to the active states and converges to the actual stationary phase in the thermodynamic limit. The threshold in finite networks can be estimated using the principal peak of the dynamical susceptibility χ defined in Eq. (3.32).

Figures 4.2(a) and (c) confirm the agreement between simulations on annealed networks

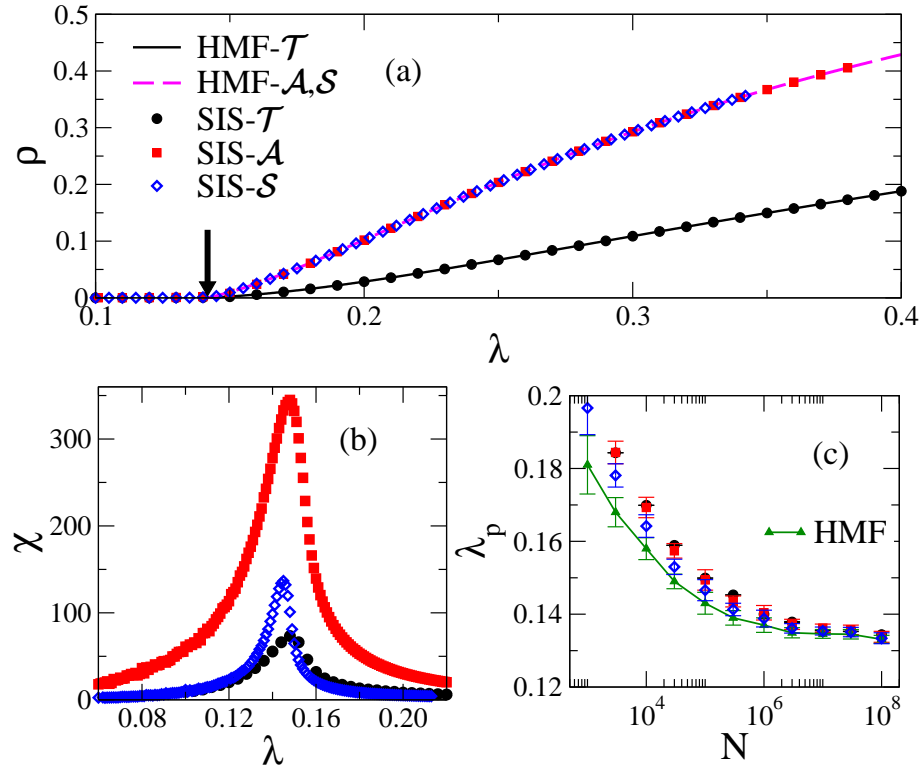


Figure 4.2: Comparison of HMF theory and simulations on annealed networks with $N = 10^5$ vertices, degree distribution $P(k) \sim k^{-3.5}$, minimal degree $k_{\min} = 3$, and upper cutoff $k_c = \sqrt{N}$. (a) QS density and (b) susceptibility versus infection rate curves are shown. Lines in (a) are numerical solutions of Eq. (4.5) in the stationary regime and the arrow indicates the HMF epidemic threshold $\lambda_c^{\text{HMF}} = \langle k \rangle / \langle k^2 \rangle$. (c) Finite-size dependence of the threshold estimated via susceptibility and HMF theory. The curves correspond to averages over 10 independent network realizations.

and HMF theory for the stationary fractions of infected vertices, and epidemic threshold, respectively, in all investigated models. However, fluctuations in the order parameter are different, as shown by the susceptibility curves in Fig. 4.2(b). SIS- \mathcal{S} and SIS- \mathcal{T} models have similar values before and after the epidemic threshold, but not around it, with the latter having stronger fluctuations. The SIS- \mathcal{A} has larger fluctuations in the whole investigated region of λ since all neighbors of an infected individual are infected at the same time. This effect will be further discussed in Sec. 4.2.3.

4.1.4 Simulations on quenched random networks

The models were investigated on quenched random networks with power-law $P(k) \sim k^{-\gamma}$ degree distributions by using the UCM algorithm [82] with minimum degree $k_0 = 3$ and structural cutoff $k_c = \sqrt{N}$, ensuring the lack of correlations [70]. These networks also allow a comparison with the HMF threshold predicted by Eq. (4.10). The simulations were performed

by using particular cases of the optimized Gillespie algorithm [34], whose implementations are defined in Appendix A.

In Fig. 4.3 we compare the epidemic thresholds found using the susceptibility peaks and those predicted by the HMF and QMF theories. For $\gamma < 5/2$, represented by $\gamma = 2.3$ in Fig. 4.3(a), all models have approximately the same thresholds well described by both HMF and QMF theories, which have already been reported for the standard SIS model [26]. Different results appear when $\gamma > 5/2$. For $5/2 < \gamma < 3$, represented by $\gamma = 2.7$ in Fig. 4.3(b), SIS- \mathcal{T} and SIS- \mathcal{A} models have the same thresholds, whose scaling is very well fitted by the HMF theory and deviates from QMF. The threshold of the standard SIS- \mathcal{S} vanishes with a scaling deviating from both HMF and QMF theories.

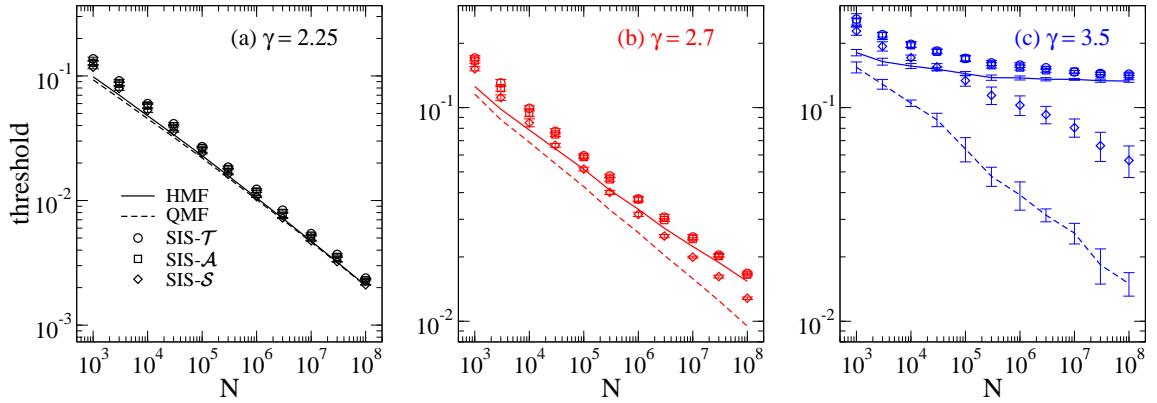


Figure 4.3: Epidemic thresholds for SIS models on UCM networks with $k_0 = 3$, $k_c = \sqrt{N}$, and different degree exponents (a) $\gamma = 2.25$, (b) 2.7, and (c) 3.5. Solid and dashed lines correspond to HMF and QMF theories, respectively. Curves are averages over 10 network realizations. Negligible error bars in mean-field theories are not shown.

The difference between the models is even stronger for $\gamma > 3$, represented by $\gamma = 3.5$ in Fig. 4.3(c). Again, SIS- \mathcal{T} and SIS- \mathcal{A} dynamics present a finite threshold in very satisfactory accordance with HMF theory and contrasting with the standard SIS- \mathcal{S} model that presents the threshold approaching zero as the size increases. Note, however, that the thresholds of SIS- \mathcal{S} have a scaling incompatible with QMF for the investigated size range that cannot be reckoned by neither pairwise QMF [97] or HMF [20, 33] theories. The latter still predicts a finite threshold, inconsistent with simulations and the rigorous results [9] for SIS- \mathcal{S} .

Pairwise mean-field theories for SIS- \mathcal{S}

HMF and QMF theories can be improved by a pairwise approximation, considering the dynamical correlations between states of neighbors. In the case of the SIS- \mathcal{S} with $\mu = 1$, the

pairwise HMF (PHMF) threshold [33] is

$$\lambda_c^{\text{PHMF}} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \quad (4.12)$$

for uncorrelated networks, while the pairwise QMF (PQMF) threshold is obtained when the leading eigenvalue of the matrix [97],

$$L_{ij} = -\left(1 + \frac{\lambda^2 k_i}{2\lambda + 2}\right) \delta_{ij} + \frac{\lambda(2 + \lambda)}{2\lambda + 2} A_{ij}, \quad (4.13)$$

is zero. Figure 4.4 compares the PHMF and PQMF thresholds for UCM networks with the results of stochastic simulations. For $\gamma = 2.7$, the stochastic simulations are well explained by PQMF, but not by PHMF. For $\gamma = 3.5$, both fail to describe the threshold scaling.

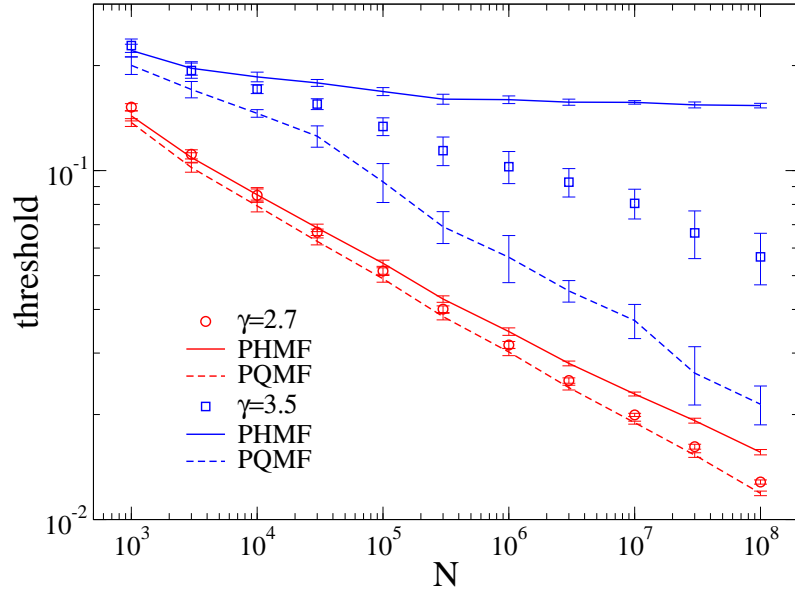


Figure 4.4: Comparison of pairwise approximations with simulations (symbols) for epidemic thresholds of the SIS-S on UCM networks, given by Eqs. (4.12) [PHMF, solid lines] and (4.13) [PQMF, dashed lines]. Data correspond to averages over 10 network samples.

4.2 Activation mechanisms in networks

A criterion based on the mutual reinfection between hubs involves the recovering time (lifespan) τ_k , which is the time that an epidemic process lasts in a star graph with k leaves of degree 1, mimicking the hubs of the networks, and the time $\tau^{(\text{inf})}$ needed for the infection transmission between hubs. If $\tau_k \gg \tau^{(\text{inf})}$, the hubs remain active for enough time to infect each other, and the epidemic is said to be triggered by the mutual activation of hubs. As a

result, the epidemic threshold vanishes in the thermodynamic limit, since hubs can keep the active epidemic state for any value of $\lambda > 0$. Nonetheless, if $\tau_k \lesssim \tau^{(\text{inf})}$, hubs are not capable of sustaining the epidemics alone, and a transition to an active state can be triggered only by a collective activation. As a consequence, the epidemic threshold is finite in this case. To clarify the antagonistic results for $\gamma > 3$ and to understand the relationship between the structure of networks and activation mechanisms of the different SIS dynamics, we consider the recovering time of the epidemics on star graphs for small values of λ .

4.2.1 Approximated expressions for epidemic lifespan on star graphs

To obtain approximated expressions for the lifespan of the SIS models in a star graph with k leaves, let us consider a discrete time dynamics based on Ref. [16]:

- (i) At time $t = 0$, the center is infected and all leaves susceptible.
- (ii) At time $t = t_1$, the center is healed and n leaves are simultaneously infected with probability $P_1(n|k)$.
- (iii) At time $t = t_1 + t_2$, the center is reinfected and all leaves become susceptible simultaneously with probability $P_2(n)$.

The probability that this dynamics survives after these three steps is

$$Q = \sum_{n=1}^k P_2(n)P_1(n|k), \quad (4.14)$$

and the probability that the dynamics reach the absorbing-state after the s -th step is $Q^{s-1}(1-Q)$.

So, the average number of steps is

$$\langle s \rangle = \sum_{s=0}^{\infty} s Q^{s-1} (1-Q) = \frac{1}{1-Q}. \quad (4.15)$$

Next, we define the times t_i and probabilities P_i ($i = 1, 2$) for each model. The steps for SIS- \mathcal{S} are reproduced here from Refs. [15, 16] and used as guides for SIS- \mathcal{T} and SIS- \mathcal{A} models. Let $t_1 = t_2 = 1/\mu$ be the average time that a vertex takes to be healed. The probability that the center infects a leaf while it is still infected is $p = \lambda/(\mu + \lambda)$ [15], that happens with the same

probability for each leaf. So, the probability that n leaves are infected at time $t = t_1$ is

$$P_1(n|k) = \binom{k}{n} p^n (1-p)^{k-n}. \quad (4.16)$$

The time t_2 is related to the probability that at least one leaf reinfects the center before healing and happens with probability

$$P_2(n) = 1 - (1-p)^n. \quad (4.17)$$

Inserting Eqs. (4.16) and (4.17) in (4.14), we have

$$Q = 1 - (1-p^2)^k \approx 1 - \exp\left(-k \frac{\lambda^2}{\mu^2}\right), \quad (4.18)$$

an approximation valid for $\lambda \ll \mu$. Now, inserting (4.18) in (4.15), we have

$$\tau_k^S = (t_1 + t_2) \langle s \rangle \approx \frac{2}{\mu} \exp\left(\frac{k\lambda^2}{\mu^2}\right) \quad (4.19)$$

for the standard SIS dynamics. As a result, exponential growth with the size of the star is expected.

For the case of SIS- \mathcal{A} dynamics, since all leaves are infected at the same time by an infected center with probability $p = \lambda/(\mu + \lambda)$ we have $P_1(n|k) = p\delta_{nk}$, while $P_2(n)$ does not change. Then, we have $Q = p[1 - (1-p)^k]$, which leads to $\langle s \rangle \approx 1$ and the lifespan is

$$\tau_k^{\mathcal{A}} = (t_1 + t_2) \langle s \rangle \approx \frac{2}{\mu} \quad (4.20)$$

for $\lambda \ll \mu$, a constant value.

Finally, since in SIS- \mathcal{T} dynamics the infection depends on the existence of at least one infected neighbor, the expression for $P_1(n|k)$ is the same of SIS- \mathcal{S} , while the probability that the center is reinfects is $P_2(n) = p$, irrespective of $n > 0$. So, $\langle s \rangle \approx 1$ as in SIS- \mathcal{A} . However, since the infection rate of the center is independent of how many leaves are infected, we should use the average time for all leaves to be healed instead of the average time for a single leaf to heal. Considering the healing processes of each leaf as being an independent Poisson process and neglecting the possibility of reinfections of leaves during this process, the average time for n

leaves to be healed is

$$t_2^{(n)} = \int_0^\infty t \left[n (1 - e^{-\mu t})^{n-1} e^{-\mu t} \right] \mu dt \approx \frac{\sqrt{2\pi}}{\mu e} \ln n \approx \frac{0.92}{\mu} \ln n. \quad (4.21)$$

The term between brackets is the probability that one single leaf is infected at time t , μdt is the probability that it heals at time t , and the saddle point approximation was used to compute the integral assuming $n \gg 1$. So, replacing n by the average number of infected leaves in part (ii), $\langle n \rangle = pk$, to estimate $t_2 = \langle t_2^{(n)} \rangle \approx t_2^{(\langle n \rangle)}$, we obtain

$$\tau_k^{\mathcal{T}} \approx \frac{1 + 0.92 \ln(pk)}{\mu} \simeq \frac{0.92}{\mu} \ln k \quad (4.22)$$

for $\lambda \ll \mu$. The prediction is a logarithmic increase with the star size.

4.2.2 Activation mechanisms of the modified dynamics on quenched PL networks

Now we scrutinize the activation mechanism for all models on quenched networks with power-law degree distributions with exponent γ , whose analytical and numerical thresholds were presented in Sec. 4.1.4. We have two different regimes, $2 < \gamma < 3$ and $\gamma > 3$, that will be considered separately in the next sections, discussing the mutual activation of hubs or collective activation mechanisms on these networks.

$\gamma > 3$

Figure 4.5(a) shows the lifespan for the three SIS models analyzed as a function of the size of stars graphs. For the SIS- \mathcal{S} model, exponential growth is obtained, as predicted by Eq. (4.19). However, both SIS- \mathcal{T} and SIS- \mathcal{A} models present lifespans increasing very slowly with graph size, consistent with logarithmic growths. In fact, the approximated discrete-time dynamics presented before result in a logarithmic lifespan for SIS- \mathcal{T} , Eq. (4.22), and finite for SIS- \mathcal{A} , Eq. (4.20).

An upper bound for the long-range infection times of hubs of degrees k and k' , denoted by $\tau_{kk'}^{(\text{inf})}$, for uncorrelated networks can be obtained, resulting in the same value [15, 16] for all investigated SIS models, given by

$$\tau_{kk'}^{(\text{inf})} \leq \tau_{kk'} = \frac{1}{\lambda} \left[\frac{N \langle k \rangle}{kk'} \right]^{b(\lambda)}, \quad (4.23)$$

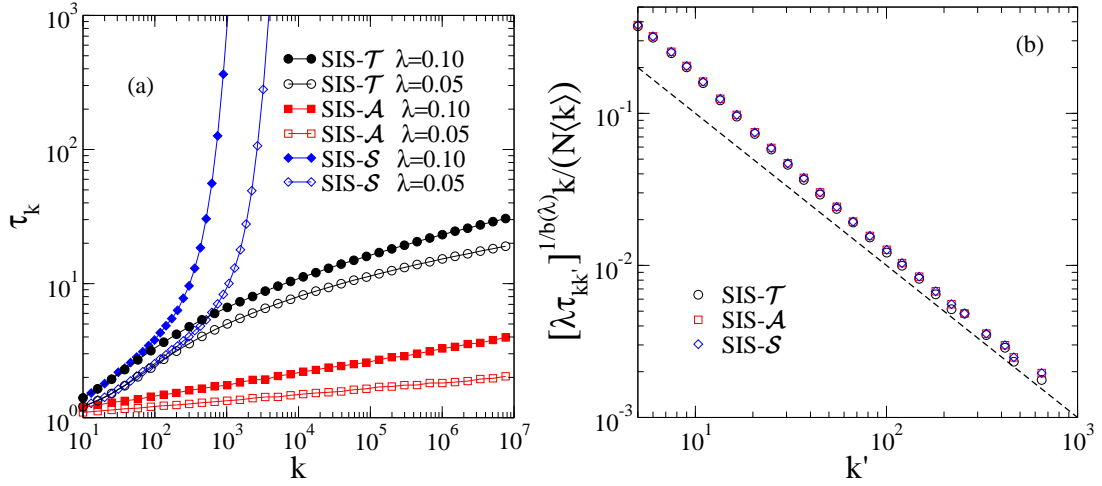


Figure 4.5: (a) Activity lifespan for epidemic processes on star graphs. The initial condition is the center infected and all leaves susceptible. The number of runs varies from 10^3 to 10^5 , the larger number the smaller λ . (b) Mutual reinfection of hubs scaled according to Eq. (4.23). The degree exponent is $\gamma = 3.5$, the size is $N = 10^6$ and infection rate is $\lambda = 0.05$. The vertex kept infected has degree $k = 50$. The dashed line is the prediction of the right-hand side of Eq. (4.23).

in which $b(\lambda) = \ln(1 + \mu/\lambda)/\ln \kappa$ and $\kappa = \langle k^2 \rangle / \langle k \rangle - 1$. Even being rigorously an upper bound, the right-hand side of Eq (4.23) works precisely for $\lambda \ll \mu$ and $\gamma > 3$, such that we can adopt $\tau_{kk'}^{(\text{inf})} \approx \tau_{kk'}$ as done for SIS- \mathcal{S} [15] and other epidemic models [16]. This agreement is confirmed in Fig. 4.5(b) for $\gamma = 3.5$. The simulation is performed by keeping a single vertex of degree $k = 50$ always active, and computing the time the infection takes to reach for the first time each vertex of the network, up to a maximum time $t_{\text{max}} = 10^{10}$. Vertices that were not reached are not included in the averages but they represent a tiny fraction.

With the approximation given by the right-hand side of Eq. (4.23), we have that $\tau_{kk'}^{(\text{inf})} \gtrsim \tau_{k_{\text{max}}, k_{\text{max}}}^{(\text{inf})}$ where k_{max} is the largest degree of the network that scales as $\langle k_{\text{max}} \rangle \sim N^{1/(\gamma-1)}$ for UCM networks with $\gamma > 3$ [70]. Also, we have that $b(\lambda)$ is finite since κ converges to a constant as $N \rightarrow \infty$ for $\gamma > 3$, providing an algebraic increase of $\tau_{kk'}^{(\text{inf})}$ with N . The condition $\tau^{(\text{inf})} \gg \tau_k^{(\mathcal{T}, \mathcal{A})}$ is obeyed such that epidemics in the modified SIS models cannot be activated by hubs when $\lambda \ll \mu$ and a collective phase transition at finite threshold is expected [16] in contrast with $\tau^{(\text{inf})} \ll \tau_k^{(\mathcal{S})}$ of the standard SIS, in which the hub activation mechanism is at work and the threshold is null in the thermodynamic limit.

$2 < \gamma < 3$

For $\gamma < 3$, the hubs are sufficiently close [109] to infect each other even if their activity lifespans are not too large (exponential) and the threshold goes to zero for all models as $N \rightarrow \infty$.

However, there exists a difference in the threshold scaling for $\gamma = 2.7$ but does not for $\gamma = 2.25$. It has been claimed [13] that the most effective spreaders in an epidemic processes lie in a subset containing the innermost core of the networks identified by the maximal index of the k -core decomposition [24, 67]. For SIS- \mathcal{S} , this mechanism is claimed to hold for uncorrelated networks with $\gamma < 5/2$ but the case $5/2 < \gamma < 3$ has activation ruled by the hubs [110]. Since hubs cannot be activated in isolation for arbitrarily small λ in both SIS- \mathcal{A} and SIS- \mathcal{T} , we propose that the epidemic threshold should be ruled by the subgraph identified by the maximal k -core for the whole range of scale-free networks with $2 < \gamma < 3$.

To check what happens for SIS- \mathcal{T} and SIS- \mathcal{A} , we perform simulations in the subgraphs composed by the maximum k -core and in star graphs of size $k_{\max} \approx \sqrt{N}$, the cutoff of the PL networks considered here. Figure 4.6(a) shows that both SIS- \mathcal{S} and SIS- \mathcal{A} have the same activation thresholds in the maximum k -core for $\gamma = 2.25$ and 2.7, although the star activation is different. The results for SIS- \mathcal{T} are essentially the same for SIS- \mathcal{A} . As a consequence, for $\gamma = 2.25$ the maximum k -core is activated before the hubs for all models, being responsible for the endemic state. For $\gamma = 2.7$, hubs are activated before the maximum k -core only for SIS- \mathcal{S} , while the latter is activated first for SIS- \mathcal{A} and SIS- \mathcal{T} models.

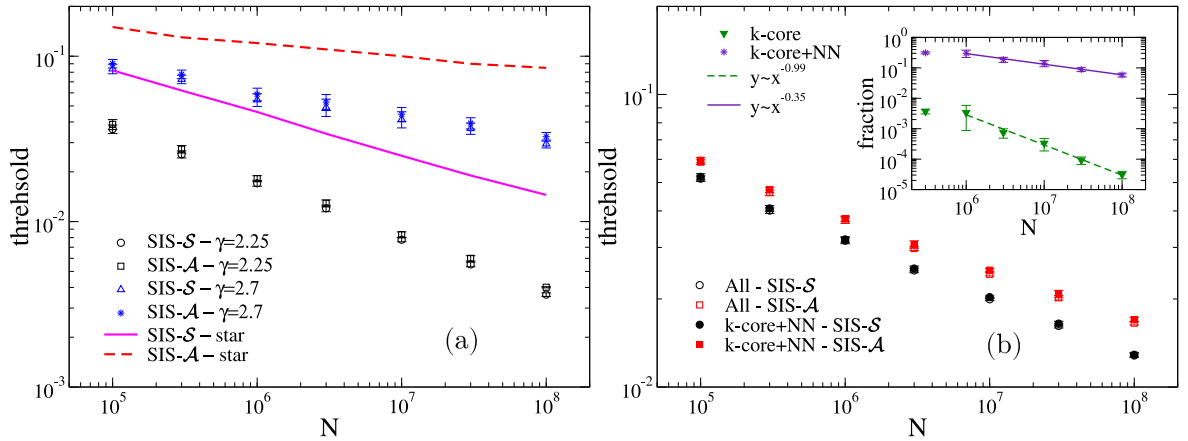


Figure 4.6: Epidemic thresholds for SIS- \mathcal{S} and SIS- \mathcal{A} running on (a) the maximum k -core subgraph of networks with degree exponents $\gamma = 2.25$ and 2.7, and on (b) a subgraph with the maximum k -core plus the nearest-neighbor (NN) vertices of an UCM network with $\gamma = 2.7$. In (a) the simulation results on star graphs with $k_{\max} \approx \sqrt{N}$ leaves are also presented. The inset in (b) shows the fraction of the network that belongs to the maximal k -core including or not its NNs, and lines are power-law regressions. The averages were done over 10 network realizations.

It is important to stress that the thresholds for $\gamma = 2.7$ in the entire network are smaller than those calculated using only the maximal k -core or star centered on the largest hub even with these subgraphs being associated with the activation of the epidemics. We performed

simulations in a subgraph with the maximal k -core plus their nearest-neighbors, which still represents a sub-extensive fraction of the network as shown in the inset of Fig. 4.6(b). The epidemic thresholds in this subset are essentially the same as those of the whole network for all models, as shown in Fig. 4.6(b) for SIS- \mathcal{A} and SIS- \mathcal{S} . The trimming of edges reduces the epidemic activity in the subset containing only the maximal k -core while the k -core mediates the mutual interactions among hubs in the activation driven by them. We see that a large fraction of the network is redundant for the epidemic threshold independently if hub (SIS- \mathcal{S}) or k -core activation (SIS- \mathcal{A} and \mathcal{T}) is at work. In both cases, the relevant region to reproduce the numerical threshold includes the maximal k -core plus its nearest-neighbors.

Returning to the case $\gamma > 3$, in these UCM networks the k -core structure is given by a single component containing the whole network [67]. Then, since hubs do not sustain the activity for $\lambda \ll \mu$ in models SIS- \mathcal{T} and SIS- \mathcal{A} , a collective phase transition involving a finite fraction of the network is obtained, happening in a finite threshold [16]. Table 4.2 summarizes these results, together with what is known [16] for the SIRS and contact process [25].

Table 4.2: Activation mechanisms for different epidemic models presenting active steady states on uncorrelated networks with degree distribution $P(k) \sim k^{-\gamma}$.

Model	$2 < \gamma < 5/2$	$5/2 < \gamma < 3$	$\gamma > 3$
SIS- \mathcal{S}	Max k -core	Hub	Hub
SIS- \mathcal{T}	Max k -core	Max k -core	Collective
SIS- \mathcal{A}	Max k -core	Max k -core	Collective
SIRS	Max k -core	Max k -core	Collective
CP	Collective	Collective	Collective

4.2.3 Finite-size scaling of critical quantities

The finite-size scaling (FSS) of the epidemic threshold is fundamental to characterize the phase transition and critical exponents [25, 94]. The analyses were already done for the SIS- \mathcal{S} [16, 26] model and it is important to investigate what happens in the modified dynamics on annealed and quenched networks. In this case, we consider that the QS quantities follow power-laws with the size of the network as

$$\rho \sim N^{-\nu} \quad \text{and} \quad \chi \sim N^{\phi}, \quad (4.24)$$

where ν and ϕ are the exponents. The analyses were performed both for quenched and annealed networks. Here we consider a hard cutoff in the degree distribution $k_c \sim N^{1/\gamma}$ for $\gamma = 3.5$ that

prevents outliers in the degree distribution and multiple peaks in the susceptibility curves of quenched networks [17, 26] making, thus, the determination of the transition point much more accurate. In the other cases, a cutoff $k_c = \sqrt{N}$ was used. Figure 4.7 shows the results for simulations and Table 4.3 the estimated exponents by power-law regressions for $N \geq 10^6$. Uncertainties were calculated using different fit regions aiming at establishing equivalences or discrepancies between annealed and quenched simulations rather than accurate estimates of the asymptotic exponents.

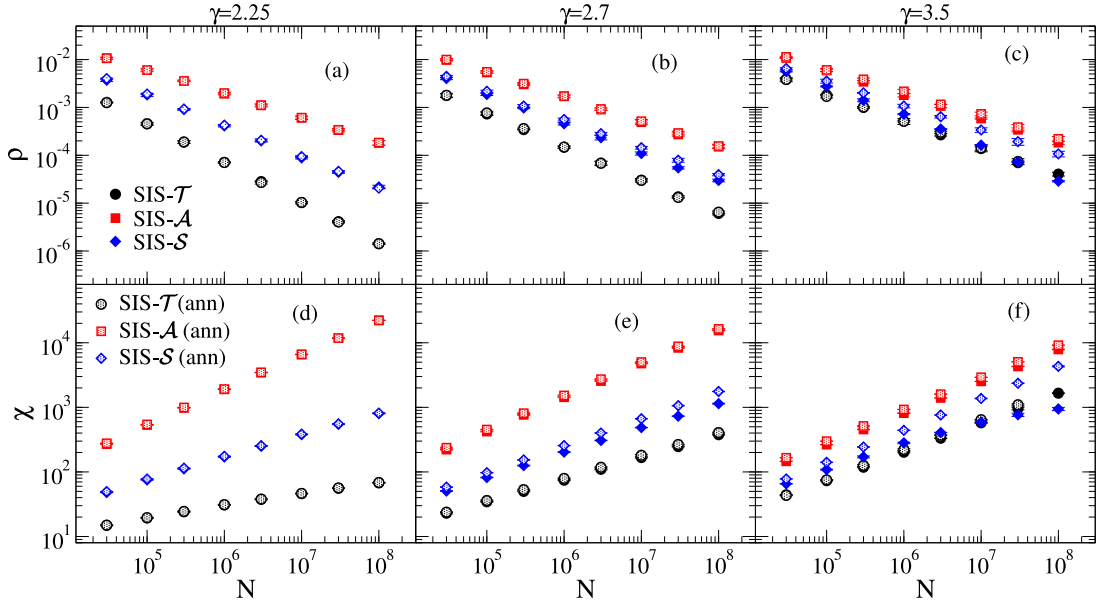


Figure 4.7: Finite-size scaling of the critical QS quantities for SIS models on UCM networks with different degree exponents. The QS densities of infected vertices are shown in (a)-(c) while the QS susceptibilities are shown in (d)-(f). Data correspond to averages over 10 network realizations and error bars are smaller than symbols.

The FSS of the critical quantities provides a scenario in consonance with that observed for the thresholds. The FSS of both SIS- \mathcal{T} and \mathcal{A} are in full agreement with annealed simulations showing their mean-field behaviors for all values of γ investigated. Moreover, the agreement between quenched and annealed networks is also found for SIS- \mathcal{S} for $\gamma < 5/2$. For $\gamma > 5/2$, the dichotomy for SIS- \mathcal{S} is again present. A significant difference in the scaling happens for $\gamma = 2.7$ and a sharp difference is obtained for $\gamma = 3.5$. In the latter, we can see a susceptibility curve of the SIS- \mathcal{S} bending downwardly for the quenched network, which has been associated with a smearing of the phase transition [43], while in the annealed network a power-law typical of an ordinary critical phase transition is seen. No sign of smearing is observed for SIS- \mathcal{T} and \mathcal{A} .

The FSS provides different exponents for distinct models. So, despite being described by the same mean-field equations, the role played by stochastic fluctuations depends on the model.

Table 4.3: Critical exponents of the FSS for the SIS models on UCM (ν and ϕ) and annealed (ν_{ann} and ϕ_{ann}) networks. Exponents for SIS- \mathcal{S} with $\gamma = 3.5$ are missing due to the smearing of the transition in quenched networks.

Model	$\gamma = 2.25$		$\gamma = 2.7$		$\gamma = 3.5$	
	ν	ν_{ann}	ν	ν_{ann}	ν	ν_{ann}
SIS- \mathcal{T}	0.845(6)	0.84(2)	0.697(4)	0.692(6)	0.55(1)	0.555(3)
SIS- \mathcal{A}	0.519(9)	0.517(4)	0.52(1)	0.515(9)	0.499(6)	0.49(3)
SIS- \mathcal{S}	0.63(2)	0.655(2)	0.60(2)	0.57(1)	—	0.506(7)
	ϕ	ϕ_{ann}	ϕ	ϕ_{ann}	ϕ	ϕ_{ann}
SIS- \mathcal{T}	0.167(2)	0.169(1)	0.353(1)	0.352(1)	0.458(1)	0.467(3)
SIS- \mathcal{A}	0.530(2)	0.528(2)	0.514(1)	0.513(1)	0.494(1)	0.497(1)
SIS- \mathcal{S}	0.329(5)	0.329(4)	0.372(1)	0.421(1)	—	0.496(1)

4.2.4 Accuracy of the QMF theory and modified dynamics on real networks

We simulated the three models in 102 networks based on a list of Refs. [73, 111]. We also computed correlations with a measure of the level of heterogeneity of each network, defined as $\eta = \langle k \rangle^2 / \langle k^2 \rangle$, that decreases with the heterogeneity (darker colors in Fig. 4.8). As seen in synthetic networks, epidemic thresholds for SIS- \mathcal{A} and SIS- \mathcal{T} tend to agree with each other, with exception of few networks, as in Fig. 4.8(a). In general, epidemic thresholds of SIS- \mathcal{A} are smaller than those of the SIS- \mathcal{T} dynamics, while in some networks they are essentially the same.

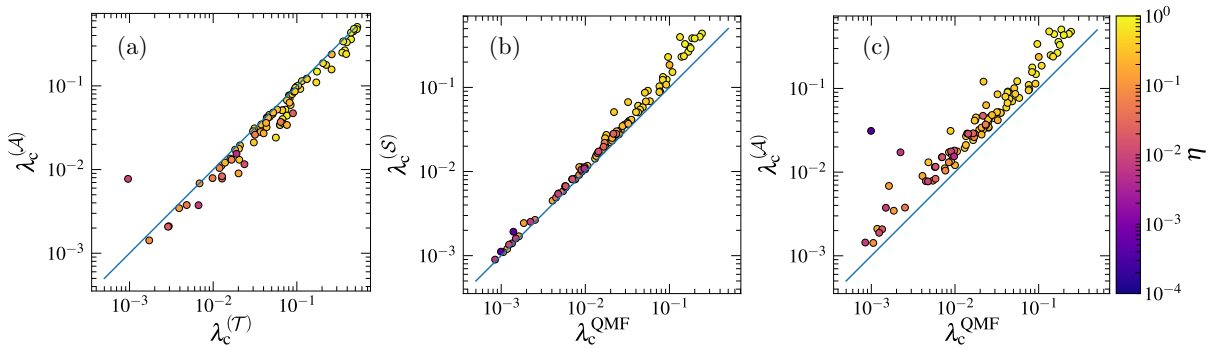


Figure 4.8: (a) Epidemic thresholds for SIS- \mathcal{T} and SIS- \mathcal{A} models. Comparison between QMF theory and simulations for (b) SIS- \mathcal{S} and (c) SIS- \mathcal{A} dynamics. Each point is one of the 102 investigated real networks, and colors are related to their level of heterogeneity $\eta = \langle k \rangle^2 / \langle k^2 \rangle$.

Next, we compare the accuracy of the QMF theory for the epidemic thresholds of SIS- \mathcal{S} , in Fig. 4.8(b), and SIS- \mathcal{A} in Fig. 4.8(c). In the first case, the QMF theory predicts quite well the epidemic thresholds of the investigated real networks in the standard dynamics, except for networks with a lower level of heterogeneity. Finally, for the SIS- \mathcal{A} we find a correlation between the epidemic thresholds predicted by QMF and the ones from simulations, but the agreement is not good. For further discussions regarding the accuracy of QMF, HMF, and pair-wise theories on real networks, see Ref. [73].

4.3 Summary and discussions

The conception of theoretical frameworks for epidemic processes frequently passes over the model's fine-tuning due to the belief that universality takes over and all central features, related to the leading properties and symmetries of a system, will be obtained irrespective of the specific details. However, this does not seem to be always the case when the substrate carrying out the process is a complex network. The standard SIS model, called SIS- \mathcal{S} in this work, is an example that behaves very differently from most of other related processes.

We investigated two slightly different versions of this model, presenting exactly the same thresholds of the SIS- \mathcal{S} in both QMF and HMF theories. Stochastic simulations on uncorrelated synthetic networks, however, show a dual scenario where the three models have essentially the same vanishing thresholds for $\gamma < 5/2$ but disparate results are found for $\gamma > 5/2$. The dissonance is explained in terms of self-sustained, long-lived activation of hubs for any finite value of λ [15] that holds for SIS- \mathcal{S} but does not for the other models.

We also analyzed the activation mechanisms of the epidemic phase on uncorrelated networks. While the activation for SIS- \mathcal{S} occurs in the innermost, densely connected core of the network, determined by the largest index of a k -core decomposition, for $\gamma < 5/2$ and in hubs for $\gamma > 5/2$ [110], this happens in the hubs for the whole range of scale-free networks with $2 < \gamma < 3$ for SIS- \mathcal{A} and SIS- \mathcal{T} . The aforementioned dichotomy is also observed in the finite-size scaling of the quasi-stationary density and susceptibility computed at the epidemic threshold. Agreements between simulations on quenched and annealed versions of the investigated networks are observed for SIS- \mathcal{A} and SIS- \mathcal{T} irrespective of the degree exponent. In turn, they deviate in the hub activated regime with $\gamma > 5/2$ in SIS- \mathcal{S} , being more marked for $\gamma > 3$ where the transition observed for quenched networks seems to be smeared [43], in contrast with a regular critical transition in the annealed case.

An interesting point observed in our analysis is that the HMF theory was more accurate than QMF theory in all investigated cases, except for SIS- \mathcal{S} . Dynamical correlations are neglected in both approaches assuming that the states of interacting vertices, in the case of QMF, or interacting compartments, in the case of HMF, are independent. This approximation becomes more problematic for QMF since we explicitly reckon the interactions with the actual nearest-neighbors of each vertex and assume that their states are independent. The leading approximation in HMF is to assume that the probability to be infected depends only on the

vertex degree, neglecting the local structure of the network. As an effect, HMF theory may not be able to capture localized activity due to specific motifs as those observed for star sub-graphs in the SIS- S model. Finally, QMF theory is not a genuine mean-field approach since it does not present mixing of vertices while HMF does through the degree compartmentalization. Our results thus reinforces the belief that mean-field approaches with heterogeneous mixing are suitable approximations for most dynamical processes on networks with a small-world property.

Chapter 5

Griffiths phases on non-hierarchical modular random networks

Related publications:

Griffiths effects of the susceptible-infected-susceptible epidemic model on random power-law networks [43]

Wesley Cota, Silvio C. Ferreira, Géza Ódor

Physical Review E **93**, 032322 (2016)

Griffiths phases in infinite-dimensional, non-hierarchical modular networks [44]

Wesley Cota, Géza Ódor, Silvio C. Ferreira

Scientific Reports **8**, 9144 (2018)

A fundamental question is why and how a complex system is led to criticality, in which spatial and temporal correlations diverge [36, 37]. As a very important example, experimental evidence suggests that the brain operates close to criticality [38, 45, 112–114], with optimization of information processing capabilities and sensibility to external stimuli as a result of chemical interactions. Simple models on homogeneous substrates [38, 39] have been frequently used to answer this question with self-organizing criticality [40] or an evolutionary selection mechanism [41]. In any case, heterogeneities on the networks mediating these interactions are fundamental in the behavior of dynamical processes, and it is a big challenge to understand how disorder induced by these heterogeneities affects the critical state.

5.1 Effects of quenched disorder on criticality of complex systems

In condensed matter physics, quenched disorder, with timescales much longer than those of the dynamics, can lead to the so-called Griffiths phases (GPs) [42] with dynamical criticality and high sensitivity to external stimuli in an extended parameter space [96]. These phases are

consequences of rare regions (RRs), that are sub-extensive local and active domains, sustaining activity for long times. Let us start with an example of this phenomenon in a ferromagnetic system with quenched disorder.

5.1.1 Griffiths phases on ferromagnetic systems

In a ferromagnetic system, normally there exists a critical temperature T_c above which the spontaneous magnetization does not happen and the free energy is an analytical function of the temperature T and magnetic field h . Critical systems can be sensitive to quenched disorder when its dimension is sufficiently small [96]. In 1969, Griffiths showed [42] that for a diluted ferromagnetic system in Fig. 5.1, even when the spontaneous magnetization is zero the free energy continues non-analytical in an extended region of the parameter space called Griffiths phase. Consider, for example, a 2D lattice ($d = 2$) in which vertices can have a spin with probability q or a vacancy with probability $p = 1 - q$, shown in Fig. 5.1(a). The occupation fraction q is fixed and independent of T and h , defining a system with quenched disorder. As a result, the critical temperature T_c will now be a function of p (or q). For $p = 0$, the critical temperature is T_c^0 , the same of the system without disorder, as seen in Fig. 5.1(b). As p increases from 0, the system reaches the percolation threshold p_c above which isolated patches of the original system coexist but are rare, exemplified by the shaded area in Fig. 5.1(a).

In real systems, the disorder can be represented by vacancies, as in the previous example, but also impurities and atom position displacement in the lattices. The important fact is that rare but large fluctuations, in general, can lead to Griffiths singularities close to the original critical region in the phase diagram, resulting in a GP. In non-equilibrium quantum systems, for example, these regions can have even stronger effects, such as power-law singularities in the

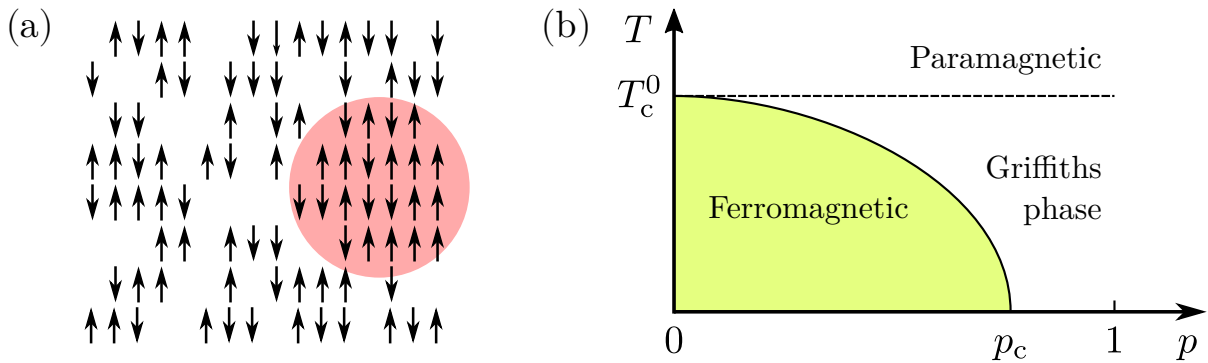


Figure 5.1: (a) Example of a diluted ferromagnetic system. The shaded region does not present vacancies, in which spins are removed with probability p . (b) Schematic phase diagram of a diluted system, presenting a GP. Adapted from Figs. 1 and 2 of Ref. [96].

free energy or even the destruction of the phase transition [96]. In the previous example, we can calculate [96] the probability of finding a region devoid of vacancies. A region of volume L_{RR}^d , being d the system dimension, happens with probability

$$w(L_{\text{RR}}) \sim (1 - p)^{L_{\text{RR}}^d} = \exp\left(-\tilde{p}L_{\text{RR}}^d\right),$$

with $\tilde{p} = -\ln(1 - p)$. Although these regions are exponentially rare, their contributions

$$M_{\text{RR}} \sim \exp\left(-\tilde{p}b^d/h\right)$$

for the magnetization are not, even in the paramagnetic phase, where b is a constant. In the following sections, we show that similar effects can also happen with dynamical processes on complex networks.

5.1.2 Griffiths phases in the epidemic spreading

Heterogeneity is an intrinsic feature of complex networks, in which the process rates are dependent on the vertices' properties. In the SIS model, for example, hubs have a larger infection capacity than the other vertices, representing a kind of disorder in the system. Moreover, vertices can have different healing rates, whose consequences will be discussed in Sec. 5.3. This intrinsically disordered nature of networks calls for analogs of GP and RR phenomena. This issue has recently been investigated [115, 116] and GPs have been found in the contact process (CP) [105] on finite-dimensional networks. The difference between finite- and infinite dimensional networks was discussed in Sec. 2.1.2.

To illustrate a Griffiths phase on dynamical processes, consider a spreading process with a transition between inactive and active phases, such as in the SIS model. The control parameter is the infection rate λ , while the order parameter is the fraction of infected vertices ρ . A global active phase exists for $\lambda > \lambda_c$, with non-zero density ρ for $t \rightarrow \infty$ (stationary limit), or zero for $\lambda < \lambda_0$. Rare regions are responsible for a very slow relaxation for $\lambda_0 < \lambda < \lambda_c$, leading to scale invariance even in the inactive phase ($\rho = 0$ for $t \rightarrow \infty$ in this case). The susceptibility χ is a kind of response function of the system, with large values when dynamical fluctuations are high, portraying a high sensibility to external stimuli. This regime constitutes the hallmarks of a Griffiths phase and Fig. 5.2 shows its qualitative behavior.

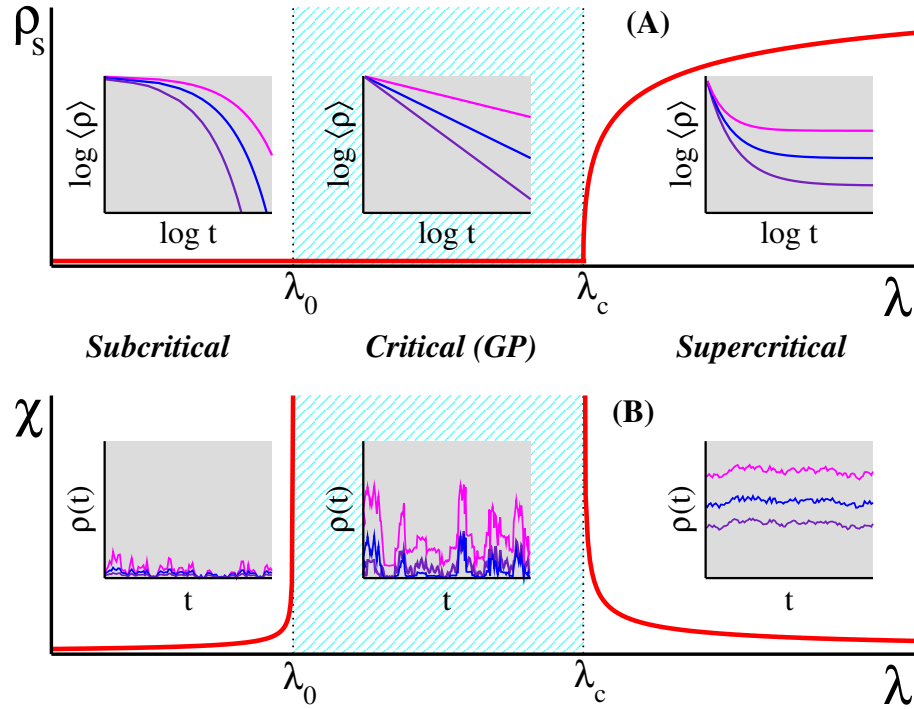


Figure 5.2: Griffiths phases and dynamical criticality on spreading processes. The top graph shows the stationary density ρ_s as a function of the infection rate λ . The order parameter has a finite value above the critical point λ_c and goes to zero as $\lambda \rightarrow \lambda_c$. This density is null for $\lambda < \lambda_0$. In both cases, the density decays in time are typically given by exponential curves, with finite characteristic times. The asymptotic density is still null in the critical region, but the decay is slow, typically described by a power law. The bottom graph shows the dynamical susceptibility [17, 43] as a function of the infection rate. Fluctuations are finite outside the critical region, while diverge as it is approached.

5.1.3 Conjectures about Griffiths phases on small-world networks

As discussed in Chapter 4, the standard SIS dynamics is puzzling due to the highly fluctuating size and the number of hubs realized in a network sample [17, 26, 27, 100] in finite random networks. Indeed, in the standard SIS dynamics, the effective finite-size epidemic threshold is finite since $k \ll 1/\lambda^2$ for sufficiently small values of λ and star subgraphs alone cannot sustain a long-term activity, as discussed in Sec. 4.2.1; see Eq. (4.19). Several network realizations have just a few vertices with a degree much larger than the rest of the network and these outliers can sustain localized epidemics, with different activation thresholds for very long times, producing multiple activations [17].

Neglecting interactions among hubs, Lee *et al.* [27] predicted that the threshold for an endemic phase, in which a finite fraction of infected vertices is present, should take place at a finite value for $\gamma > 3$, where $\langle k^2 \rangle$ is finite. According to Lee *et al.* [27], the subcritical region is ruled by a GP, with an ultraslow (logarithmic) decay of activity, in odds with the rigorous

results of a null threshold for infinite PL networks [9]. This proposition was supported by numerical simulations on very large, but finite, random PL networks [17]. Finally, in finite PL networks, a vanishing epidemic threshold is predicted by the QMF theory [11, 117]; see Sec. 4.1.2. In such models, Griffiths effects were also shown in the localized phase for $\gamma > 3$ [118]. RR effects, localization, and heavy-tailed dynamics have also been shown in spreading models defined on weighted PL networks by suppressing hub infection via disassortative weight schemes [119, 120], in random networks [121, 122], or aging Barabasi-Albert graphs [121].

There exists a hypothesis based on these spreading models that affirm that heterogeneity becomes irrelevant in the thermodynamic limit in the case of networks of infinite dimension [115], that is the case of small-world networks. In this work, an investigation is done to check if both the hierarchy and finite dimensions are necessary conditions for GPs. First, we start by considering an *ensemble* of independent samples of random networks with power-law degree distribution. We show that the averaging over many independent graph realizations exhibits a slow dynamics, analogous to GPs, in an interval of control parameter $\lambda_1 < \lambda < \lambda_2$. This region is delimited by two transitions: The former is related to the activation of the most connected hub of the network, while the latter is related to a smeared phase transition [96]. In this chapter, we perform analysis over independent, finite networks and results indicate that this region shrinks as the size of the network increases and disappears in the thermodynamic limit, implying the absence of GPs. This is in agreement with the conjecture that finite dimensionality is required for the existence of GPs [115]. However, we also investigate these effects in weakly interconnected networks, forming a non-hierarchical modular structure, and find robust GPs that seem to hold in the thermodynamic limit. This discussion is presented in Sec. 5.3.

5.2 Independent power-law networks with cutoffs

To investigate the Griffiths effects and their relation with the presence or not of outliers we considered three different cutoffs k_c for the degree distribution: $k_c = N$ (free), $k_c = k_0 N^{0.9/(\gamma-1)}$ (hard) and $k_c = \sqrt{N}$ (structural). The first one, also called natural cutoff, leads to degree distributions in which both the average and the standard deviation of the highly fluctuating natural cutoff diverge as $N^{1/(\gamma-1)}$ [17]. Conversely, the second one is engineered to render distributions without very large gaps in their tails of the degree distribution, since the factor 0.9 guarantees that $\sqrt{\langle k_{\max}^2 \rangle - \langle k_{\max} \rangle^2} / \langle k_{\max} \rangle \rightarrow 0$ as $N \rightarrow \infty$ [17]. The structural cutoff is not

fluctuating and guarantees the absence of degree correlations for $\gamma < 3$ and becomes equivalent to the absence of a cutoff for $\gamma > 3$ and $N \rightarrow \infty$ [70]. Figure 5.3 shows the qualitative behavior of $P(k)$ for different cutoffs and values of γ . Graph edges are generated randomly, forbidding multiple and self-connections, and using the configuration model defined in Sec. 2.3.2. All simulations were performed for $k_0 = 3$. Three ranges of the degree exponents were considered separately: $\gamma > 3$, for which localization is conjectured by the QMF theory; $\gamma < 5/2$, where delocalized [27] or a k -core [110] epidemic spreading activation are predicted; and $5/2 < \gamma < 3$ strict SF regime, where localization is conjectured by the QMF theory [100, 110]. It is important to stress that in this and next chapters we only consider the standard SIS- S model investigated in Chapter 4.

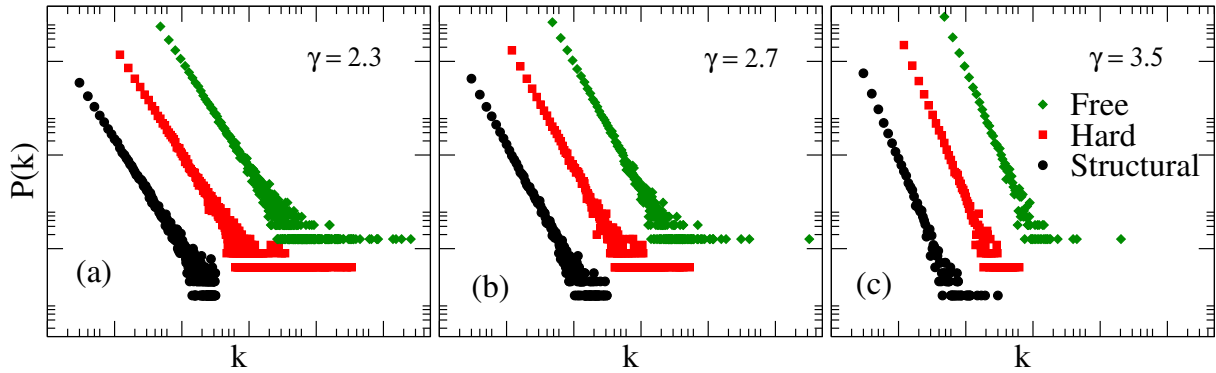


Figure 5.3: Qualitative behavior of curves for degree distributions considering different cutoffs (indicated in legends) and exponents (a) $\gamma = 2.3$, (b) 2.7, and (c) 3.5. Note that the presence of outliers is common using the free cutoff, while the hard cutoff is able to omit them.

We performed both standard (over time) and quasi-stationary (QS) analysis; see Sec. 3.3.3 for details. Differently from previous analysis of SIS on SF networks [11, 15, 17, 26], we aim at the average behavior of a large number M of independent realizations. Firstly, we start performing a time-dependent analysis and discuss an optimal fluctuation theory in Sec. 5.2.3 and a finite-size analysis in Sec. 5.2.4. Later, in Sec. 5.2.5, we present the results and discussions for the QS analysis.

5.2.1 Time-dependent analysis

In the decay analysis, 10 to 100 independent runs were performed for each network, being the largest number of runs used for the smallest values of the infection rate. In this section, we discuss the results for the time-dependent analysis on networks with the different cutoffs by starting from a fully infected initial state.

Free cutoff

Simulations for networks with free cutoff are shown in Fig. 5.4. We observe an extremely slow logarithmic decay in an extended region of the control parameter λ . For $\gamma = 3.5$, the decay is very well fitted by

$$\rho \sim (\ln t)^{2-\gamma}. \quad (5.1)$$

For $\gamma = 2.7$, in the SF regime with large fluctuations on the degree distribution, we also see a logarithmic decay $\rho \sim (\ln t)^{-\alpha}$, with a varying exponent α that is not quantitatively well described by Eq. (5.1). The origin of the logarithmic decay given by Eq. (5.1) is related to the presence of the outliers in the network and will be analytically discussed in Sec. 5.2.3 using an optimal fluctuation theory.

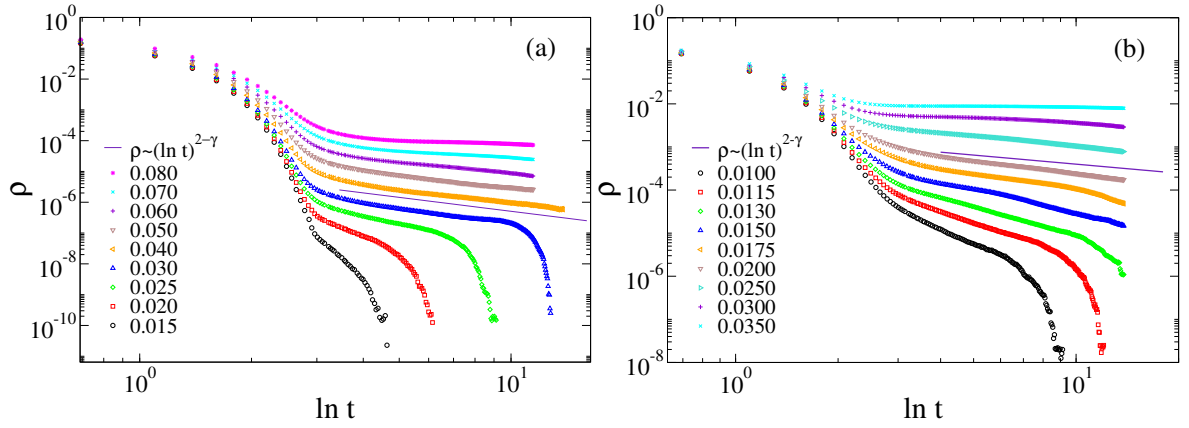


Figure 5.4: Decay of density of infected vertices for networks with a free upper cutoff ($k_c = N$) using (a) $\gamma = 3.5$ and (b) $\gamma = 2.7$ for networks with sizes $N = 10^7$ and $N = 10^5$, respectively. The number of samples were up to $M = 500$ and 4000 for $\gamma = 3.5$ and 2.7 , respectively. Lines are predictions of the optimal fluctuation theory in Sec. 5.2.3. The values of λ are indicated in the legends.

Hard cutoff

The localization of the QMF theory for $\gamma > 5/2$ is concentrated around the largest hub [123]. So, the role played by the hubs can be evidenced by damping their number and fluctuations. Evolution of the density of infected vertices for a hard cutoff of $P(k)$ is shown in Fig. 5.5 for $\gamma = 3.5$ and 2.7 . The data indicate PL decay with non-universal exponents at long times for both degree exponents. Regressions fits: $\rho \sim t^{-\alpha(\lambda)}$ at $\gamma = 3.5$ resulted in α varying from 0.70 to 0.17 by increasing λ from 0.088 to 0.095. Similar range of exponents were found for $\gamma = 2.7$, varying λ from 0.030 to 0.0365.

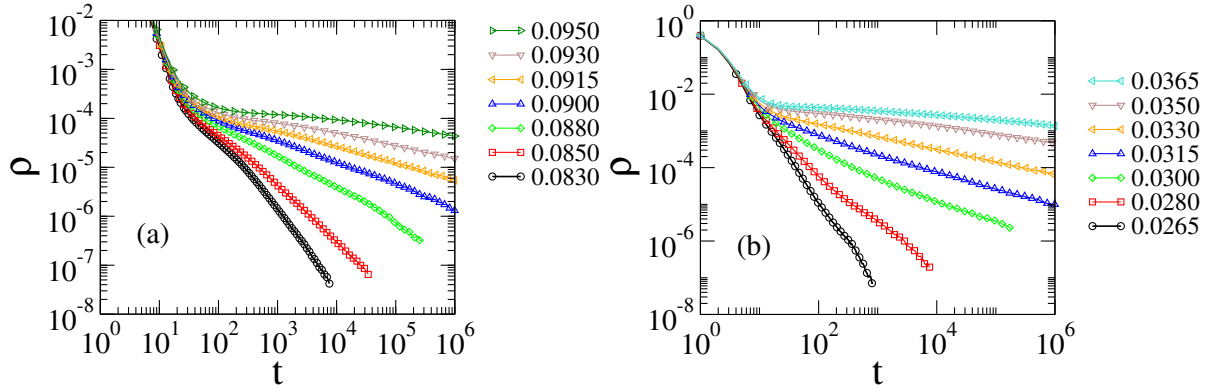


Figure 5.5: Decay of density of infected vertices for (a) $\gamma = 3.5$ and (b) $\gamma = 2.7$ for networks of sizes $N = 10^7$ and 10^5 , respectively, using hard cutoff $k_c = k_0 N^{0.9/(\gamma-1)}$. The numbers of independent networks are $M = 500$ and 4000 for $\gamma = 3.5$ and 2.7 , respectively. The values of λ are indicated in the legends.

Structural cutoff

We also simulated the density decay in networks with the structural cutoff $k_c = \sqrt{N}$ with $\gamma < 3$ since otherwise it is equivalent to the natural one. This cutoff leads to the uncorrelated configuration model (UCM) [82], that has been used in many analyses of SIS on SF networks [11, 15, 26, 27, 97, 110]. Power-law decays in time are still observed, but the extended region is reduced compared with hard cutoffs, see Fig. 5.6.

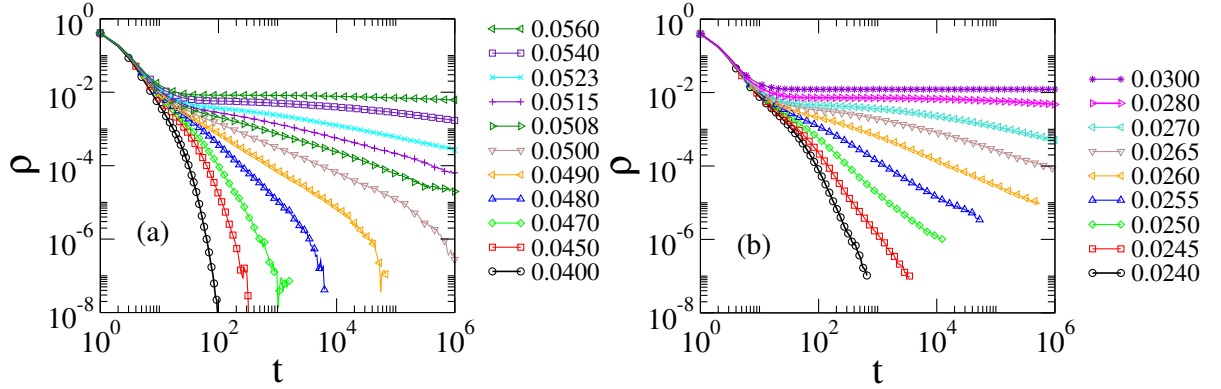


Figure 5.6: Decay of density of infected vertices for (a) $\gamma = 2.7$ and (b) $\gamma = 2.3$ for networks of size $N = 10^5$ using structural cutoff $k_c = \sqrt{N}$. The values of λ are indicated in the legends.

5.2.2 Sample-to-sample fluctuations

The origin of the observed slow decays is the sample-to-sample fluctuations, rather than the occurrence of rare regions in the same network. In Fig. 5.7 we present the decay of the density for 50 networks with all parameters fixed to values for which slow decays are observed in the averaged curves. Several curves are subcritical, while others behave super-critically before

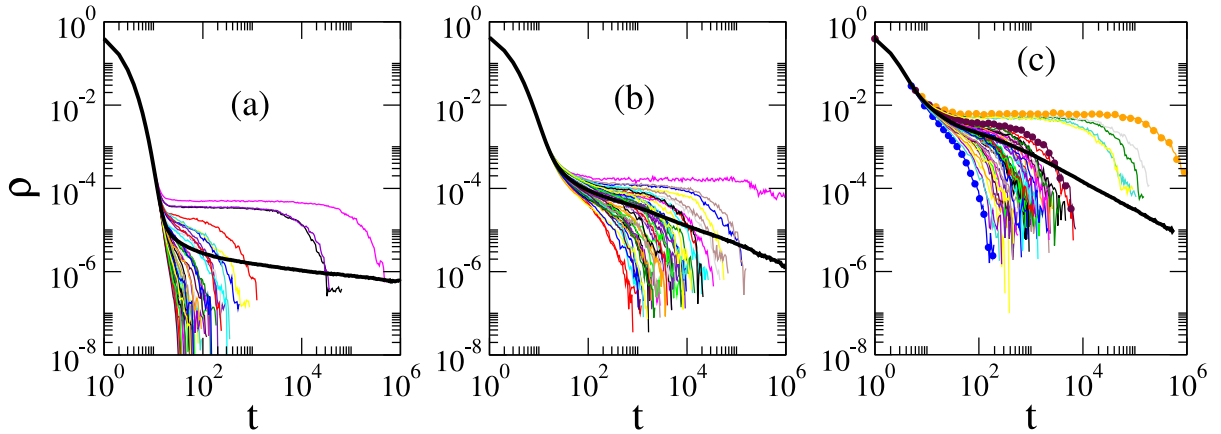


Figure 5.7: Sample-to-sample fluctuation of the evolution of SIS on PL networks: (a) $\gamma = 3.5$, $N = 10^7$, free cutoff, and $\lambda = 0.04$; (b) $\gamma = 3.5$, hard cutoff $k_c = k_0 N^{0.9/(\gamma-1)}$, $N = 10^7$, and $\lambda = 0.09$; (c) $\gamma = 2.3$, structural cutoff $k_c = \sqrt{N}$, $N = 10^5$, and $\lambda = 0.026$. Curves for 50 independent networks are shown. Thick lines represent the average of $M = 500$ and 2000 samples for $\gamma = 3.5$ and 2.3, respectively. Three samples in (c) are highlighted with symbols and discussed in the main text.

falling into the absorbing-state. There exist two main mechanisms for this large variation, and the leading one depends on γ and the cutoff used.

For a free cutoff, Fig. 5.7(a), the size of hubs fluctuates greatly and the presence of outliers, creating local active domains, determines if the dynamics goes to a metastable state in that sample. In the cases of hard or structural cutoffs, Figs. 5.7(b) and (c), respectively, these fluctuations are smaller and another mechanism takes place, which is the variation of the overall heterogeneity of the network. This can be measured by the average degree of the nearest neighbors of the vertices k_{nn} . For the structural cutoff case this becomes $k_{nn} = \langle k^2 \rangle / \langle k \rangle$ [70] (UCM network), whose inverse provides a very precise estimate of the SIS epidemic threshold for $\gamma < 5/2$; see Sec. 4.1.3.

Density decays for networks with $\gamma = 2.3$, using a structural cutoff, are shown in Fig. 5.7(c). Three samples highlighted with symbols possess $\langle k \rangle / \langle k^2 \rangle = 0.0270$, 0.0243, and 0.0235, and the larger values the lower densities. These values must be compared with the infection rate $\lambda = 0.026$ used in all samples. We see that samples for which $\langle k \rangle / \langle k^2 \rangle \lesssim \lambda$ are supercritical and those where $\langle k \rangle / \langle k^2 \rangle \gtrsim \lambda$ are subcritical. An optimal fluctuation theory to explain these slow dynamics is presented in the next section.

5.2.3 Optimal fluctuation theory

The hypothesis drawn in Sec. 5.2.1, in which the slow decay is originated from sample-to-sample fluctuations of the effective thresholds in finite-size networks can be put in a mathemat-

ical grounds by approximating the sample average with an integral

$$\bar{\rho} = \int_0^\lambda \rho(\lambda, \lambda_c) \mathcal{P}(\lambda_c) e^{-t/\tau(\lambda, \lambda_c)} d\lambda_c. \quad (5.2)$$

Here, $\rho(\lambda, \lambda_c)$ is the quasi-static density as a function of $\lambda > \lambda_c$ ($\rho \equiv 0$ for $\lambda < \lambda_c$), $\tau(\lambda, \lambda_c)$ is the lifetime of the dynamical processes and $\mathcal{P}(\lambda_c)$ is the probability density that a randomly selected sample has a threshold at λ_c .

We assume that in a free cutoff network with $\gamma > 5/2$ the activation happens at the most connected hub [110]; see Table 4.2. Consider a star subgraph, centered on the vertex of maximal degree k_{\max} , which forms an independently activated domain in a network with N vertices. Using QMF theory, the threshold in such a star graph is $\lambda_c \sim 1/\sqrt{k_{\max}}$ [26, 97]. The density in a star of size k is¹ $\rho_{\text{star}} \approx \lambda$ for $\lambda \gtrsim \lambda_c$, implying that

$$\rho \approx \frac{\lambda k_{\max}}{N}. \quad (5.3)$$

The lifespan of the activity in a star in case of SIS dynamics is given by Eq. (4.19) in the previous chapter, or simply $\tau \approx \tau_0 \exp(a\lambda^2 k_{\max})$ where a and τ_0 are constants. Finally, the probability of a given threshold is $\mathcal{P}(\lambda_c) d\lambda_c = \Pi(k_{\max}) dk_{\max}$, where

$$\Pi(k_{\max}) \simeq N \exp\left(-cN k_{\max}^{-\gamma+1}\right) k_{\max}^{-\gamma} \quad (5.4)$$

is the probability of the largest degree to be k_{\max} in a PL network with N vertices [70]. Here c is a constant depending on $P(k)$. Plugging Eqs. (5.3)-(5.4) into Eq. (5.2), we obtain

$$\bar{\rho} \sim \lambda \int_{1/\lambda^2}^{\infty} k_{\max}^{-\gamma+1} \exp\left(-cN k_{\max}^{-\gamma+1}\right) \exp(-t/\tau) dk_{\max}. \quad (5.5)$$

If $1/\lambda^2 \gg N^{1/(\gamma-1)}$, then the exponential suppresses the integral and a standard subcritical phase with exponential decay is expected. For $1/\lambda^2 \lesssim N^{1/(\gamma-1)}$, the first exponential is approximately 1. After an integration by parts this integral can be easily evaluated using the saddle-point method to return $\bar{\rho} \sim (\ln t)^{2-\gamma}$, exactly the result of Eq. (5.1). This decay is the same found by Lee *et al.* [27], in a theory of SIS dynamics for infinite PL networks, with non-interacting hubs. This predicts ultraslow decay instead of a stationary endemic state, contradicting the exact result of the null epidemic threshold for SIS irrespective of γ [9] and a question is why our simulations

¹The actual density in a star must increase a $\rho \sim \lambda^{\beta_{\text{star}}}$, where $\beta_{\text{star}} > \beta_{\text{QMF}} = 1$.

match this theory. The assumption that stars form independent domains of activity is incorrect in principle, since the lifetime of epidemics on stars can be sufficiently large to permit mutual infection of hubs, even if they are not directly connected due to the small-world property [15, 16]. However, several stars that contributed to the average epidemic activity in our simulations are observed in different realizations of networks and are thus actually independent, being the reason for the agreement.

In the SF regime at $\gamma = 2.7$ the QMF still predicts localization, but there is a high probability that several activated hubs occur in the same network sample even if their size is finite. Thus neglecting the multiplicity of activated hubs as well as the interaction among them [15] is not a quantitatively accurate approximation but it can capture the essentially logarithmically slow dynamics observed in simulations.

In case of hard cutoff, we do not know the form of $\mathcal{P}(\lambda_c)$. Since the fluctuations of λ_c depend on global properties of the networks, we assume their distribution to be Gaussian, with width $\sigma(N)$ and centered at $\lambda_0(N)$, tending to a delta function at $\lambda = 0$ as $N \rightarrow \infty$, in conformity with numerics. Less is known about the lifespan. Numerically, we have data consistent with $\tau \sim \exp[a(\lambda - \lambda_c)^2]$, where $a(N)$ is some function increasing with the size that we could not determine precisely. Plugging these forms into Eq. (5.2) and using the saddle-point approximation to solve the integral we found $\bar{\rho} \sim t^{-1/2a\sigma^2}$. This is a non-universal power law, in agreement with the density decay simulations, as far as $a\sigma^2$ is a non-universal constant.

5.2.4 Finite-size analysis

The slow dynamics observed in the ensemble averages is not a genuine Griffiths singularity, since it is not triggered by the slowly decaying RRs; thus we can expect that these effects disappear in the thermodynamic limit. This conjecture is confirmed for a hard cutoff in Fig. 5.8, where we show the density of infected vertices against time for different sizes for a fixed infection rate $\lambda = 0.088$. We see that the dynamics is deeply subcritical, with exponential decay of activity, for $N = 10^6$. For size $N = 10^7$ a PL regime can be observed but, finally, a saturation to a constant plateau develops at $N = 10^8$. The disappearance of the PL regimes is mainly associated with the shift of the epidemic threshold towards zero as the size increases [26, 97]. The threshold drops from approximately 0.12 for $N = 10^6$ to 0.075 for $N = 10^8$. Similar finite-size effects were observed in the CP on weighted trees [119] and the same mechanism shown here is probably also present there. The CP also exhibits strong finite-size dependence of the thresholds, approaching

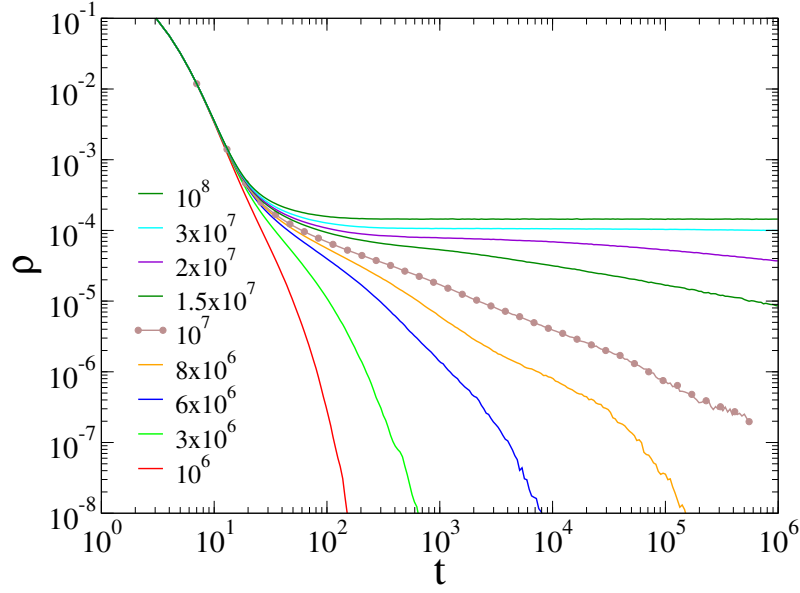


Figure 5.8: Finite-size analysis of the density decay against time for networks with $\gamma = 3.5$ and hard cutoff $k_c = k_0 N^{0.9/(\gamma-1)}$ for infection rate $\lambda = 0.088$. The number of network samples is 100 for the two largest sizes and 500 for the others. The network sizes are indicated in the legend.

the asymptotic value only at exceeding large networks [32, 33]. Moreover, the range of λ where PLs are observed decreases as $N \rightarrow \infty$ (see also Sec. 5.2.5), thus Griffiths effects disappear in the thermodynamic limit for an ensemble of independent networks.

5.2.5 Quasi-stationary analysis

Using the free and hard cutoffs on a single network with degree exponent $\gamma > 3$ we find multiple activations as λ is varied due to independent activation of different regions with different thresholds [17]. One transition happens with a vanishing threshold predicted by QMF theory $\lambda_1 \sim 1/\sqrt{k_{\max}}$ associated to the most connected vertex with degree k_{\max} . A second transition was associated with the mutual activation of hubs [15] at a value λ_2 . Again we tackle this problem by performing averages over different network realizations.

Figures 5.9(a) and (b) compares the QS density against λ for degree exponent $\gamma = 3.5$ with either free (left) or hard (right) cutoffs for networks of different sizes. The average QS density is a double sigmoid, a non-monotonically increasing function of λ , which indicates two activations. At a standard clean N dependent critical point, the logarithmic derivative of the QS density scales as [124]

$$\left. \frac{d \ln \rho}{d \lambda} \right|_{\lambda_c} \sim L^{1/\nu_{\perp}}, \quad (5.6)$$

where L is the system size and ν_{\perp} is a critical exponent associated with the divergence of

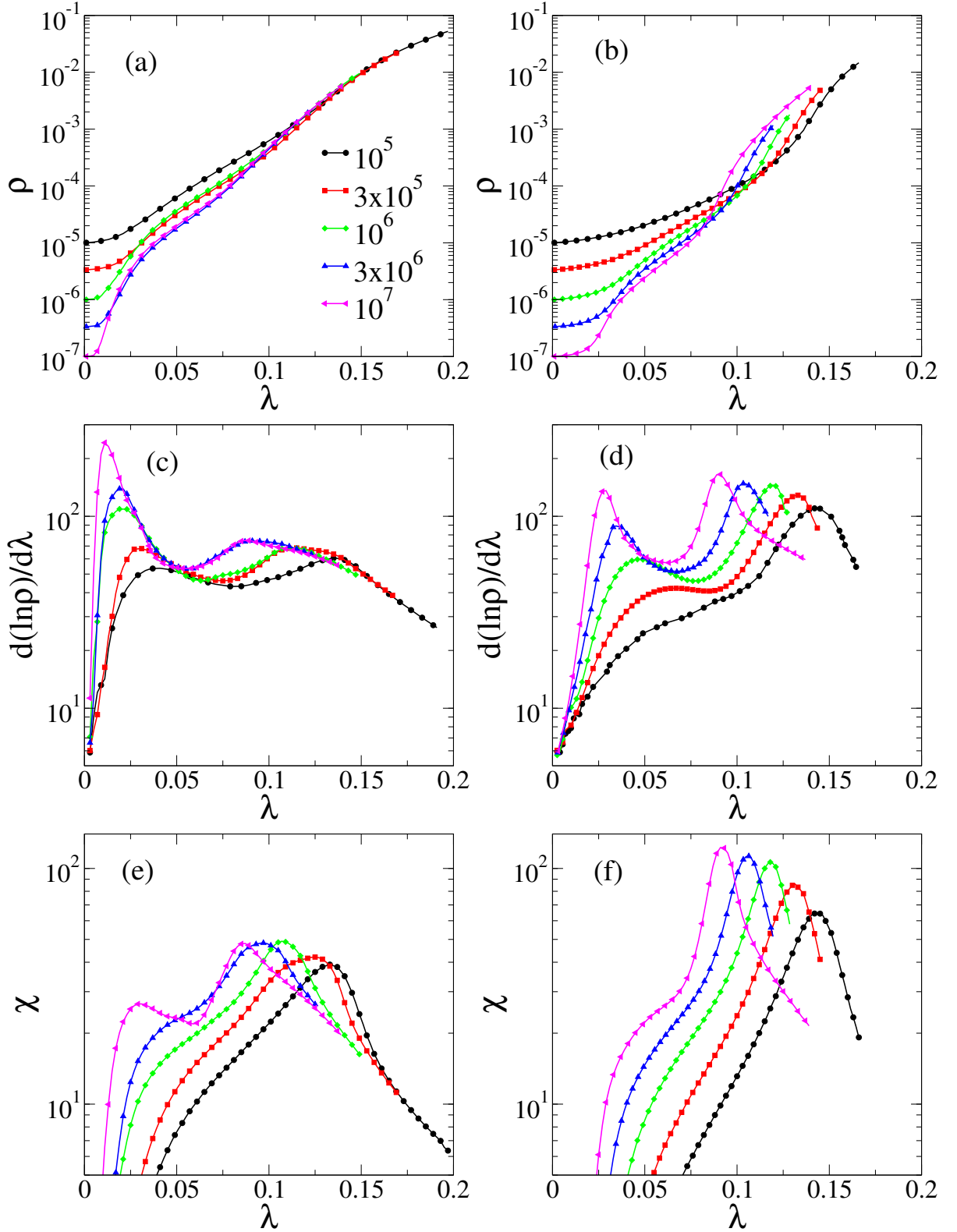


Figure 5.9: Finite size analysis of the QS state for networks with $\gamma = 3.5$. The [(a) and (b)] QS density ρ , [(c) and (d)] log-derivative of ρ , and [(e) and (f)] dynamical susceptibility are shown for free (left) and hard (right) cutoffs, respectively. The number of samples were at least $M = 50$. The network sizes are indicated in the legend.

the correlation length. This log-derivative can also be used to identify multiple transitions in epidemic spreading in networks [17] in association with the dynamical susceptibility χ , defined in Sec. 3.2.4. The latter quantifies the relative fluctuations of the order parameter as shown in Fig. 5.9(e-f).

The log-derivative of the density is shown in Figs. 5.9(c) and (d), while the susceptibility is shown in Figs. 5.9(e) and (f). Averages over the ensemble of networks wipe out the multiple transitions of single networks, leading to two observable transitions at thresholds λ_1 and $\lambda_2 > \lambda_1$. These correspond to the peaks of the log-derivatives and susceptibility; the latter is less evident for λ_1 . Notice, that the double transition identified with the hard cutoff in the log-derivative analysis starts to emerge as a shoulder in the susceptibility curves of the largest size investigated. The threshold at λ_1 can be seen in the susceptibility curves for natural cutoff only for very large sizes and is manifested as a shoulder for the other cases, including hard cutoffs.

For the free cutoff, a QS density $\rho_{\text{free}} \gg 1/N$, the minimum value allowed in a QS simulation, is observed in the interval $\lambda_1 < \lambda < \lambda_2$. This resembles a smeared phase transition [96] and the interval coincides with the region, where Griffiths effects are found in the density decay analysis. We attribute this smearing to the presence or absence of outliers in different samples. In the case of hard cutoff, the suppression of outliers leads to weaker smearing with a density $1/N \ll \rho_{\text{hard}} \ll \rho_{\text{free}}$. In a standard smeared phase transition, patches having high-enough dimensions can exhibit ordering transition independently. In principle outliers, represented by stars, are high-dimensional objects, which could be activated independently. So the basic ideas of smeared transitions are fulfilled. However, outliers plus their neighbors provide a vanishing fraction of the network and give a vanishing contribution to the global density in the thermodynamic limit.

The finite-size analyses at thresholds $\lambda_{1,2}$ are shown in Fig. 5.10. The left threshold, determined via the log-derivative, decays consistently with QMF theory and, for $\gamma > 5/2$, scales as $\lambda_1 \sim 1/\sqrt{k_{\text{max}}} \sim N^{-0.5/(\gamma-1)}$ or $N^{-0.45/(\gamma-1)}$ for natural and hard cutoffs, respectively. This coincides with the activation of the star graph centered at the most connected vertex of the network [11]. The position of the right peak of the susceptibility, which agrees with the right one of log-derivative curves, goes slowly to zero as $\lambda_2(N) \sim N^{-0.09}$ in the investigated interval while the density evaluated at λ_2 follows a power law $\rho_c(\lambda_2) \sim N^{-0.65}$. These scaling laws are, within uncertainties, independent of the cutoff. However, both response functions, log-derivative, and susceptibility, saturate with the size, confirming a smeared transition at λ_2 . The susceptibility

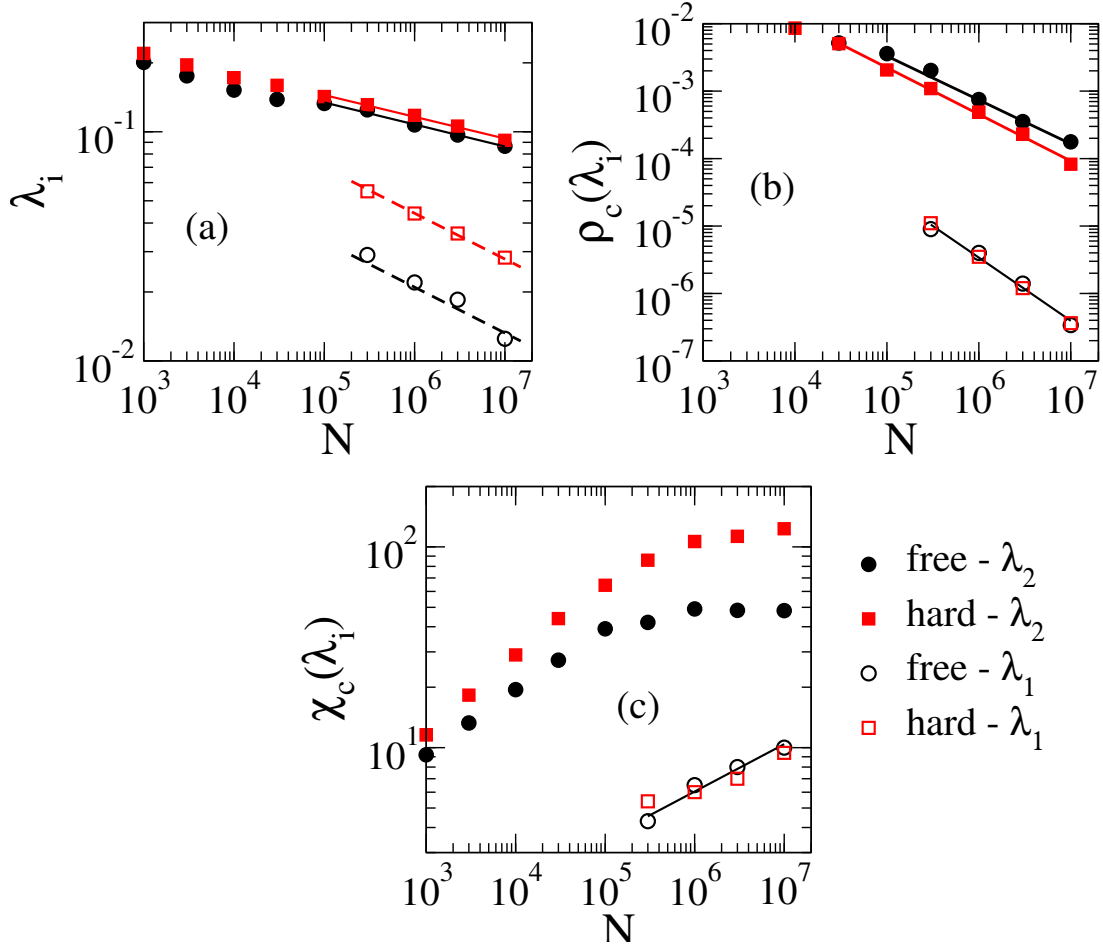


Figure 5.10: Critical quantities against networks size for $\gamma = 3.5$: (a) threshold, (b) critical density, and (c) critical susceptibility at $\lambda = \lambda_i$. Solid lines are PL regressions while the dashed lines are $\lambda_1 \sim N^{-0.2}$ and $N^{-0.18}$ predicted by the QMF theory for free and hard cutoffs, respectively.

at λ_1 , for free cutoff, increases as $\chi_c \sim N^{0.23}$, which is again consistent with the activation of the most connected vertex: For a star with k_{\max} leaves we have approximately $\chi \sim (k_{\max})^{0.55}$ (using data from Ref. [97]) and given that $k_{\max} \sim N^{1/(\gamma-1)}$, we obtain an exponent, which is close to 0.23, observed in the simulations.

The peaks at finite values of λ , observed in single networks [17], are wiped out and are not evident when averages are done. In Ref. [17] networks with up to $N = 10^8$ vertices were simulated while here we analyzed until $N = 10^7$ and performed larger ensemble averaging. So our results do not definitely discard other transitions for higher sizes but no indications of them were observed in the investigated size range.

Analyzing the behavior for $N = 10^7$ with free cutoff and $\gamma = 3.5$, we found $\rho \sim \lambda^\beta$, where $\beta \approx 2.8$ with $\lambda_1 < \lambda < \lambda_2$. Running the dynamics only on the star centered on the most connected vertex by permanently immunizing the rest of the network, we found $\beta_{\text{star}} \approx 2.0 < \beta$.

This means that the mutual activation of hubs is relevant in the interval $\lambda_1 < \lambda < \lambda_2$, leading to an exponent larger than that of a single star centered on the most connected vertex. The estimate $\beta \approx 2.8$ for $\gamma = 3.5$ is inside the rigorous bounds found by Chatterjee and Durrett [9]: $\gamma - 1 < \beta < 2\gamma - 3$.

5.3 Griffiths phases on loosely coupled networks of modules

There is evidence that brain networks, for example, are heterogeneous and contain modular and hierarchical structure, in which the modules are also composed of modular substructures in different levels [45, 125, 126]. This fact inspired Moretti and Muñoz [127] to investigate the activity spreading in hierarchical modular networks of finite dimension, in which Griffiths phases are observed [128, 129], conjecturing that the brain criticality could be explained by the quenched disorder. The empirical organization of biological networks is highly complex and subjective [45] and, therefore, it is not completely clear whether real brain networks in a cellular level are actually hierarchical [130]. Moreover, long-range connections can reduce drastically the network dimension, even consisting of subextensive fractions [131, 132]. Furthermore, modular graphs without hierarchical structure are observed in diverse important systems such as socio-technological [46, 47] or protein interaction networks [48], but the existence of extended critical regions due to the quenched disorder on such systems has not been considered extensively.

To our knowledge, no investigation has been done to scrutinize whether the hierarchy is a necessary condition for the emergence of GPs. This section aims to fill this gap, using simulations of activity spreading models on non-hierarchical modular structures. As seen in the last section, optimal fluctuation theory and simulations provided extended critical regions on heterogeneous networks of finite size constrained to averages over independent network samples. This inspired us to investigate the dynamical behavior of the continuous-time Markovian SIS model on a loosely coupled network of modules. We found extended control parameter regions with non-universal PL decays of activity in time, which are size-independent, calling for the existence of real GPs in infinite-dimensional, but loosely connected modular structures. Thus, our results point out that we can relax the requirement of hierarchical organization and large-world [115, 127] for the existence of GPs on modular networks, although these factors certainly enhance RR effects.

5.3.1 Infinite dimensional and non-hierarchical modular networks

We generated modular networks based on the benchmark model of Lancichinetti, Fortunato, and Radicchi [133]. Consider $g = 1, \dots, M$ modules where the size S_g of each group is drawn according to a distribution $Q(S_g)$. At a vertex level, the degrees are drawn from a distribution $P(k)$ with $k \in [k_0, k_c]$ where k_0 and k_c are lower and upper cutoffs of the degree distribution, respectively. The maximal number of *intermodular* edges connecting vertices of different groups is predefined as $k_g^{[\text{out}]}$ and, in general, can depend on the module g . By construction, this model produces highly modular networks if the number of intermodular connections is much smaller than the *intramodular* one, which was confirmed by the calculation of the modularity coefficient Q_{mod} , defined in Sec. 2.1.4, using the Louvain community detection algorithm [69]. Figure 5.11 shows examples of modular networks with different levels of intermodular connectivity. For $k^{[\text{out}]} \ll M$, we find $Q_{\text{mod}} \approx 1$, confirming the modular structure.

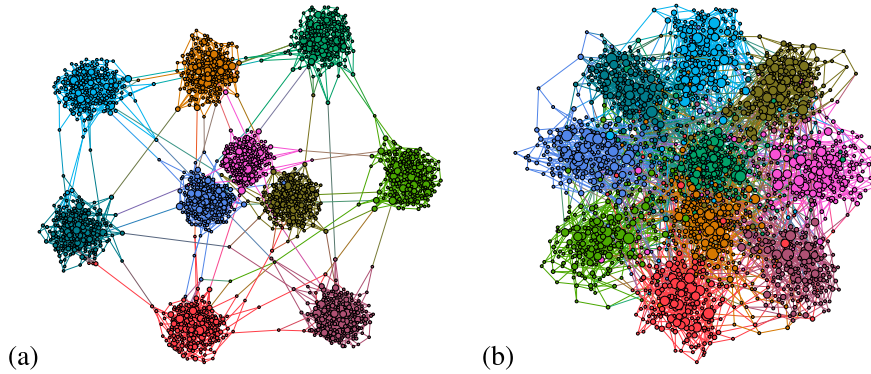


Figure 5.11: Examples of modular structures. Networks with $M = 10$ modules of same size $S = 200$ and number of intermodular connections (a) $k^{[\text{out}]} = 10$ and (b) 100 representing loosely and densely connected modular graphs, respectively. The network degree distribution is given by $P(k) \sim k^{-2.7}$ with $k_0 = 3$ and $k_c = 14$ for the lower and upper bounds cutoffs, respectively. Connected modular structures can clearly be observed. Vertices in a same community are plotted with the same color and their sizes are proportional to the vertex degree. The graph was generated using Gephi visualization tool [134].

The network is generated as follows:

- i) The number of stubs² of each vertex is drawn according to the degree distribution $P(k)$.
- ii) Two stubs are randomly chosen. If they belong to the same group, a new edge is formed. If not, an edge is formed only if the maximal number of intermodular connections in both groups is not exceeded.
- iii) Multiple or self-connections are forbidden.

²Stubs are edges that are not connected yet; see Sec. 2.3.2.

- iv) The process is iterated until all stubs are connected or it becomes impossible to form new edges without multiple or self-connections.
- v) The unconnected stubs are removed. We study only the largest connected component (LCC) which, in the present study, contains almost all vertices of the network.

The number of removed stubs is a tiny fraction (less than 0.02% of the stubs) and does not play any relevant role in the network properties that will be shown in Fig. 5.13 and Table 5.1.

Introducing topological disorder on modular networks

Complex networks are intrinsically disordered by the presence of heterogeneity in the connectivity of its elements. Examples of networks with a topological disorder were shown in Sec. 5.2, in which vertices follow a PL degree distribution with exponent γ . This heterogeneity can also happen in the level of the size of communities given by the distribution $Q(S_g)$. Here, we define two classes of networks to introduce topological disorders:

Monodisperse Modular Networks (MMNs): all modules have the same number of vertices and inter-modular connections. So, $S_g = S$ and $k_g^{[\text{out}]} = k^{[\text{out}]}$.

Polydisperse Modular Networks (PMNs): the modules size distribution is given by $Q(S_g) \sim S_g^{-\phi}$, consistent with empirical networks [133]. The upper bound size is limited by the system size ($S_g \leq N$), while the lower one is chosen such that $\langle S_g \rangle$ has a predetermined value. The intermodular connection $k_g^{[\text{out}]}$ is chosen proportionally to S_g , limiting at $k_g^{[\text{out}]} \geq 2$ to guarantee connectivity [5].

The MMN case is similar to the idea of connecting the different networks used in Sec. 5.2. In all investigated cases, the values $\langle S_g \rangle = 10^3$ and $\langle k_g^{[\text{out}]} \rangle = 5$ were used for comparisons. Since we have two levels of connectivity, one for the modules and the other for the vertices, we also define a network of modules. In this case, we treat modules as vertices, in which modules g and h are connected if one vertex of g is connected to h . In Fig. 5.12 we show one example of the usual network, formed by connections between vertices, and the network of modules obtained.

We determined the average clustering coefficient and the average shortest mean distance [5] for both vertex and module networks. Structural properties of these modular networks are shown in Fig. 5.13. The clustering coefficient, averaged over the whole network, saturates at a small finite value as the network size increases (see Fig. 5.13(a)). This is a natural consequence

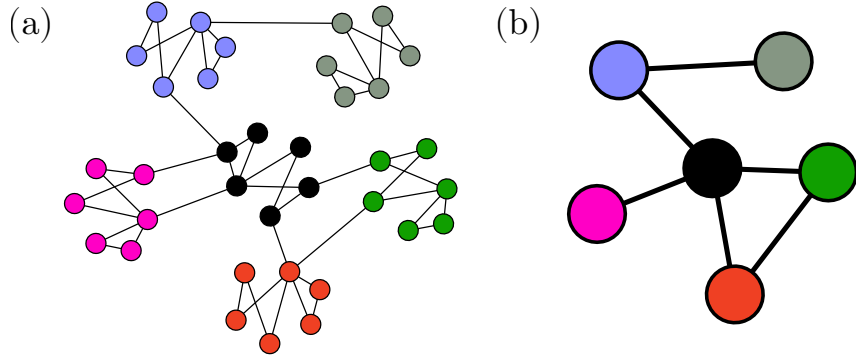


Figure 5.12: Example of (a) network of vertices and (b) network of modules. The network in (a) is transformed in (b) by treating each module as a vertex and edges are formed when an edge exists between vertices of different modules. Note that we consider the network of modules as an unweighted network. Vertices in (a) belonging to the same module have the same color.

of the modular organization of the network that forces vertices to be connected mostly within the modules which are of finite size and thus the probability to form triangles is not negligible. The clustering coefficient of the network of modules vanishes as $\langle C \rangle \sim M^{-1}$ in the cases of $\phi = 4.0$ and MMNs, while it vanishes as $\langle C \rangle \sim M^{-1/2}$ for $\phi = 2.5$. Hierarchically organized networks are clustered with coefficient independent of the size [75]. So, the analysis of Fig. 5.13 shows the lack of hierarchy in the modular networks of our investigation.

For the presented modular networks, the average shortest path increases logarithmically

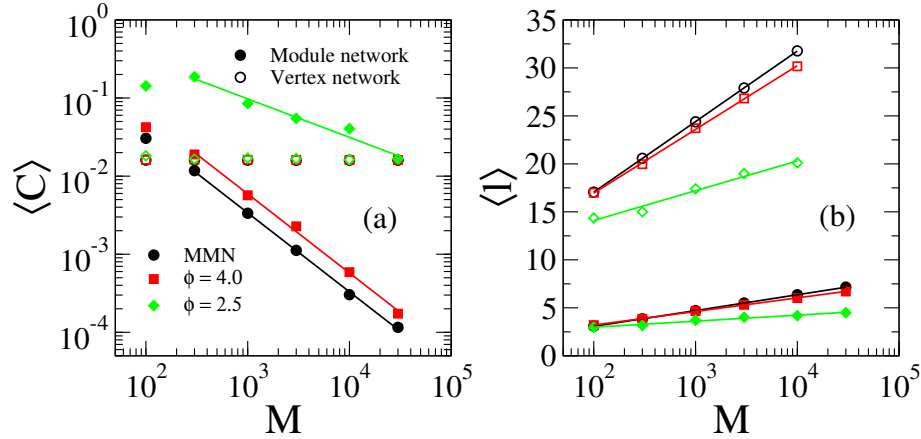


Figure 5.13: Structural properties of modular networks. (a) Clustering coefficient and (b) average shortest path as a functions of the number of modules. Open symbols correspond to the vertex network, while the closed ones represent the network of modules, in which the modules are themselves treated as vertices connected by the intermodule edges. Lines denote (a) power-law or (b) logarithmic regressions. In monodisperse modular networks (MMN), all modules have the same number of vertices $S = \langle S_g \rangle$. Networks with module size distribution $Q(S_g) \sim S_g^{-\phi}$ are obtained by fixing the minimal size S_{\min} such that the chosen average size is obtained. The parameters are $\langle k_g^{\text{out}} \rangle = 5$, $\langle S_g \rangle = 10^3$, $\gamma = 2.7$, $k_0 = 3$, and $k_c = 58$; see main text. The averages were performed over 25 independent networks.

with the size as shown in Fig. 5.13(b) and Table 5.1. So, the investigated networks have infinite dimensions besides the lack of hierarchy.

Table 5.1: Logarithmic regressions for the average shortest distance in modular networks. We analyzed both the original vertex network and the one where modules are considered as vertices connected only by the intermodular edges. Correlation coefficient of the regressions is $r^2 > 0.999$ for MMN and $\phi = 4.0$, while $r^2 = 0.99$ for $\phi = 2.5$. The parameters are $\langle k_g^{\text{out}} \rangle = 5$, $\langle S_g \rangle = 10^3$, $\gamma = 2.7$, $k_0 = 3$, and $k_c = 58$; see main text. The averages were performed over 25 independent networks.

	Vertex networks	Module networks
MMN	$\langle l \rangle = 2.35 + 3.19 \ln(M)$	$\langle l \rangle = -0.22 + 0.71 \ln(M)$
$\phi = 4.0$	$\langle l \rangle = 3.67 + 2.88 \ln(M)$	$\langle l \rangle = 0.37 + 0.61 \ln(M)$
$\phi = 2.5$	$\langle l \rangle = 7.91 + 1.35 \ln(M)$	$\langle l \rangle = 1.73 + 0.272 \ln(M)$

Intrinsic disorder by defining different healing rates

SIS dynamical simulations were performed for networks with topological disorder (MMNs and PMNs) and for networks with intrinsic disorder. In the latter case, we considered $P(k) = \delta_{k4}$ in which all modules are composed by random regular networks (RRNs) [135], while the healing rates μ_i for each vertex i were randomly chosen with equal chance between binary values $1 - \epsilon$ or $1 + \epsilon$, and the infection rates are the same ($\lambda_i = \lambda$). In order to see the finite-size effects, we considered networks of size $N \approx M \langle S_g \rangle$, with $M = 10^3$, 10^4 and 3×10^4 , with $\langle S_g \rangle = 10^3$.

5.3.2 Decay analysis

We show the density decays for a given realization of a MMN for three models of disorder in Fig. 5.14. Similar results were found for the other analyzed network realizations (up to 20). For the topological disorder, a finite-size analysis increasing the number of modules is presented in Fig. 5.14(a). The curves reveal non-universal PLs in the $0.089 \leq \lambda < 0.12$ extended region, which does not change within statistical error margins as the number of modules increases from $M = 10^3$ to 3×10^4 . Thus, contrary to the case of SIS on non-modular PL networks (Sec. 5.2), we see a GP behavior. It is important to mention that the critical regimes hold for intermediate times since the networks are still finite. Furthermore, the analysis provides numerical evidence that the transition point is also size independent. The case of strong intrinsic disorder given by $\epsilon = 0.9$, shown in Fig. 5.14(b), also presents an extended region of critical behavior with non-universal PLs preserved as the sizes are increased. It is worth noting that the SIS dynamics

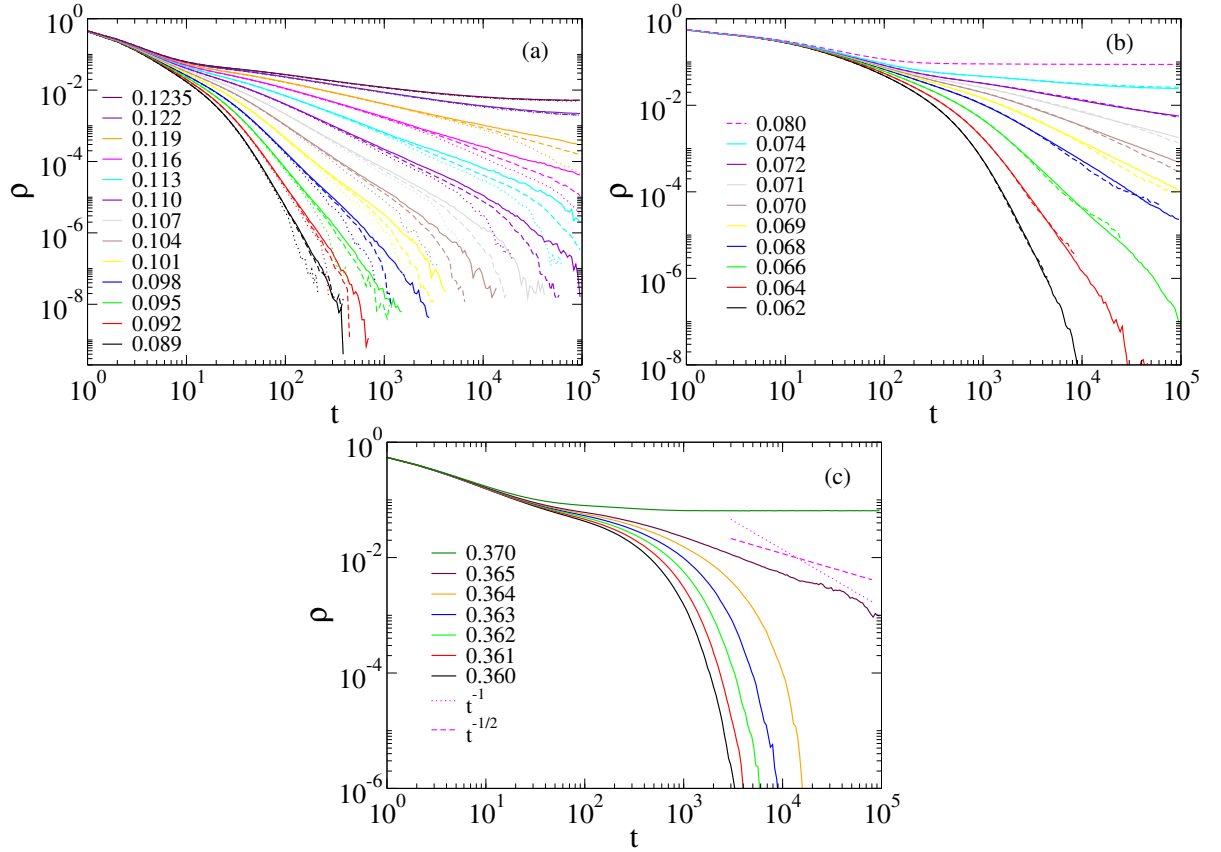


Figure 5.14: Density decay on a single MMN. (a) Decay analysis for SIS with topological disorder introduced by a degree distribution $P(k) \sim k^{-2.7}$ and $k_c = 58$. The numbers of modules are $M = 10^3$ (dotted lines), $M = 10^4$ (dashed lines), and $M = 3 \times 10^4$ (solid lines). (b) Decay analysis for SIS with intrinsic disorder ($\epsilon = 0.9$) on MMNs of sizes $M = 10^3$ (dashed lines) and $M = 10^4$ (solid lines) where the modules are themselves RRNs. (c) SIS decay without intrinsic disorder ($\epsilon = 0$) on a single MMN of $M = 10^3$ modules, each one consisting of a RRN. (a-c) Legends indicate the values of λ and (c) exponents of the dashed and dotted curves.

on MMNs without intrinsic or topological disorder ($\mu_i = 1$), shown in Fig. 5.14(c), does not show GPs and the critical behavior is given by $\rho \sim t^{-1/2}$, instead of a regular mean-field decay [94] $\rho \sim t^{-1}$. This was also found in generalized small-world networks for which the GP shrank to a very narrow region [116]. We also investigated weaker intrinsic disorder using $\epsilon = 0.5$ and observed GPs in several networks realizations, as in Fig. 5.15(a), but in others they were weak or absent, as in Fig. 5.15(b). However, when we performed disorder realization averaging, GPs became evident for both values of ϵ , as seen in Fig. 5.15(c).

We investigated the effects of module size dispersion considering PMNs with topological disorder only. These exhibit the same truncated PL for degree distribution and average sizes of modules as those of MMNs to permit a comparison. In Figure 5.16(a), we show extended regions of λ with PL tails in the density time decays for $\phi = 4.0$, which corresponds to a heterogeneous, but finite variance distribution. These results seem qualitatively similar to those of the MMN

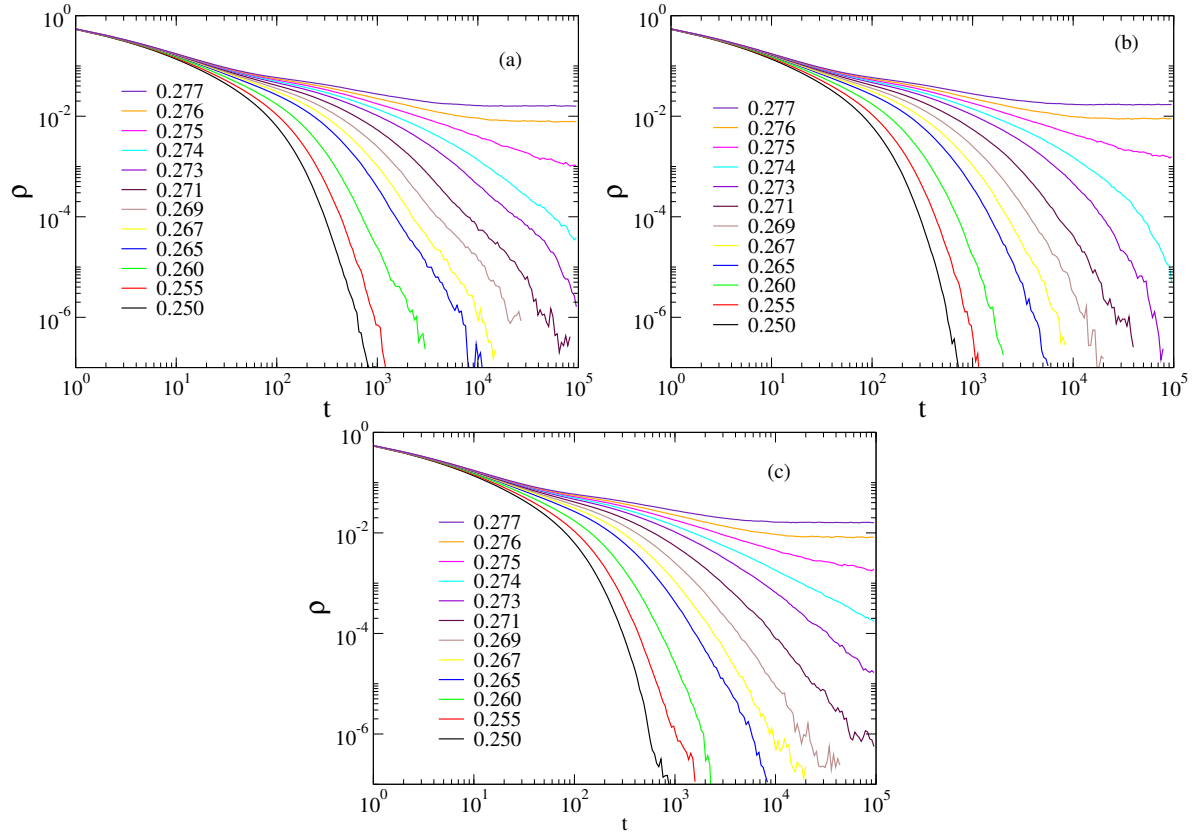


Figure 5.15: SIS decay with intrinsic disorder of strength $\varepsilon = 0.50$ on a single MMN with $M = 10^3$ modules and with uniform degree distribution $P(k) = \delta_{k4}$ (modules are random regular networks). The infection rates are shown in the legends. We show two different disorder realizations, fixed for all dynamical samples, which (a) does or (b) does not exhibit GP. (c) When a new disorder realization is generated for each dynamical sample, the average leads to GP on extended scaling law regions in the interval $0.265 \lesssim \lambda \lesssim 0.275$. 100 to 500 dynamical runs were used to compute averages.

case. Finite-size effects are stronger, but a GP occurs in the interval $0.095 < \lambda < 0.115$, which is narrower than in the monodisperse case. Noticeably, GPs are not observed for the scale-free case with $\phi = 2.5$ shown in Fig. 5.16(b), in which modules of essentially every size may appear. A finite variance of $Q(S_g)$ reduces the RR effects in comparison with MMNs, since some large modules have many intermodular connections $k^{[\text{out}]} \gg \langle k^{[\text{out}]} \rangle$ reducing their independence. For an infinite variance, the situation becomes drastic. A single module can contain a considerable large fraction of the whole network and alone rules the critical dynamics of the system becoming equivalent to the non-modular case [43, 120, 121]. Once we have established under which conditions of module size variability the GPs are robust, in the rest of the section we consider only the case of MMNs, stressing that the central conclusions are the same as in the case of a finite variance in the module size distribution.

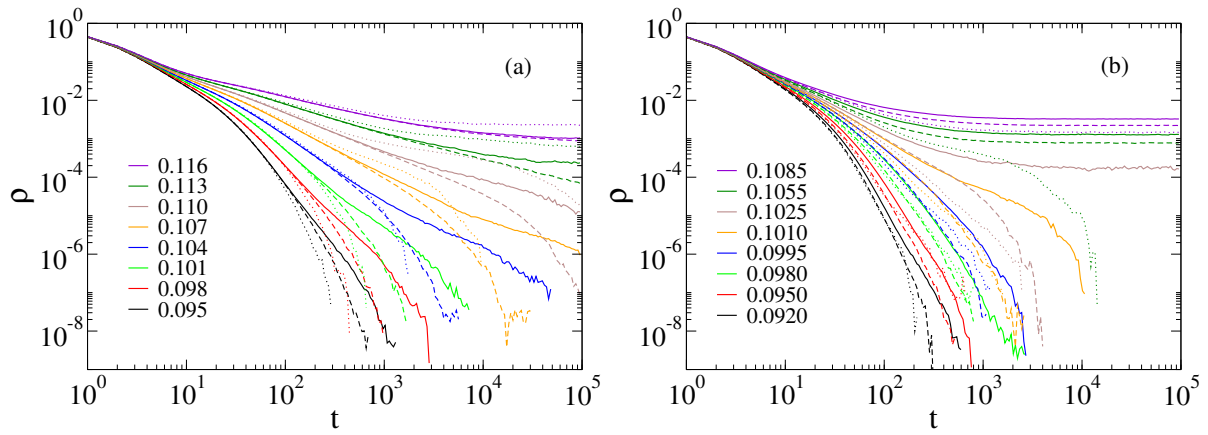


Figure 5.16: Density decay for SIS with topological disorder on a single PMN. The module size distributions with exponents (a) $\phi = 4.0$ and (b) $\phi = 2.5$ are shown. The finite size analysis is done using $M = 10^3$ (dotted lines), 10^4 (dashed lines), and 3×10^4 (solid lines). Other network parameters are given in text and the values of λ indicated in the legends.

5.3.3 Spreading analysis

Fig. 5.17(a) shows the number of active vertices as a function of time in spreading simulations on a MMN. For regular dynamical criticality, this quantity is expected to evolve as $N_a(t) \propto t^\eta$. One can see non-universal PL tails in a range similar to the one found in the density decays, including a similar exponential cutoff for long times due to the finite size of the networks.. The survival probability curves, $P_s(t)$, defined in Secs. 3.2.4 and 3.3.3, exhibit a very similar behavior [Fig. 5.17(b)] with the same exponents as those of the $\rho(t)$ decays at a given λ within the critical region, expressing that the rapidity reversal symmetry [94, 136] is unbroken by the quenched disorder. This symmetry implies, for example, that the asymptotic probability ($t \rightarrow \infty$) to find one infected vertex at a randomly chosen location is weakly dependent on the initial condition or, more precisely, $\rho(t) \propto P_s(t)$.

Due to the extended interval with PLs and the corrections, it is hard to estimate the transition point location and the time decay functional form accurately. Simple PL fitting results in $P_s(t) \propto t^{-\delta}$ with $\delta = 0.42(1)$ at $\lambda_c \simeq 0.12$. Assuming a scaling in the form $P_s(t) \sim \ln(t/t_0)^{-\tilde{\delta}}$, as in case of the absorbing-state phase transition with strong disorder in lower dimensions [137], we could obtain $\tilde{\delta} \approx 5$. Neither of these is in agreement with the regular mean-field behavior obtained for absorbing-state phase transitions with quenched disorder in high dimensions [138]. We applied an alternative method [116], which assumes that the leading correction to the scaling comes from the same scale t_0 in the critical behaviors of $P_s(t)$ and $N_a(t)$. Plotting $\ln[N_a(t)]$ against $\ln[P_s(t)]$, transition point curves must fit on a straight line. As Fig. 5.17(c) shows, this allows an

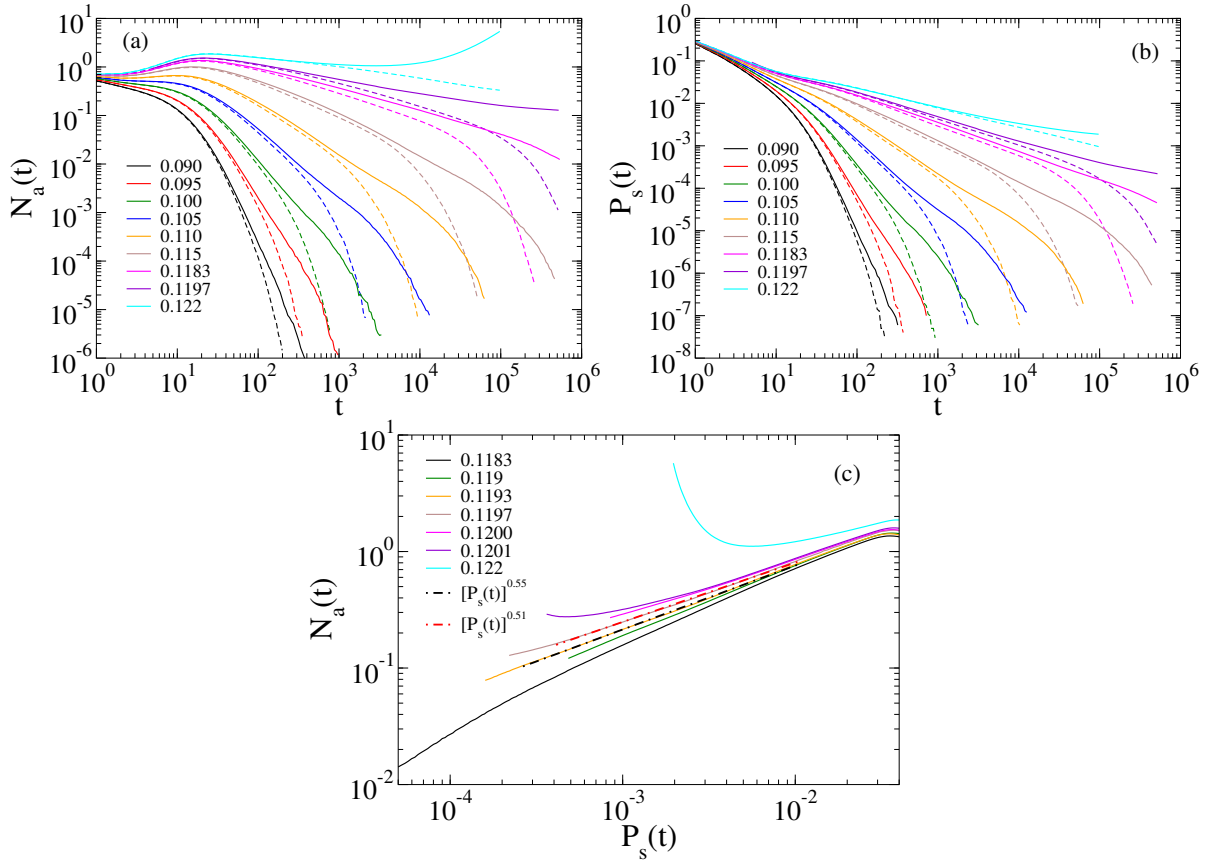


Figure 5.17: Spreading analysis for SIS on a single MMN. Only topological disorder is introduced as a truncated PL $P(k) \sim k^{-2.7}$ with $k_c = 58$. Finite size analyses of the (a) number of active vertices and (b) survival probability are done with $M = 10^3$ (dashed lines) and 10^4 (solid lines) modules. (c) Determination of the transition point in a double logarithmic plot of $N_a(t)$ vs $P_s(t)$ for $M = 10^4$ modules. Infection rates λ are indicated in the legends.

estimate for the transition point $\lambda_c = 0.1195(2)$, in which the slope is $0.53(2)$.

We also determined the avalanche size distributions $P_{\text{ava}}(s)$ in spreading simulations. The size of an avalanche is defined as the total number of vertices s activated during a spreading experiment. The results for $M = 10^3$ can be seen in Fig. 5.18. Power-law behavior occurs for the $10 < s < 10^6$ region with a variation of the exponent as a function of λ . A PL fitting for the $0.1 \leq \lambda \leq 0.1215$ region results in $P_{\text{ava}}(s) \propto s^{-\tau}$ with $1.20 \leq \tau \leq 1.52$, which encloses mean-field exponent of the directed percolation class ($\tau = 3/2$) [139]. Curiously, this mean-value is consistent with reports for activity avalanches observed in the brain [112, 140, 141]. However, it is important to remember that other mechanisms can explain the exponent [38] $\tau = 3/2$. On the other hand, since the spreading and decay exponents depend on λ , the same should happen for τ , as the consequence of the scaling relation $\tau = (1 + \eta + 2\delta)/(1 + \eta + \delta)$ for absorbing-state phase transitions [139]. Indeed, the $3/2$ exponent also appears in the avalanche mean-field exponents of many models and several universality classes [136]. However, in our case, the universality

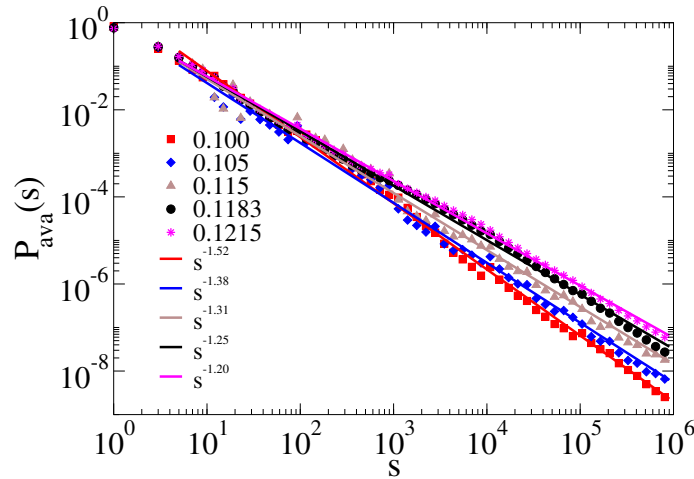


Figure 5.18: Avalanche-size distributions for SIS spreading in a single MMN with a topological disorder and $M = 10^3$ modules. Different values of λ are indicated in the legends, and simple PL tail fits are also shown.

class seems to be different, since the $3/2$ exponent is found within the GP, while at the critical point $\lambda_c \approx 0.12$ the measured τ is smaller.

At a first glance, the results presented up to this point are at odds with the conjecture that infinite-dimensional networks cannot sustain real GPs [115]. Strictly speaking, one may argue that the finite module sizes in the monodisperse case imply that RR lifespans can be huge, but bounded and thus the observed PLs correspond to very strong Griffiths effects, differing from real GPs in the sense that they disappear in the thermodynamic limit. However, we also found size-independent GPs in the polydisperse case with $\phi = 4.0$, where this size restriction does not apply. To understand this, we express the typical S_c for which at least one module of size $S > S_c$ is present as

$$M \int_{S_c}^{\infty} Q(S) dS \sim 1 \Rightarrow S_c \sim M^{1/(\phi-1)}, \quad (5.7)$$

implying that the sizes of the largest modules diverge as their number is increased. The same result can be deduced more rigorously using extreme-value theory [70].

In general, the critical dynamical regions found are robust as N is increased and have similar sizes to the ones found in hierarchical networks [127–129]. Besides, for MMNs the scaling regime improves as the network size is increased, as seen in the slope analysis in Fig. 5.19(a). Lastly, it is important to stress that the results were found for single networks realizations, without averages as in Sec. 5.2 for independent networks.

The mechanism of GPs formation can be explained by the SIS model nature and the weak connections between modules. In the case of topological disorder, the SIS activity is concentrated

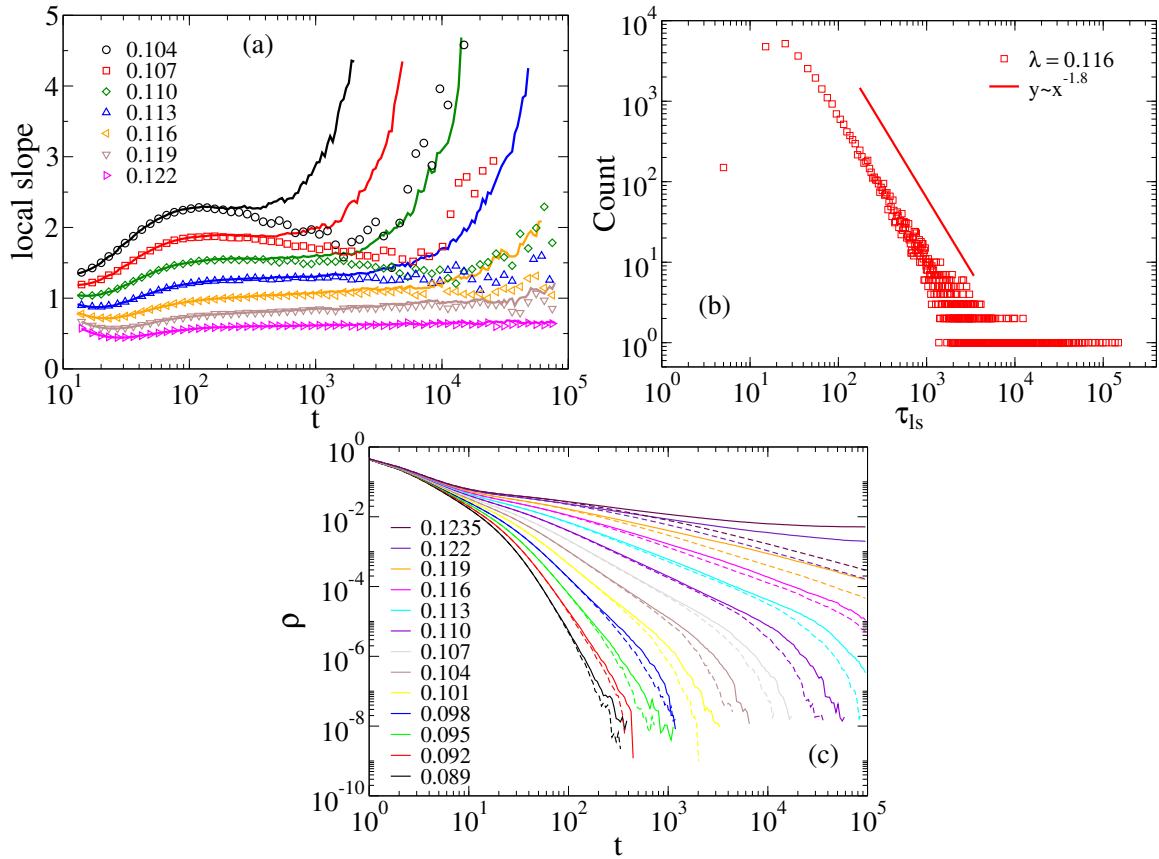


Figure 5.19: (a) Local exponent analysis for activity decay in MMNs with $M = 10^4$ (lines) and 3×10^4 (symbols) modules. The local exponent is defined as $\alpha_{\text{eff}} = -d(\ln \rho) / d(\ln t)$. After an approximately constant interval, the exponents become large, indicating an exponential decay. As the number of modules increases, exponential tails show up for larger times. (b) Lifespan distribution for individual modules in a SIS dynamics in a single MMN with $M = 3 \times 10^4$ and $\lambda = 0.116$, inside a GP. This distribution is evaluated by using bins of size $\Delta\tau_{ls} = 10$. (c) Comparison between SIS decays for modules of a single MMN interconnected (solid lines, $k_g^{\text{[out]}} = 5$) and disconnected (dotted lines, $k_g^{\text{[out]}} = 0$) with $M = 10^4$ modules. The parameters are the same used previously. Averages were done for 100 to 500 dynamical samples, depending on λ , whose values are indicated in the legends. The interconnected case has larger activity densities due to the long-range activation of the SIS model [9, 15, 16].

in localized domains, but sub-extensive, as around the hubs or maximum k -cores, as discussed in Chapter 4. However, connections between modules are random, resulting in a small probability of connecting highly active regions. So, randomness results in variability of the modules, producing lifespan distributions with high heterogeneity, as shown in Fig. 5.19(b), following a power-law $P(\tau_{ls}) \sim \tau_{ls}^{-a}$, with $a \approx 1.8$. This analysis was made by computing exponential decays as a function of time for each module, with $\rho \sim e^{-t/\tau_{ls}}$ for $t > 20$.

Active vertices in a large fraction of these modules are therefore short-lived and, from the point of view of the spreading process, behave as if they were removed. The remaining network that sustains the long-term activity can be approximated by isolated or weakly coupled patches,

providing an effective zero-dimensional substrate for the activity spreading. In Fig. 5.19(c) we compare simulations performed on a MMN with interconnected, and disconnected with $M = 10^4$ modules. The latter is the same kind of analysis performed in Sec. 5.2. One can see that GPs are even stronger in the former case, showing that connections between modules are important and modify decay profiles.

5.4 Summary and discussions

Random, scale-free networks exhibit strong, quenched inhomogeneities, and therefore rare region (RR) effects can be expected to play an important role. To see RRs of arbitrary sizes we should simulate arbitrarily large system sizes or performing sample averages over many independent network realizations. The latter way is not equivalent to the former one in scale-free networks. In models defined on networks with infinite topological dimensions, a recent hypothesis states that Griffiths phases cannot exist [115] and another important result for infinite-dimensional networks with power-law degree distributions is that SIS does not exhibit a phase transition at finite λ [9]. Real networks, on the other hand, can be very large but are always finite. Therefore, numerical analysis of different sizes is of great importance.

Here we presented extensive simulations on networks generated with the configuration model using free (fluctuating) and structural or hard (non-fluctuating) degree cutoffs. We focused on statistics over a large ensemble of networks. Contrary to the results obtained on single network realization, where multiple activations were observed [17], we observe that the network ensemble averaging exhibits Griffiths effects in an extended region of the control parameter $\lambda_1(N) < \lambda < \lambda_2(N)$, which diminishes as network size increases and disappears in the thermodynamical limit. These Griffiths effects are due to sample-to-sample fluctuations, producing non-self-averaging within the shrinking critical dynamical region, rather than the existence of RRs of actual Griffiths phases. We also observe the occurrence of a smeared transition, with saturated fluctuations of the order parameter at λ_2 . Our findings can be relevant if we consider independent realizations as graphs occurring in a sequence of uncorrelated, time-dependent networks at a given time and we measure quantities in the long-time average. Alternatively, such results can describe the behavior of systems in which power-law degree distribution in modules make up a very weakly coupled network.

Indeed, we have analyzed the activity spreading of the SIS model in loosely connected,

non-hierarchical modular networks and observed extended regions of non-universal scaling behavior for topological [Figs. 5.14(a), 5.16(a), and 5.17] and intrinsic [Figs. 5.14(b) and 5.15(a,c)] disorders, using density decay and spreading analyses. The interval of the control parameter with a dynamical critical region, which was robust under finite size analysis, showed size similar to previous studies on hierarchical networks [127–129].

To summarize, our analysis reveals the existence of stable GPs on small-world, thus infinite-dimensional substrates, conditioned to be sparsely connected in a modular structure, as an alternative for the origin of criticality on modular systems. The hierarchical modular networks, where GPs were previously observed [127, 128] are loosely connected. Hierarchy plays an important role, by increasing the distances between the modules, thus enhancing GPs, but it is not a necessary condition. The brain criticality hypothesis via GPs raised by Moretti and Muñoz [127] is strengthened by our results since building connectome networks [45, 130, 142] is far from being trivial and the hypothesis of hierarchy in the modular organization of the brain is not fully accepted. In particular, the very restrictive condition of finite dimension is fragile due to the presence of long-range connections. The model we used, conceived to be simple, allowed us to address this specific issue which would not easily be accessed in a real brain network. We expect that our results, which were not conceived for a specific system, will be important for the investigation of criticality in other modular systems beyond brain networks.

Part III

Models and applications on directed social and metapopulation networks

Chapter 6

Quantifying echo chamber effects over political communication networks

Related publication:

Quantifying echo chamber effects in information spreading over political communication networks [49]

Wesley Cota, Silvio C. Ferreira, Romualdo Pastor-Satorras, Michele Starnini

EPJ Data Science 8, 35 (2019)

Social networks in the internet form a substrate for information propagation with unprecedented capabilities, in which users are at the same time producers and consumers of content, allowing an almost instantaneous and unmediated propagation, as in Twitter¹ and Facebook². Their emergence in the last few years changed the way people behave, decide, opine, and make choices [143]. These networks are several orders of magnitude larger than those traditionally available in the social sciences [144], defining the field of computational social science [145]. In this chapter, we reconstruct a political communication network gathering data from Twitter focusing on discussions about the impeachment process of the former Brazilian President, Dilma Rousseff. Protests and manifestations both in favor or against Rousseff left important digital fingerprints in social networks, involving a large set of individuals in a period of almost one year. We investigate how the political position of users affect information diffusion across the network.

6.1 Online social networks

Data containing the content and interactions on Twitter are available with a very detailed level of information. Each message or *tweet* sent on Twitter has details such as the users involved

¹Online social network available at <https://twitter.com/>.

²Online social network available at <https://facebook.com/>.

(author and users mentioned), geographical information, and the exact moment in which it was posted with a precision of milliseconds. Different from the static networks studied in the last chapters, we consider explicitly the time dimension in these Twitter networks, investigating the effects of the temporal dimension on the information spreading. Hashtags, which are words introduced by a hash symbol #, are used to define topics or even sentiments about a subject. In the impeachment process of Dilma Rousseff, for example, users used hashtags in favor, against, or related to the discussion. By analyzing the content of tweets containing hashtags, we can classify the interactions between users expressing different sentiments and quantify the so-called echo chamber effects.

6.1.1 Echo chambers in social networks

Online communications networks are marked by their degree of homophily, the property that individuals prefer to interact with others similar to them, sharing views and opinions [146–148]. This leads to a natural polarization in groups or communities with different perspectives [149, 150]. Mutual interactions of these users can create echo chambers [50], which are composed of users reinforcing their beliefs discussing with each other. With the large availability of data on online social networks, it has been possible to analyze social polarization and echo chambers in big datasets in different contexts, such as US and French elections [151], secular vs. Islamist discussions during the 2011 Egyptian revolution [152, 153], or the 15M movement in Spain during 2011 [154].

Echo chambers have been shown to percolate to the offline realm [155], to be related to the spreading of misinformation [156, 157], or the development of ideological radicalism [158]. Recent studies, however, have challenged the impact of echo chambers and partisan segregation in communication networks over online social media [159, 160]. In this work, we collect data from Twitter to study echo chamber effects in discussions about politics in Brazil.

6.1.2 Obtaining data from Twitter

The Twitter social network allows anyone with a developer account to collect data via an Application Programming Interface (API) in real-time filtering by keywords (hashtags or not), users, or geographical location [161]. With this, a large amount of data can be extracted to reconstruct large communication networks to quantify polarization and echo chambers. Users can send and receive texts in form of tweets, containing at most 280 characters. All tweets are

updated in real-time, without the need for approval, on the authors' pages. Special characters define interactions, mentioning users by prepending @ to their usernames, or a topic tagging by using a word starting with #, defining a hashtag. Figure 6.1 shows a hypothetical example of a tweet created by a user `user1` mentioning `user2` and using a hashtag `#CésarLattes`. Every time a user is mentioned in a tweet, she/he is warned by the Twitter platform. Hyperlinks are created for each mentioned user and used hashtags, allowing individuals to see the personal pages of others and a list of tweets containing the same hashtag in reverse chronological order. Some of these hashtags can become "trending topics", meaning that many users are using them at the same time interval.



Figure 6.1: An individual with username `user1` mentions another user `user2`, and tag her/his tweet with a hashtag `#CésarLattes`, highlighted in blue as hyperlink. Users can follow others by clicking in the "Follow" button. In this example, the Twitter API returns the tweet's content as `@user2 "Everything I did is not worth a Mozart symphony." #CésarLattes`.

Users can follow other users on the platform, receiving their tweets in a single web page with the latest tweets, even when they are not mentioned by others. Each tweet, whether self-authored or not, can be retweeted, replicating the message to her/his followers. In this case, the API platform returns a text containing `RT @` followed by the username of the author of the original tweet. Note that a retweet can be considered as a kind of implicit mention. It is important to mention that Twitter has recently implemented new interactions like quotation, that were not considered in this study.

Individuals registered on Twitter have only two privacy choices: to keep their profile public, meaning that anyone, registered or not, can see her/his tweets; or having a private profile, allowing only her/his followers to see their tweets. As of 2019, Twitter stated it had 330 million monthly active users, and more than 85% of American users keep their profile public [162]. Therefore, public tweets can be collected on large scale without violating privacy policies. In this work, we collected tweets published between March and December 2016 by using keywords related to the impeachment process of Dilma Rousseff [163]. Table 6.1 shows the 323 keywords

Table 6.1: List of the 323 keywords used to collect tweets related to the impeachment process of Dilma Rousseff between March and December 2016.

13marbrasilnasruas	13marco	13marco2016brasilnasruas	13marcobrasilnasruas	13marcoeunaovou	13marcoeuvou
13marcoouvamosouelevolta	13mareuvou	16ago	16agosto	16deago	16deagoevou
16deagosto	17abrilpovonasruas	17deabril	18marco	18mareuvou	31jul
31juleuvou	31julho	31julhoavantebrasil	31julhoconfirmado	31julhoevou	31julhopelobrasil
31julvamos	31mar	31mareuvou	acaboudilma	acordabril	adeusquerida
aecio	aeciogolpista	aeciomedroso	aecionacadeia	aquemptenceaescola	autorizaplanejamento
avantetemer	bandidoviraministro	bhnasruas	bolsomito	bolsonaro	bolsonaro2018
boratemer	brasilapoialavajato	brasilapoiatemer	brasilcontraogolpe	brasilcontrastf	brasilianasruas
brasilnasruas	brasilpaisdeladroses	brasilsemp	brazilnocorrupt	cadeia2ainstancia	caixa 2
caixa2	camara	camarasemp	censuranuncamais	cinogolpista	constituicao
contraogolpeedia18	contrapec	contrapec241	corrupcao	coxinha	culturapelademocracia
cunhagolpista	cunhanacadeia	democracia	democracaja	deputados	derrubargolpenasruas
desejoprotemer	desligaogolpe	desligatv	dia13mareuvou	dia16	dia17abril
dia17impeachment	dia18.03	dia18.e.nossavez	dia18.nossavez	dia18.nossavez	dia31juleuvou
dia31vaisermayor	dilma	dilmabandida	dilmacaixa2	dilmacaradarenuncia	dilmaculpada
dilmafeiaobrasilteodeia	dilmafica	dilmaguerreira100	dilmais	dilmajaera	dilmanaomerepresenta
dilmanovamente	dilmanuncamais	dilmare	diretasja	diretasja2018	ditaduratemer
eduardocunhagolpe	eduardocunhagolpista	eleicoesgerais	esquentagrevegeral	estamos todos com lula	eugritomoro
euquerodilmapresa	euquerolulapreso	felizaniiversariomoro	ficadilma	ficalula	ficaquerida
ficatemer	fimdopt	forabandidos	foracomunismo	foracoxinhas	foracunha
foradilma	foragolpistas	foraladiao	foralula	forapt	foraserra
forastf	foratemer	foratemerolimpico	foratemerrio2016	fueratemer	globogolpista
golpe	golpequinaopassa	golpeday	golpenao	golpenuncamais	golpista
golpistasday	grevedia29	grevegeral	impeachment	impeachmentday	impeachmentdilma
impeachmentja	jantardotemer	jantartemer	jucagolpista	lavajato	lewandowskipetralha
libertemzedirceu	ligacao dilma	ligacao lula	lula	lula2018	lulaacabou
lulacasacivil	lulacovarde	luladenunciado	lulaestamoscomvoce	lulaestamoscontigo	lulaeuconfio
lulaeufendo	lulaeufespeito	lulafica	lulagolpista	lulaisworththefight	lulala
lulaladenovo	lulalidermundial	lulalixomundial	lulaministro	lulaministroja	lulanacadeia
lulanacadeiaja	lulanapapuda	lulanuncamais	lulapajaula	lulaperseguidopolitico	lulapersiste
lulapresidente	lulapreso	lularesiste	lulareu	lulavalealuta	lulavergonhanacional
lulavolta	lutarsempre	lutepelas10medidas	lutodilma	lutopelademocracia	lutopelobrasil
lutopt	lutosempre	mandato	marchadascoxinhas	marchadoscortuados	marchadoscortuados
mastenhoconviccao	mblogolpista	mexeucomlulameucomigo	micheltemer	mobilizacaototal	moralistassem moral
moropresidente	mortadeladay	mudabril	naoagolpe	naovaitergolpe	naovoupraria
nenhumdireitoamenos	novaeleicao	novaseleicoes	obrigadomf	ocupab	ocupabril
ocupadilma	ocupabrazil	ocupacopacabana	ocupapoliada	ocupapaulista	ocupario
ocuparj	ocupasaopaulo	ocupasp	ocupatudo	ocupatudocontraogolpe	ouvaioelafica
ouvaioelevolta	ouvamosouelafica	ouvamosouelevolta	ouvocevaioelafica	panelaco	passadilma
pec 241	pec 55	pec215	pec241	pec55	pecdamorte
pecdofindomundo	pelademocracia	petrobras	pl2431.11	planalto	pmdbgolpista
povocomlula	presaledopovo	psdb	psdbteupassadotecondena	pt	ptacabou
ptdesmoronando	ptexit	quedado planalto	quedaplanalto	queremosdilmare	renangolpista
renantemealavajato	renunciadadilma	renunciatemer	renunciadilma	respeiteasurnas	rippt
rjnasruas	saotodosgolpistas	senado	senadores	sessaodoi impeachment	simpelo impeachment
somostodosgolpistas	somostodoslula	somostodosmoro	somostodospt	soscoupinbrazil	soupt
souptpq	souptsoudilma	souptsoulula	spnasruas	standwithlula	stf
stopcoupinbrazil	tchaudilmavez	tchauquerida	tchauqueridaday	teimadilma	temer
temereglobounosnogolpe	temergolpista	temergolpistafranco	temerjamais	temermelhorquept	thauquerido
toconilma	toconilula	todoscomilma	todoscomlula	todosnasruas31julho	todosruadial3
vaadilma	vaiterforatemersim	vaiterimpeachment	vaiterlula	vaiterlulasim	vaiterluta
vaitervia	vamostirarbrasildovermelho	vazatemer	vemprademocracia	vempraria	vempraria3mar
vempraria17abril	vempraria18mar	vempraria31jul	vempraria31julho	vemprariaubrasil	voltadilma
voltadilmapresidenta	voltalula	voltaquerida	vomitacojantardotemer	votacao impeachment	

used to collect tweets. A Python library called *Twython* [164], version 3.4.0, was used to make the connection to the Twitter API and all usernames were anonymized before any analysis.

During the data collection, protests, and important events occurred, with messages exchanged on Twitter about these issues. Table 6.2 and Fig. 6.2 show some of these events and their activity on Twitter, respectively. All the data were filtered by using a procedure that will be discussed in the next sections. A total of 48 212 722 tweets were collected, and 12 322 322 of them contained at least one hashtag. The number of tweets with hashtags collected for each day is shown in Fig. 6.2. As expected, this activity is not constant over time, varying every day,

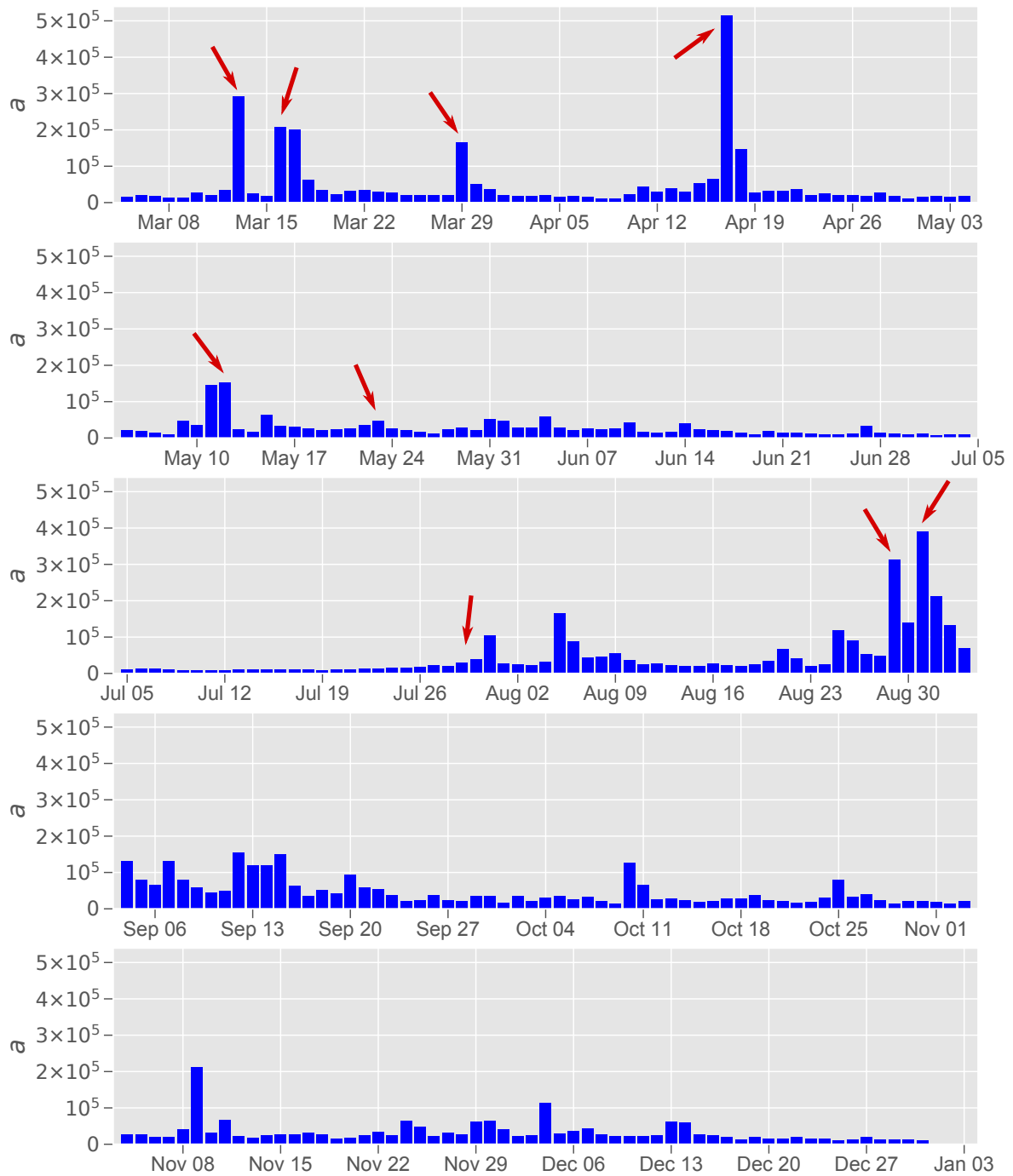


Figure 6.2: Activity of tweets with hashtags collected as a function of the day. High activity can be observed around some events, reported in Table 6.2, which are indicated by arrows. The high activity in November 9th coincide with Trump's victory in USA election which is, in principle, not related to the process we are investigating. This peak disappears when we consider only the strongly connected component of the communication network; see Sec. 6.2.2.

with peaks associated with important events described in Table 6.2. The maximum number of tweets containing hashtags collected on the same day was more than 500 thousand on April 17, 2016, when the impeachment process was approved.

Table 6.2: Some important dates and events during the impeachment process of President Dilma Rousseff, indicated by arrows in Fig. 6.2. The sentiment of the majority of interactions (belonging to the strongly connected component of the political communication network, see Sec. 6.2) collected on that day is shown in the rightmost column.

Date	Event	Activity
Sun Mar 13	Biggest street manifestation against the government spread out in more than 250 cities.	-1
Wed Mar 16	Supreme court permits the constitution of a commission on the chamber of deputies	+1
Tue Mar 29	MDB (Brazilian political party "Movimento Democrático Brasileiro") left the government	+1
Sun Apr 17	Deputy chamber approves impeachment with 367 votes against 137	+1
Thu May 12	Rousseff leaves the presidency after Senate approval	-1
Mon May 23	Audio of Senator Romero Jucá saying " <i>Estancar a sangria</i> "	+1
Fri Jul 29	Rousseff delivers final arguments in the Deputy chamber	-1
Mon Aug 29	Rousseff's defense in Senate	+1
Wed Aug 31	Senate approves impeachment with 62 to 20 votes	+1

6.2 Reconstruction of political communication networks on Twitter

After the data collection, a filtering procedure was developed to reconstruct the network of interactions of users discussing the impeachment process on Twitter. In this section, we describe all these procedures, starting with the classification of hashtags by different sentiments, then filtering out tweets not relevant for the process, and in the end, a network based procedure to ensure paths of information between any pair of users.

6.2.1 Classification of hashtags

Hashtags can be used to define the political position of users [165]. In this work, we assign to each hashtag t one of four sentiments s_t :

- i) not related to the impeachment process ($s_t = \mathbf{x}$),
- ii) pro-impeachment ($s_t = -1$),
- iii) anti-impeachment ($s_t = +1$), or
- iv) neutral and/or related ($s_t = 0$).

The last one includes tweets whose sentiments are not clearly polarized but are related to or represent both political positions. We performed a manual classification of tweets containing these hashtags [150, 153, 166, 167]. Considering a list with 495 most frequent hashtags in our dataset, four volunteers made the categorization. An interactive page was created, available

at <http://labs.wesleycota.com/twitter>, containing all these hashtags and the four classifications. The following question was presented: *“How do you think that these hashtags were used in tweets related to the process of the impeachment of the president Dilma Rousseff along the year of 2016?”*. The volunteers were asked to read tweets containing these hashtags before choosing their classification for each hashtag. A screenshot of the interactive page is presented in Fig. 6.3.

Como você acha que essas hashtags devem ser usadas em um tweet relacionado ao processo de Impeachment da Dilma em 2016?

O contexto é apenas os eventos ocorridos durante o ano de 2016. Caso a hashtag não seja relevante ao processo, coloque em "fora de contexto". No final da pesquisa, digite seu nome (apenas o primeiro é suficiente) e clique em Salvar.

Basta arrastar a hashtag para o devido bloco. Não funciona em celular! Testado no Firefox e Chrome.

Caso não tenha tempo e já tenha classificado algumas hashtags, clique em salvar e **os dados poderão ser recuperados posteriormente** a partir do código que será mostrado. Automaticamente, caso NÃO limpe o cache do navegador, poderá clicar em "Recuperar dados já salvos" para continuar. Caso não funcione, entre em contato comigo.

Se tiver alguma dúvida sobre determinada hashtag, digite-a no campo seguinte e tecle enter para ver alguns tweets relacionados:

Seu nome:

[Fazer nova separação \(do zero\)](#)

Não classificada	Contra a Dilma	A favor da Dilma	Neutro ou ambos	FORA DE CONTEXTO
<input type="text" value="moroamigodecunha"/>	<input type="text" value="liberadelacaojanot"/>	<input type="text" value="forcadilma"/>	<input type="text" value="forastf"/>	<input type="text" value="pt"/>
<input type="text" value="acordabrasil"/>	<input type="text" value="quedadoplanalto"/>	<input type="text" value="opovoquerdemocracia"/>	<input type="text" value="stfacovardado"/>	<input type="text" value="31mar"/>
<input type="text" value="aragaopetralhao"/>	<input type="text" value="forarodrigomaia"/>	<input type="text" value="ocupapaulista"/>	<input type="text" value="mudabrasil"/>	<input type="text" value="13marco"/>
<input type="text" value="deolhonostf"/>	<input type="text" value="4dezeuvou"/>	<input type="text" value="dia31vaisermaior"/>	<input type="text" value="stfvergonhanacional"/>	<input type="text" value="listatripliceagu"/>
	<input type="text" value="quedaplanalto"/>	<input type="text" value="renunciadacunha"/>	<input type="text" value="ocupasp"/>	<input type="text" value="17deabril"/>

Figure 6.3: Screenshot of the hashtag classification web interface, available at <http://labs.wesleycota.com/twitter>. Hashtags can be moved between four categories, and a field allows to search for tweets created between March and December 2016 to decide their classifications.

The final classification of each hashtag was determined by the majority of the volunteers' opinions. From the total of 495 hashtags, 321 (64.8%) had a full agreement, while for 443 (89.5%) of them, at least three of four volunteers agreed. These 443 hashtags are reported in Tables B.1 to B.3 of Appendix B, colored according to their classification: blue for hashtags used in tweets that convey pro-impeachment sentiments, red for anti-impeachment sentiments, grey for neutral sentiments, yellow for not related hashtags. Dark (light) colors have been used to indicate full (partial) agreement. Divergent opinions were given for 52 (10.5%) hashtags, reported in Table B.4 of Appendix B. Table 6.3 reports the summary of the classification.

6.2.2 Filtering process of the dataset to obtain the networks

After the classification, 404 hashtags with at least 3/4 agreement were used to reconstruct the political communication (PC) network removing tweets with the 39 classified as out of context and the 52 without agreement. In summary, we had 200 anti-impeachment, 185 pro-impeachment, and 20 neutral hashtags. The PC network reconstructed this way is hereafter referred to as the 20-neutral network.

To check if the main results are robust with respect to the hashtags classification, we con-

Table 6.3: Final number of hashtags for each category reported in Tables B.1 to B.3 of Appendix B. The symbols in superscript between parenthesis correspond to the ones used in the tables. Three different levels of agreement are listed: full agreement (full); 3/4 agreement (partial); and less than 3 agreements (divergent). In the modified classification to reconstruct the 72-neutral network, we include the 52 hashtags with divergent classification in the neutral class, see main text.

	full	partial (*)	divergent (²)	total
-1	139	45	—	184
0	3	17	0 (52)	20 (72)
+1	163	37	—	200
×	16	23	52 (0)	91 (39)
total	321	122	52	495

structed also a different PC network where the 52 hashtags for which an agreement was not achieved were considered as neutral. The corresponding modified 72-neutral network is defined by 456 hashtags, 72 considered as neutral, with only the remaining 39 classified as not related being filtered out.

Next, tweets without mentions (explicit or not) were also filtered out, since it is not possible to extract interactions from them. Tweets containing hashtags with opposite sentiments (pro- and anti-impeachment) in the same tweet were also filtered out. From the total of 12 322 322 tweets with hashtags, we had:

- 2 911 655 (23.629%) without mentions (text in form `@user`),
- 7 486 459 (60.76%) with at least one hashtag of the 20-neutral classification,
- 9 908 405 (80.41%) with at least one hashtag of the 72-neutral classification,
- 74 111 (0.6%) with at least two hashtags of opposite sentiments in the same tweet.

The validity of the hashtag classification method is strengthened by the fact that only 0.6% of the collected tweets had hashtags with opposite sentiments. In the case of the 20-neutral, the total number of mentions was 5 050 291, from which 2 327 787 (46.092%) were retweets (RTs). For 72-neutral, 7 596 888 mentions, with 3 837 204 (50.510%) RTs.

However, RTs do not involve an explicit effort of content production and do not convey a specific communication target. For this reason, here we discard RTs from our analysis and focus on tweets that include an explicit mention to another user, to establish or continue a discussion on some topic, carrying even personal messages [168]. This choice allows us to single out

only actual social interactions between users, to reconstruct a communication network in which people actually exchange information, discuss, and form their opinion reacting in real-time to ongoing political events. Discarding RTs, we obtained $N = 285\,670$ users and 2 722 504 explicit mentions for the **20-neutral**, and $N = 437\,728$ users and 3 759 684 explicit mentions for **72-neutral** network. Hereafter, we consider only networks obtained with explicit mentions, i.e., disregarding retweets.

Finally, a temporal network \mathcal{G} was constructed, defined by the set of N vertices (users), $\mathcal{N} = \{1, 2, \dots, N\}$, and the set of interactions $\mathcal{E} = \{e_1, e_2, \dots, e_E\}$. In this case, an interaction between users i and j ($i, j \in \mathcal{N}$) occurs at time t when user i mentions j in a tweet with sentiment s_t , being represented mathematically as $e_t = (i, j, t, s_t)$. Multiple mentions (for different users) in the same tweet imply multiple simultaneous interactions, and the interactions are instantaneous (without duration) and not symmetrical, defining a temporal directed network.

An integrated weighted directed network [169] was obtained from \mathcal{G} , defined by directed edges between i and j whose weights are defined by the total number of times that i mentioned j . If this weight is zero, the edge does not exist, as defined in Chapter 2. Twitter is known to be populated by social bots, which contribute to the spreading of misinformation and poison political debate [170]. To reduce the number of bots, that also interact with humans [23, 171], the strongly connected components [172, 173] (SCCs) of the integrated networks were obtained. This guarantees that in the resulting network the information spreading starting in one user can reach all others; see Sec. 2.1.2. The largest SCCs have $N = 31\,412$ vertices, $E = 833\,123$ links, and $W = 1\,552\,389$ interactions for **20-neutral** network, and $N = 39\,525$ vertices, $E = 1\,063\,699$ links, and $W = 2\,056\,448$ interactions in the **72-neutral** network, see Table 6.4. The other SCCs were much smaller in size compared to the larger one.

In the largest SCC of the **20-neutral** network, the number of interactions with at least one pro-impeachment, neutral, or anti-impeachment hashtag was 1 126 150, 144 405, and 756 498, respectively, showing only a slight tendency for pro-impeachment hashtags, while the number of users was 20 200, 10 821, and 22 566, respectively, showing a remarkable balance. We present the number of tweets containing the 100 most popular hashtags in Fig. 6.4, showing that pro-impeachment hashtags were the most popular but the anti-impeachment ones were more numerous in the top 100.

Hereafter, we will consider only the set of users and interactions belonging to the largest SCCs in the temporal networks. Only results for the **20-neutral** network will be presented,

Table 6.4: Properties of the 20-neutral, 72-neutral integrated networks considering the largest SCC (top) and whole network (bottom). N is the total number of vertices; E is the number of links; $\langle k_{\text{out}}^n \rangle$ is the n -th moment of the number of links; W is the number of interactions; $\langle a^n \rangle$ is the n -th moment of the activity. The average weight of the links is denoted by $\langle W_{ij} \rangle$; N_+ (N_-) is the number of vertices with overall positive (negative) polarization.

Largest strongly connected component:

	N	E	$\langle k_{\text{out}} \rangle$	$\langle k_{\text{out}}^2 \rangle$	W	$\langle a \rangle$	$\langle a^2 \rangle$	$\langle W_{ij} \rangle$	N_+	N_-
20-neutral	31 412	833 123	26.5	4 728	1 552 389	49.4	44 163	1.86	13 925	16 257
72-neutral	39 525	1 063 699	26.9	5 252	2 056 448	52.0	50 111	1.93	16 352	18 340

Whole network:

	N	E	$\langle k_{\text{out}} \rangle$	$\langle k_{\text{out}}^2 \rangle$	W	$\langle a \rangle$	$\langle a^2 \rangle$	$\langle W_{ij} \rangle$	N_+	N_-
20-neutral	285 670	1 696 841	5.94	818.2	2 722 504	9.53	8 242.83	1.60	101 250	125 591
72-neutral	437 728	2 341 473	5.35	768.0	3 759 684	8.59	7 404.21	1.61	101 250	125 591

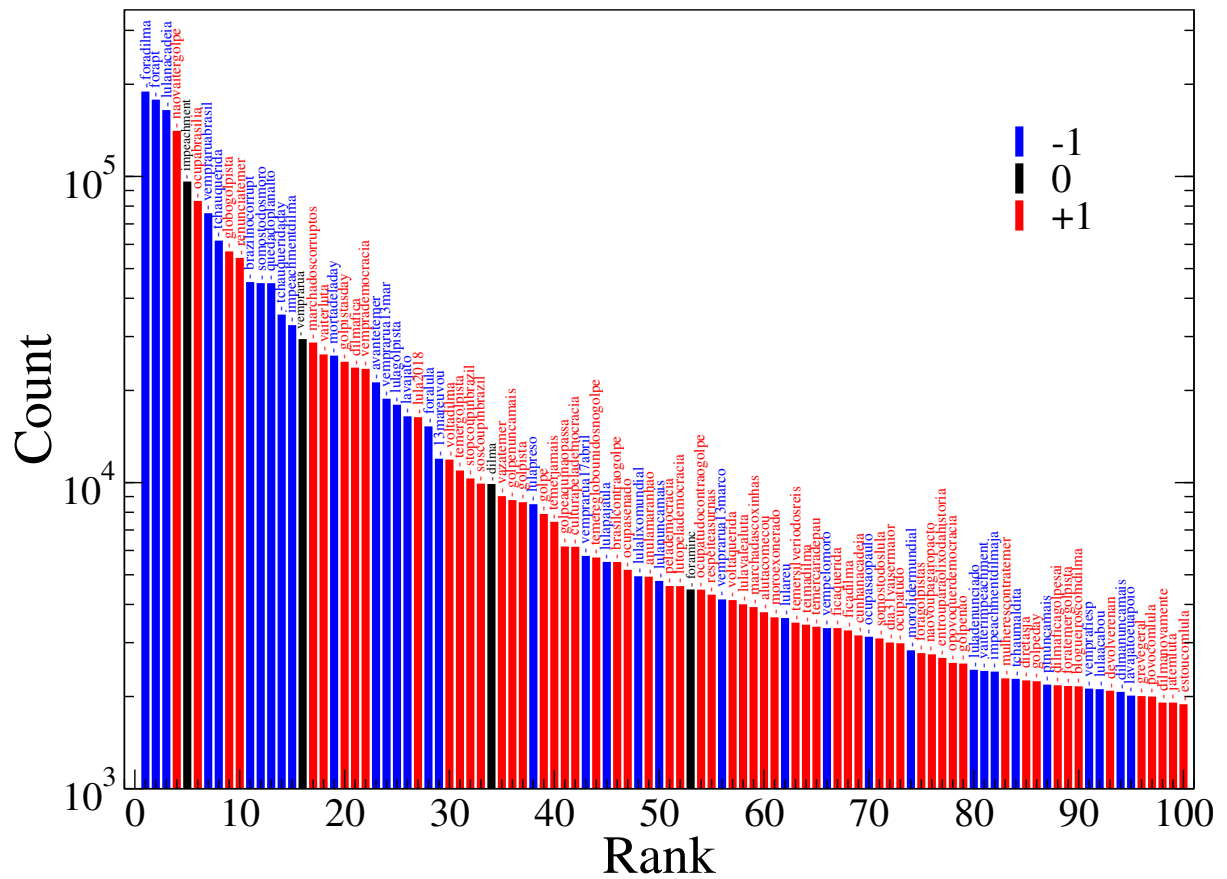


Figure 6.4: Usage count for the 100 most popular hashtags in the SCC of the 20-neutral network. Only manually classified hashtags as pro- (−1), anti-impeachment (+1), or neutral (0) are shown, with colors indicated in the legend.

while those for the 72-neutral network are shown in Appendix B.3 as a robustness analysis of the classification. On March 29, 2016, the largest fraction activity of anti-impeachment (+1) tweets was reached, when political party MDB declared the end of support to Rousseff's government; see Table 6.2. The largest fraction activity of pro-impeachment (−1) was registered

on June 4 and July 29, when Rousseff presented her final defense in the Deputy's chamber. These results are shown in Fig. 6.5.

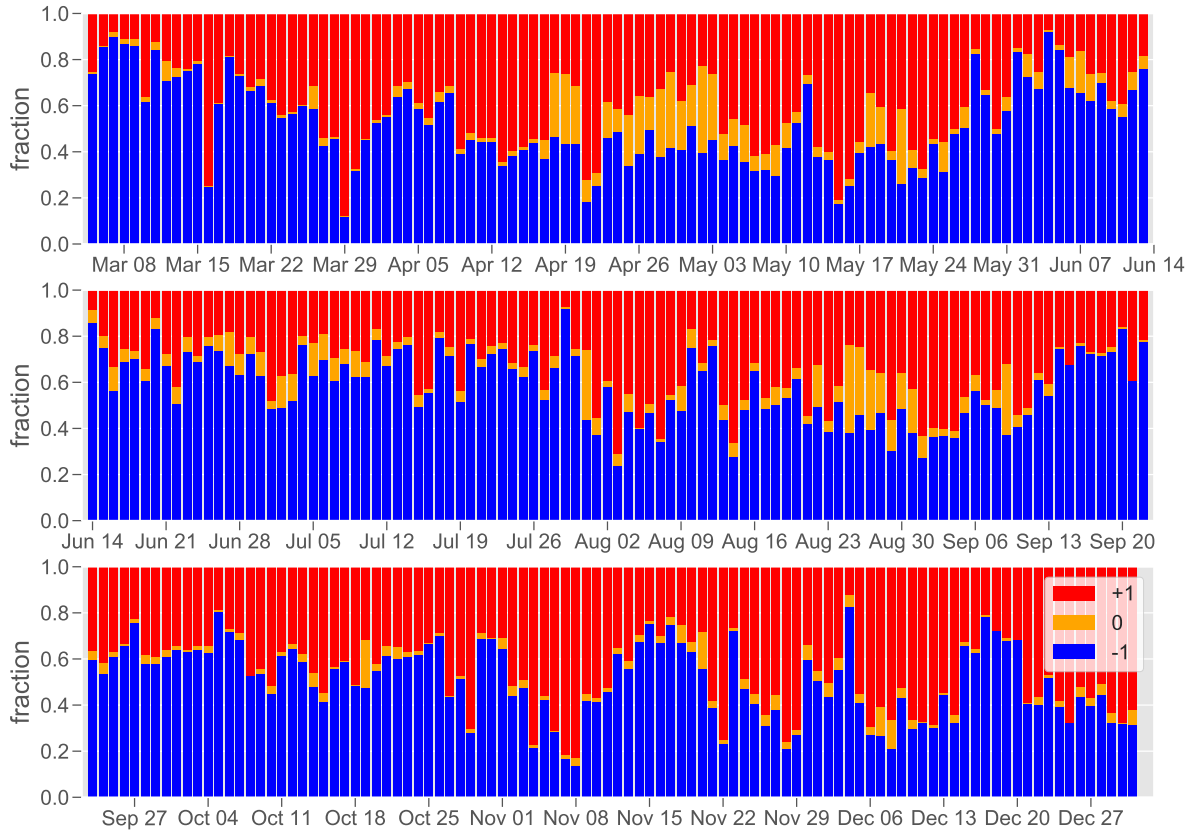


Figure 6.5: Activity frequency of tweets for the SCC of the 20-neutral network. The legend indicates the colors corresponding to the activity for -1 , 0 and $+1$ interactions.

6.2.3 Properties of the temporal and integrated networks

The distributions of activity of tweets with a given sentiment are similar, as shown in Figs. 6.6(a-c), with power-law in the form $P(a) \sim a^{-b}$, with $b \approx 2$. The same happens for the distributions of burstiness, the time interval τ between two sequential interactions of a user, Figs. 6.6(d-f), that follow

$$F(\tau) \approx \frac{\kappa\alpha}{(1 + \kappa\tau)^{\alpha+1}},$$

with parameters close to $\kappa = 0.05$ and $\alpha = 0.25$, as shown in Fig. 6.6(f).

Concerning the integrated network, typical properties of real networks were analyzed for degree distribution $P(k)$, weight distribution $P(W_{ij})$, clustering coefficient $C(k)$ and average nearest-neighbor degree $k_{nn}(k)$; see Fig. 6.7. Both degree and weight distributions follow power-laws, while $C(k)$ and $k_{nn}(k)$ show that the integrated networks are disassortative. In the

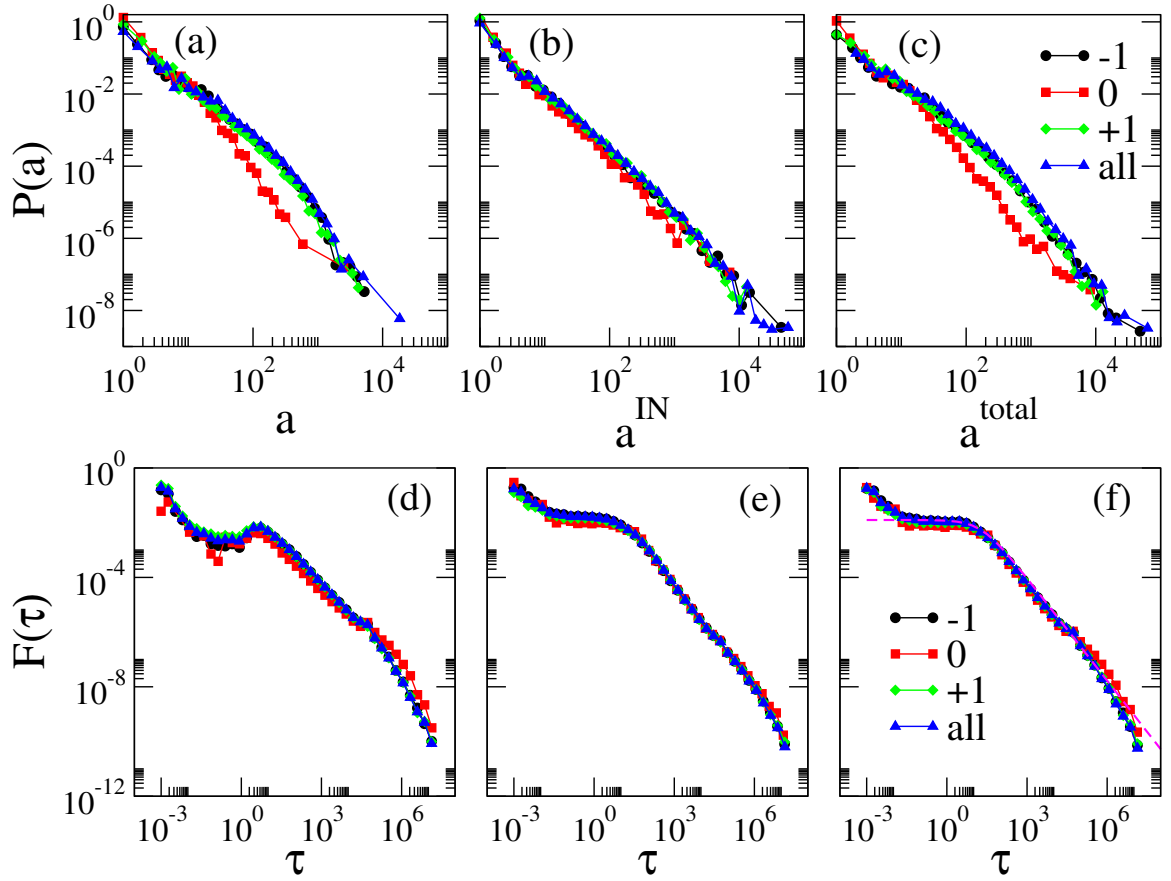


Figure 6.6: Distributions of activity a and burstiness τ (a,d) of senders, (b,e) receivers, and (c,f) total, of interactions with sentiments $-1, 0, +1$ and all. We consider $a^{\text{total}} = a_i + a_i^{\text{IN}}$. The dashed curve in (f) is given by $F(\tau) = \kappa\alpha/(1 + \kappa\tau)^{(\alpha+1)}$, with $\kappa = 0.05$ and $\alpha = 0.25$.

case of the clustering coefficient, we consider the undirected version of the network.

6.3 Defining the political position as a continuous variable

We can assign to a user i a set of tweets with sentiments $\mathcal{S}_i = \{s_1, s_2, \dots, s_{a_i-1}, s_{a_i}\}$, where a_i is her/his activity, and define the average sentiment of tweets, or political position P_i , as

$$P_i \equiv \frac{\sum_{t=1}^{a_i} s_t}{a_i}, \quad (6.1)$$

that is bounded in the interval $[-1, 1]$. In Fig. 6.8(a) the distribution of political position is presented, showing that users are split in two groups of different orientations, while few of them present neutral position ($P \approx 0$), being also asymmetric in relation to $P = 0$: for $P > 0$ most of them have extreme position $P \approx +1$, while for $P < 0$ there is a decreasing variation, with different levels of negative average position. The number of users with positive values N_+

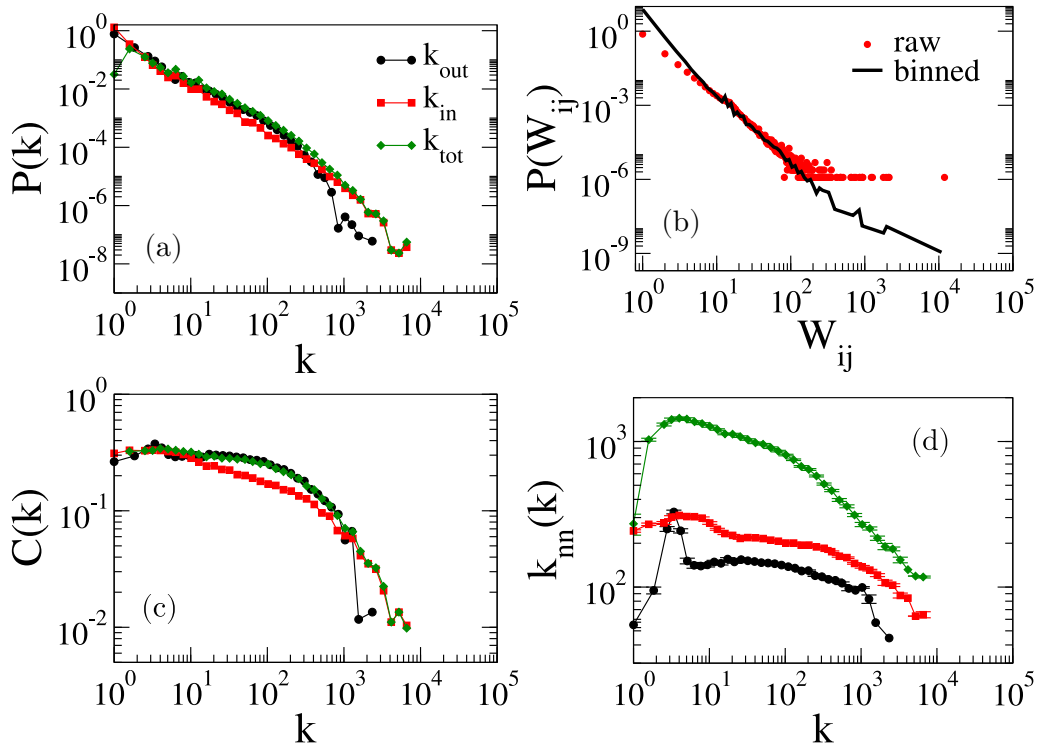


Figure 6.7: Structural properties of the SCC integrated network: (a) degree, (b) weight, (c) clustering and (d) average nearest-neighbor degree distributions.

and negative N_- are of the same order, as seen in Table 6.4. The average sentiment of a user is inherently correlated to her/his activity. In a scenario in which users send tweets with opposite sentiments with same probability, this distribution would be binomial. Figure 6.8(b) shows that users more active are also the most extreme. For pro-impeachment position, more active users have $P \approx -0.75$, while for anti-impeachment the activity is almost constant for $0 < P < 0.5$, and increasing for $P > 0.5$.

Figure 6.8(c) shows a visualization of the integrated network: a user i is represented by a circle, with color given by her/his political position P_i , and lines represent interactions with other users. Two communities of opposite political positions are clearly seen in the graphical representation, while users with neutral position are making “bridges” between these groups. The representation was obtained using the *ForceAtlas2* algorithm [174] in the Gephi software [134]. A way to quantify this visual observation is by the identification of the community structure [175] using the Louvain algorithm [69] to compute the modularity Q_{mod} , as done in Sec. 5.3; see Sec. 2.1.4. Figure 6.8(d) confirms the existence of two big communities, both with approximately 10^4 users, and opposite positions of values $P_+ \approx 0.84$ and $P_- \approx -0.69$; see Table B.5. However, users with pro-impeachment position also form other communities with considerable sizes and moderate positions, differently from users with anti-impeachment

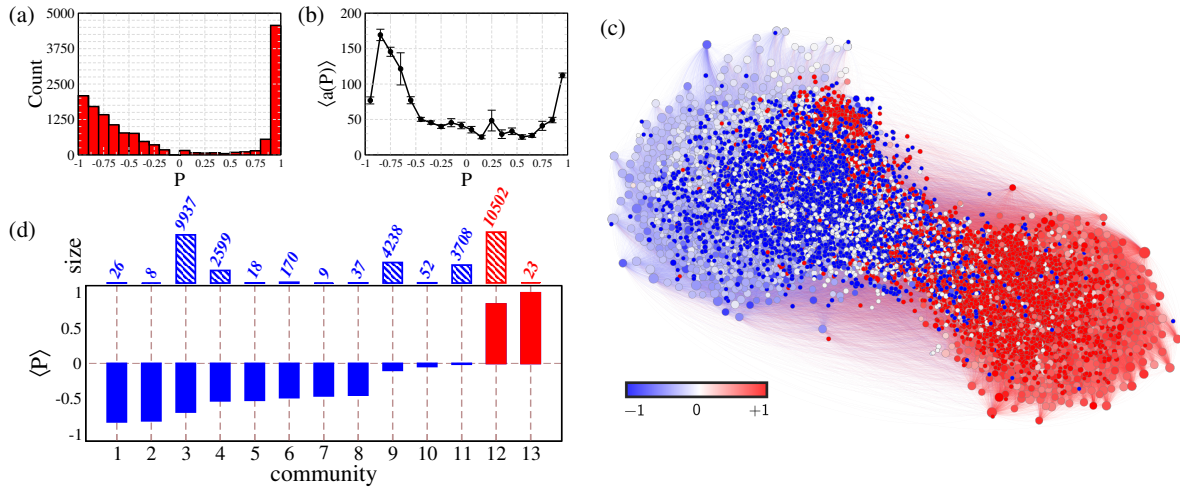


Figure 6.8: (a) Number of users as a function of polarization. (b) Average activity as unction of polarization. Only users with activity $a \geq 10$ in the SCC are considered for (a) and (b). (c) Visualization of the time-aggregated representation of the PC network, formed by $N = 31\,412$ users in the SCC. The size of vertices increases (non-linearly) with their degree. Colors represent political polarization, as defined by (6.1), blue for negative, red for positive, and white for neutral polarization. (d) Community size and average polarization of different communities identified by the Louvain algorithm [69].

positions. We will discuss the consequences of these structures in the sequence.

6.3.1 Content diversity and political position

Another interesting quantity is the content diversity of tweets sent by user i , defined as the variance of sentiments expressed by i ,

$$\omega = \frac{\sum_{t=1}^a (s_t - P)^2}{a}. \quad (6.2)$$

The sentiments can be $-1, 0$ or $+1$. So, for a given user, we denote n_+ as the number of tweets with $s_t = +1$, n_- with $s_t = -1$ and n_0 with $s_t = 0$. The total number of tweets is then $a = n_+ + n_- + n_0$. So, Eq. (6.1) becomes $P = (n_+ - n_-)/a$ and Eq. (6.2) is $\omega = [n_+(1 - P)^2 + n_-(1 + P)^2 + n_0P^2]/a$. In Fig. 6.9(a) we plot contour maps of the content diversity ω of users in the SCC of the 20-neutral network with activity $a \in [10, 100]$. By this density plot we see a large number of pro-impeachment users with large content diversity, while users with anti-impeachment positions are clustered together with reduced content diversity. Let us consider three limiting scenarios in which users with $P > 0$ do not tweet pro-impeachment messages, and those with $P < 0$ anti-impeachment messages. We then can write ω as a function of the political position P :

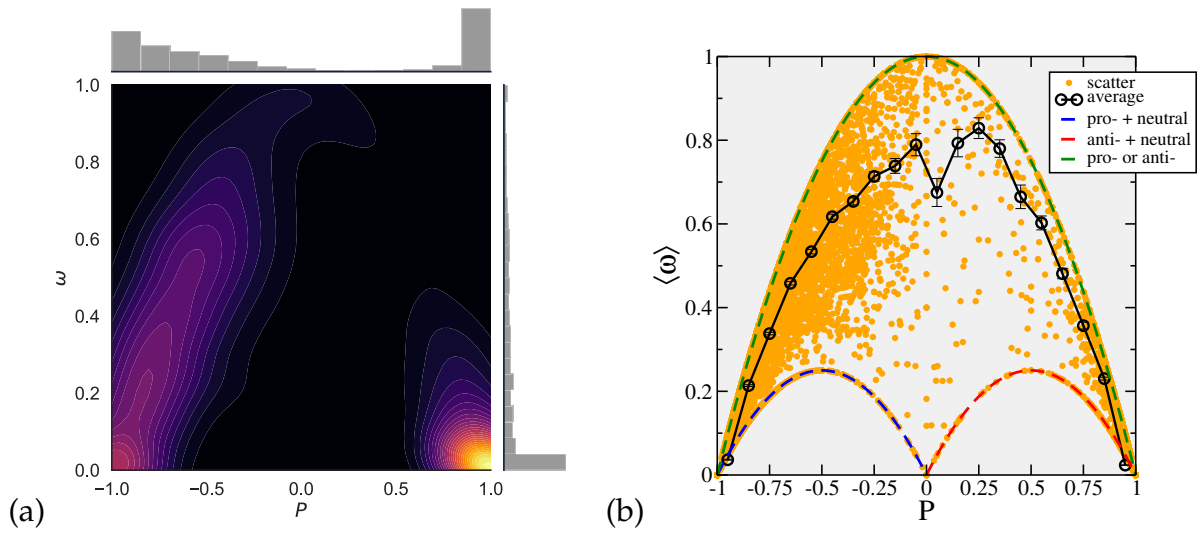


Figure 6.9: (a) Contour maps and (b) scatter plot for the content diversity ω of tweets sent by users with political position P . In (b) points represent users and their content diversity, dashed lines the predictions by Eqs. (6.3) to (6.5) and circles the average value as a function of the political position P . Only users with $a \in [10, 100]$ (corresponding to 11 386 users) are considered.

- $n_+ = 0$: users send only neutral or pro-impeachment tweets. In this case, $P \leq 0$. We have

$$n_- = -Pa \text{ and } n_0 = a + Pa. \text{ So,}$$

$$\omega(P)|_{n_+=0} = -P(1 + P), \quad -1 \leq P \leq 0. \quad (6.3)$$

- $n_- = 0$: users send only neutral or anti-impeachment tweets. In this case, $P \geq 0$. We have

$$n_- = Pa \text{ and } n_0 = a - Pa. \text{ So,}$$

$$\omega(P)|_{n_-=0} = P(1 - P), \quad 0 \leq P \leq 1. \quad (6.4)$$

- $n_0 = 0$: users send only polarized tweets. In this case, we have $Pa = n_+ - n_-$ and $a = n_+ + n_-$. So, we have $n_+ = a(1 + P)/2$ and $n_- = a(1 - P)/2$, which leads to

$$\omega(P)|_{n_0=0} = 1 - P^2, \quad -1 \leq P \leq 1. \quad (6.5)$$

In Fig. 6.9(b) we show a scatter plot version of Fig. 6.9(a) for comparison with the last three limiting cases, together with the average $\langle \omega \rangle(P)$. Users with extreme positions have small content diversity by definition since the majority of tweets sent by them are either pro- or anti-impeachment. However, users with intermediate positions behave differently for pro- and anti-impeachment average positions. While pro-impeachment users tend to be in a scenario in which those with larger content diversity tend to use neutral hashtags, few anti-impeachment users appear using them. This can be related to the breaking of echo chambers that will be discussed later.

6.3.2 Structural evidences of echo chambers

One way to quantify the existence of echo chambers is by the relationship between the political position a user expresses and the sentiments of tweets she/he receives, or with the average position of its neighbors in the integrated network. In politics, echo chambers are characterized by users sharing messages with similar opinions [50]. From the structural point of view in a network, this translates in a vertex i with position P_i connected with vertices with positions similar to P_i , receiving with larger probability messages with average sentiment similar to P_i . We then define for each user i the average position of received tweets P_i^{IN} , applying (6.1) to the set of tweets of any user $j \neq i$ mentioning i . The results are shown in Fig. 6.10(a). Analogously, the average political position of the neighbors of i , P_i^{NN} , can be defined by its successors,

$$P_i^{\text{NN}} \equiv \frac{1}{k_{\text{out},i}} \sum_j A_{ij} P_j,$$

or predecessors,

$$P_{\text{in},i}^{\text{NN}} \equiv \frac{1}{k_{\text{in},i}} \sum_j A_{ji} P_j.$$

Results for both cases are similar, as seen in Fig. 6.10(b-c).

Figure 6.10 shows the correlations between the political position of a user i and of its neighbors. These strong correlations confirm the existence of echo chambers. Yet, it also shows that densities are not symmetric between anti- and pro-impeachment: for $P > 0$ users are more concentrated in a small region, while for $P < 0$ are more spread out. While users with pro-impeachment position exchange information with pairs of distinct positions, anti-impeachment exchange with extreme users. This is in line with the community structure shown in Fig. 6.8(d).

6.3.3 Spreading capacity as a function of political position

The presence of echo chambers can have an impact on how information is disseminated through PC networks. A usual methodology to gauge these effects is by computing the number of RTs that a given user can achieve [176–178]. Influential users could attract more interest from others, accomplishing a larger number of RTs. In Figure 6.11 the number of RTs of users is presented as a function of both her/his activity and political position. As expected, this number is strongly correlated with her/his activity, since a user producing many tweets has a larger chance of being retweeted, in a homogeneous assumption of equal probabilities. Instead, if we

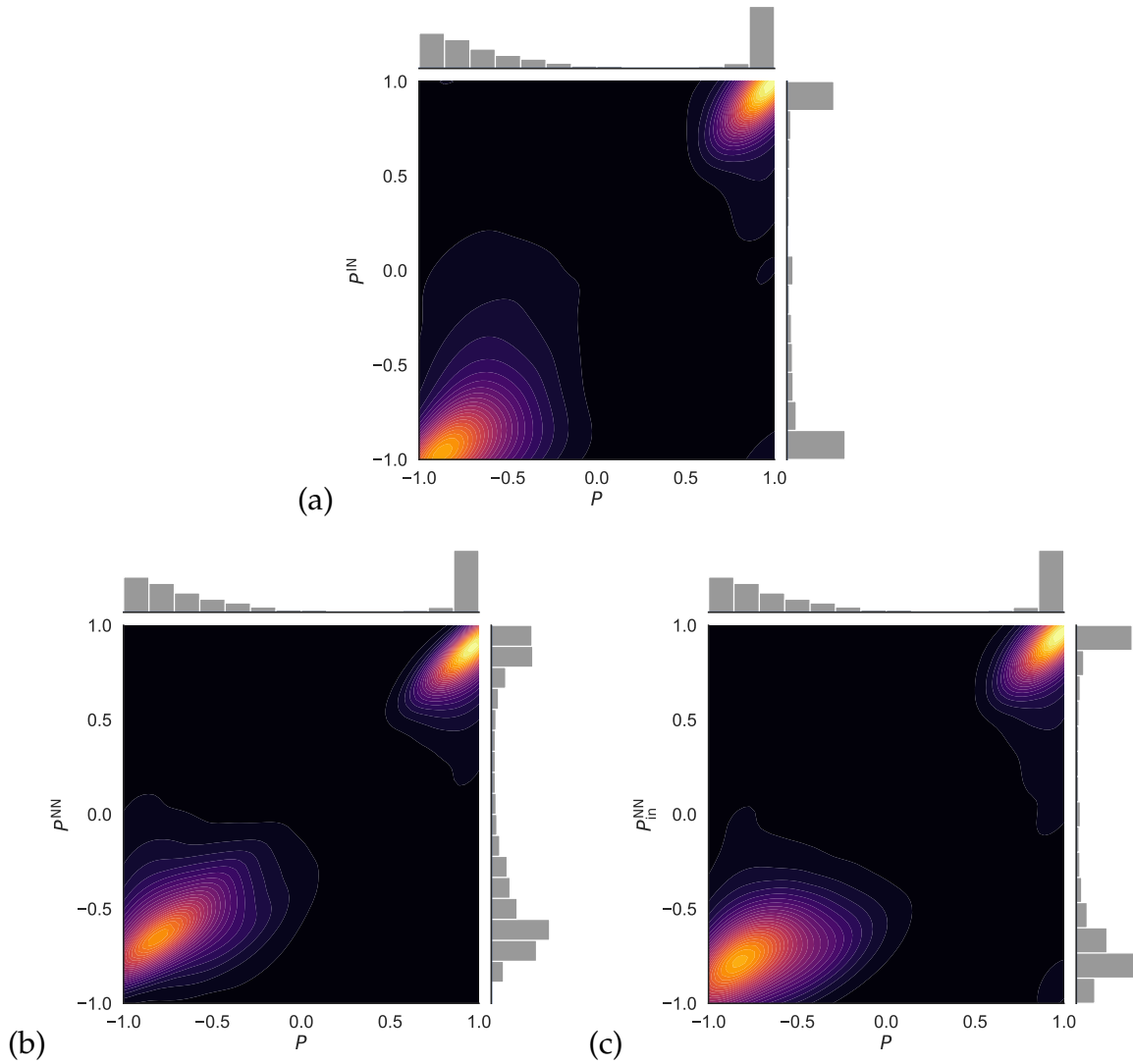


Figure 6.10: Contour maps for average (a) political position of tweets received P^{IN} , (b) sentiment of successors P^{NN} and (c) predecessors $P^{\text{NN}}_{\text{in},i}$ against the political position P of a user for the 20-neutral network. Colors represent the density of users: the lighter the larger the number of users. Probability distribution of P , P^{NN} and $P^{\text{NN}}_{\text{in},i}$ are plotted in the axes. Only users with activity $a \geq 10$ (corresponding to 14 813 users) are considered.

consider this number normalized by the activity, there is a lack of evident correlations with the political position, as shown in Fig. 6.11(c).

Here we propose a different methodology by considering spreading models of SIS and SIR-like [88], that have been used to study information diffusion [179, 180]. In this context, information spreading can be studied both as a simple contagion or as complex contagion, like in the adoption of new ideas and technologies [181]. The first case is similar to an epidemics, in which the infection or receiving of information or opinion happens when at least one individual tries to infect or inform other, as in the SIS and SIR models. In the case of complex contagion, infection or adoption is only accepted by social reinforcement, when more than one interaction

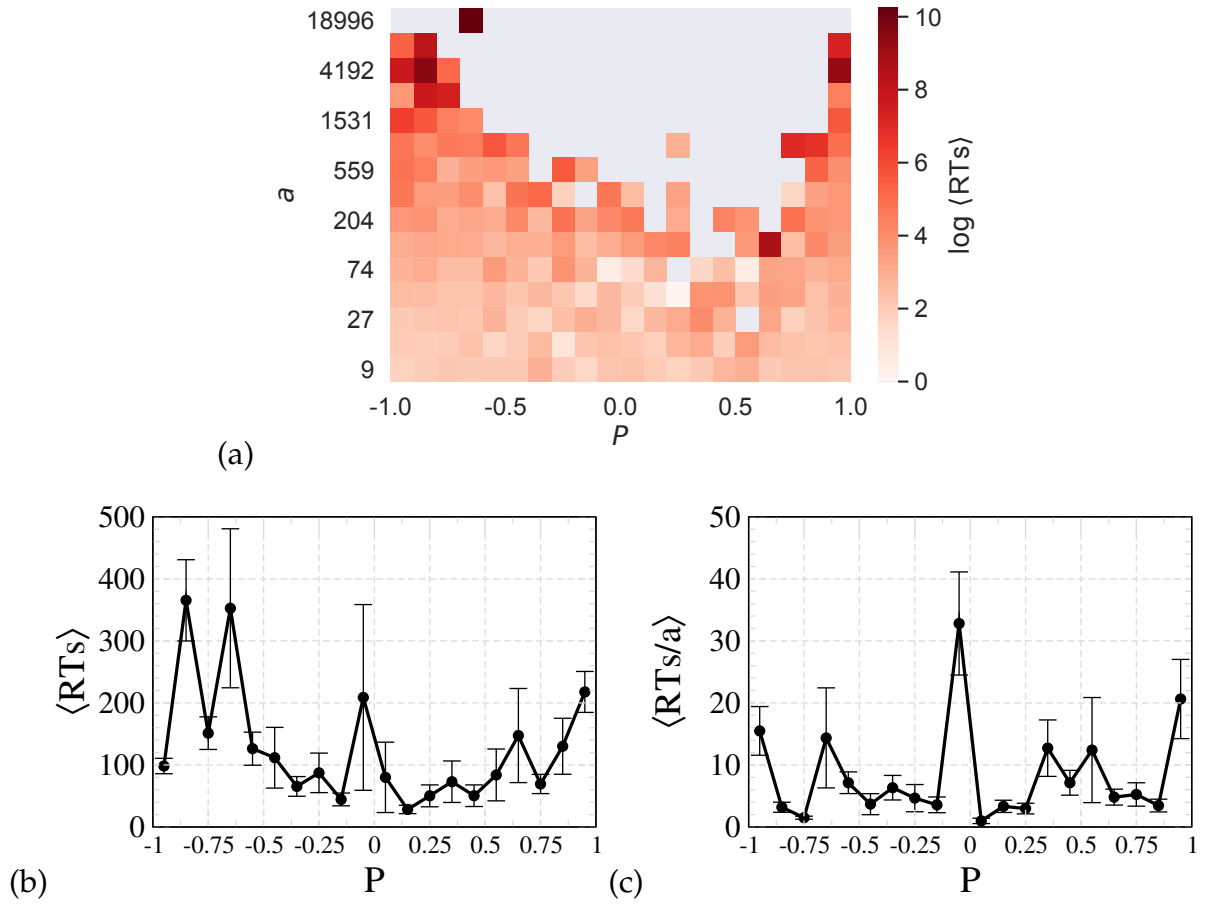


Figure 6.11: Number of retweets (RTs) received by users of the 20-neutral network in the classified data: (a) heat map of the number of RTs of users as a function of their political position P and activity a , (b) average number of RTs and (c) normalized by the user activity as a function of the political position. Error bars represent the standard error.

is needed to convince. In both cases, a susceptible vertex is unaware of the information, while an infected one is aware and can spread it. A recovered vertex, in the case of a SIR-like model, does not engage in discussions. The healing is given by a parameter $\tau = \bar{\mu}^{-1}$, fixed for all users. After a time τ of infection, a vertex becomes susceptible or recovered, for the SIS and SIR cases, respectively. Note that in this case, we need to consider a discrete dynamics, defined by the timestamps of tweets, and transmission probabilities $\bar{\lambda}$ and time τ instead of rates.

Let us begin with simple contagion models SIS and SIR. The complex contagion case will be presented later in Appendix B.5, with similar results as a robustness analysis. Figure 6.12 presents a schematic representation of both dynamics. The initial condition consists of a single individual i , referred to as source, infected while all others are susceptible. We follow the explicit temporal evolution of edges for each timestep t . So, the infection can happen only when at a time t a user i mentions j while i is infected. The set of users infected at least once

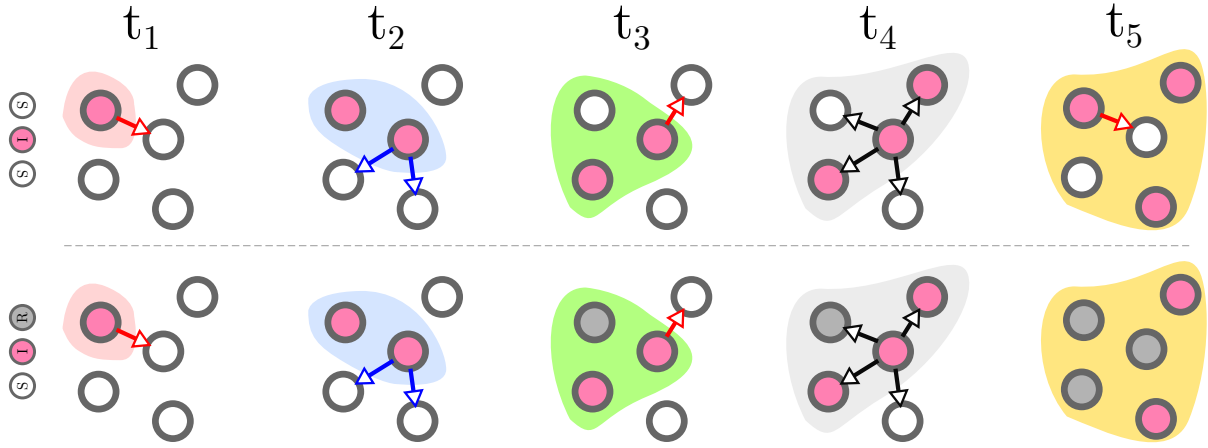


Figure 6.12: Top: SIS dynamics. Bottom: SIR dynamics. In both cases, at time t_1 the source (the first vertex infected in the dynamics) tries to transmit the information to the user mentioned in a tweet. A cascade is formed, increasing the set of influence of the source (shaded areas), defined by the set of users that have been infected once in this dynamics. Infected vertices can become susceptible (SIS, white vertices) or recovered (SIR, gray vertices). In the SIS these vertices can be informed again, but do not increase the set of influence since they had been infected before. For the SIR dynamics, recovered individuals cannot be reinfected, reducing the spreading capacity of the source by blocking possible paths of information spreading.

in the dynamics starting from source i is called the set of influence of i , denoted by \mathcal{I}_i [182]. This set represents the spreading capacity of i , or spreadability, that will be a function of the transmission probability $\bar{\lambda}$ and healing time τ :

$$S_i(\bar{\lambda}, \tau) \equiv \frac{|\mathcal{I}_i(\bar{\lambda}, \tau)|}{N}, \quad (6.6)$$

running the SIS or SIR dynamics with i as source, with averages by different stochastic realizations. In Fig. 6.13 color maps of the average spreadability of users as a function of their political position P and activity a are presented for the SIS model with healing time $\tau = 7$ days and different infection probabilities $\bar{\lambda}$. As expected, the greater the activity of users, the greater is their spreading capacities (darker colors), but are not homogeneous in relation to the political position. The spreadability is smaller for users with anti-impeachment positions while it has a maximum for $P \approx -0,5$ with pro-impeachment position.

As pro- and anti-impeachment groups have similar sizes, the spreading capacity cannot be explained only by prevalence of users with different average sentiments. Also, users with extreme position $|P| \approx 1$ have smaller spreadability than those with intermediate pro-impeachment positions. We can understand that by characterizing the set of users influenced by the spreading dynamics. With this aim, we define the average of the polarization of the set

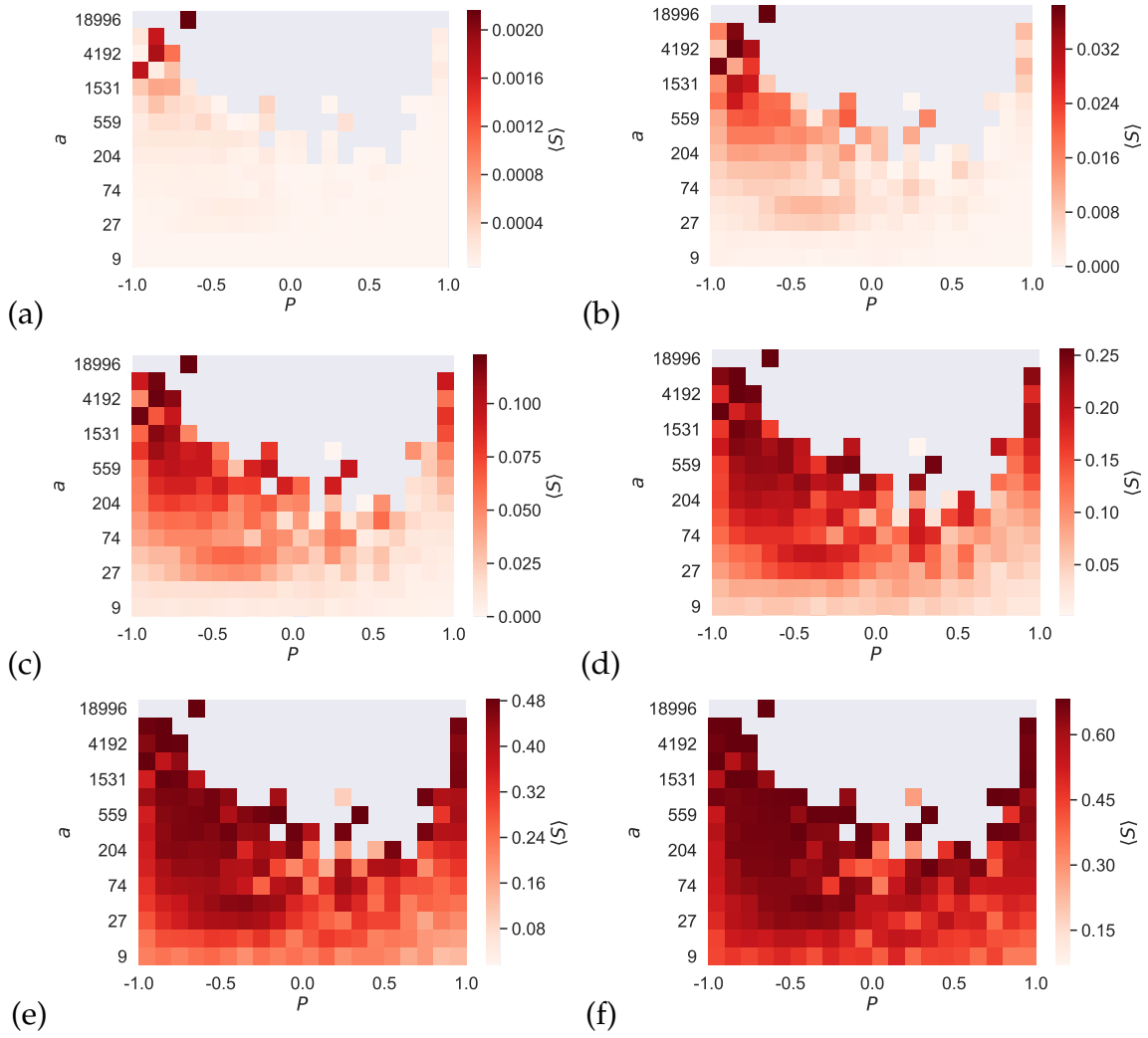


Figure 6.13: Heat maps of the average spreading capacity $\langle S \rangle$ of users generated with the SIS model as a function of their political position P and activity a for temporal network with healing time $\tau = 7$ days for the 20-neutral network and transmission probability (a) $\bar{\lambda} = 0.01$, (b) $\bar{\lambda} = 0.05$, (c) $\bar{\lambda} = 0.1$, (d) $\bar{\lambda} = 0.2$, (e) $\bar{\lambda} = 0.5$, and (f) $\bar{\lambda} = 1$. Averages were performed over 100 runs.

of influence of i as

$$\mu_i \equiv \sum_{j \in \mathcal{I}_i} \frac{P_j}{|\mathcal{I}_i|}, \quad (6.7)$$

and the variance as

$$\sigma_i \equiv \sum_{j \in \mathcal{I}_i} \frac{(P_j - \mu_i)^2}{|\mathcal{I}_i|}. \quad (6.8)$$

While the average μ_i represents the average sentiment of users influenced by i , the variance σ_i quantifies the diversity of this set. To isolate the effect of the political position of sources i , we analyze the spreadability for users with similar activity as, for example, $a \in [10, 100]$. The number of users inside this interval of activity is almost uniform, and qualitative results are

robust to this choice; see Fig. 6.14.

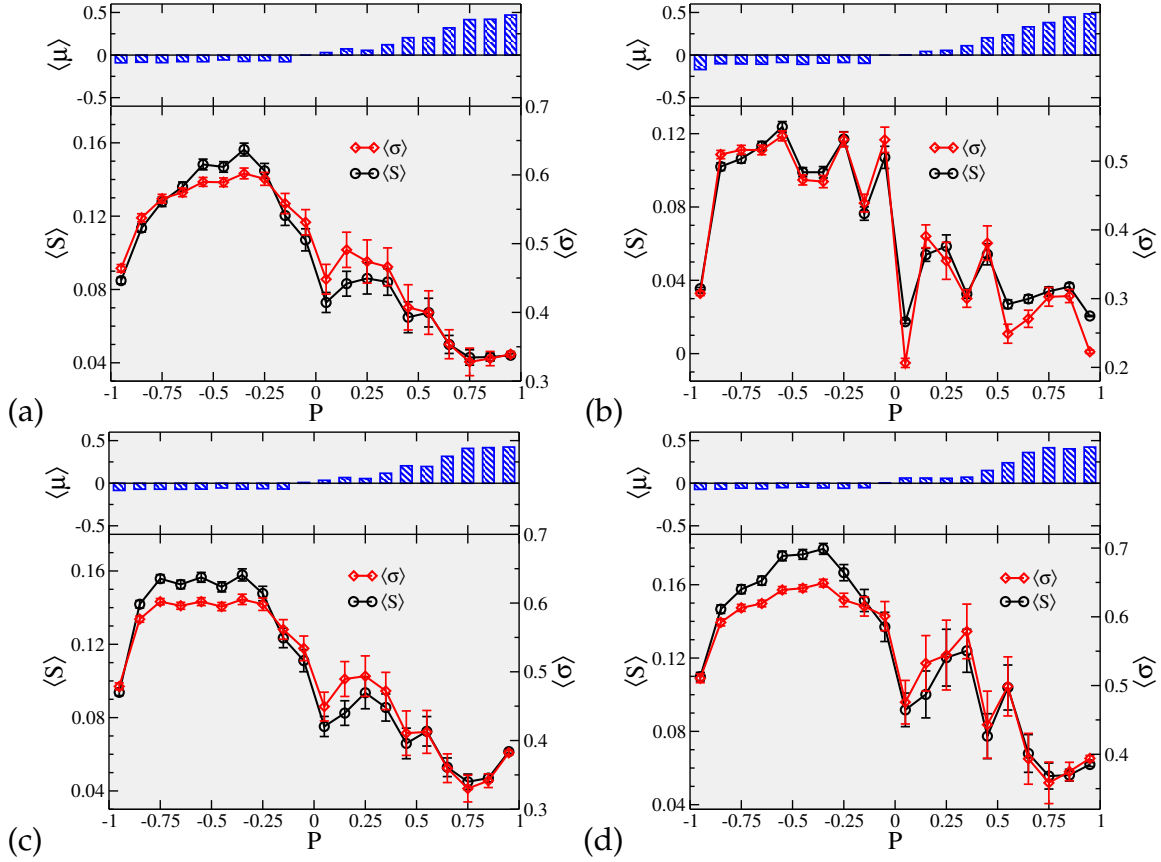


Figure 6.14: Average spreading capacity $\langle S \rangle$ (black, left axes), diversity σ (red, right axes), and average political position μ (top panel) of the set of influence \mathcal{I} , as a function of the political position P , for SIS model with transmission probability $\bar{\lambda} = 0.2$ and $\tau = 7$ days for the temporal 20-neutral network. Only users with activity (a) $a \in [10,100]$, (b) $a \in [1,100]$, (c) $a \in [10,500]$ and (d) $a \in [20,200]$ are considered, in a total of 11 386, 27 985, 14 313 and 10 409 users, respectively. Averages were performed over 100 runs.

Analyzing the spreading capacity as a function of P , $\langle S(P) \rangle$, results for SIS dynamics with $\tau = 7$ days and $\bar{\lambda} = 0.20$ are presented in Fig. 6.14(a). In the top panel, the average political position of users influenced by sources of position P , $\langle \mu(P) \rangle$, is correlated to P , quantifying the echo chamber effects: users tend to influence other users with a similar average political position. The average $\langle \mu(P) \rangle$, indeed, gauges the strength of the echo chambers: the more the value of $\langle \mu(P) \rangle$ is close to P , the stronger the echo chamber effect. Furthermore, one can note differences between users with pro- and anti-impeachment positions, μ is almost constant for negative values of P , so echo chamber effects are smaller, while μ is growing almost linearly for positive P , indicating stronger echo chambers effects.

Even more interesting, the behavior of diversity σ_i of the users reached by i strongly depends on her/his political position P_i . The curve of the average diversity as a function of the position,

$\langle\sigma(P)\rangle$, follows a behavior remarkably similar to the average spreadability of users with position P , $\langle S(P)\rangle$, in all cases of Fig. 6.14. The strict correlation observed between $\langle\sigma(P)\rangle$ and $\langle S(P)\rangle$ indicates that if a user is able to reach a diverse audience, formed by users that do not share her/his average sentiments, then the size of her/his set of influence is much larger. This means that individuals with large spreadability are able to break their echo chambers. Note that this result is not trivial since the sizes of the echo chambers are much bigger than the number of users reached. Moreover, the value of $\langle\sigma(P)\rangle$ is statistically significant and does not depend on the number of users considered in the average, as shown in Figs. 6.14(b-d). For instance, there are much more users with extreme orientations ($|P| \simeq 1$) than users with intermediate position ($P \simeq -0.5$), yet it holds $\langle\sigma(P \simeq -0.5)\rangle \gg \langle\sigma(|P| \simeq 1)\rangle$. Furthermore, given the larger number of users considered, error bars for $\langle\sigma(|P| \simeq 1)\rangle$ are smaller than the ones for $\langle\sigma(P \simeq -0.5)\rangle$. Analysis for different choices of parameters $\bar{\lambda}$ and τ , and for the SIR dynamics are shown in Appendix B.4.

More realistic models of information diffusion, such as complex contagions, independent cascade, and linear threshold models [183–186], could be used to estimate the spreadability of individuals. We have checked numerically that a modification of the classic Watts threshold model for complex contagion [183] to the framework of temporal networks [187] leads to the same behavior observed in the SIR and SIS models; see Appendix B.5.

6.4 Summary and discussions

The effects of echo chambers on the openness of online political debate have been argued by the scientific community. Their impact in real communication networks, however, remains poorly understood. The main contribution of this work is twofold: i) we quantify the presence of echo chambers in the Twitter discussion about the impeachment of former Brazilian President Dilma Rousseff, showing that communities of users expressing opposite political positions emerge in the topological structure of the communication network, and ii) we gauge the effects of such echo chambers on information spreading, showing that the capability of users to spread the content they produce depends on their political orientation. The use of spreading models allows us to characterize the internal structure of echo-chambers, showing that users belonging to the same echo chamber, with different convictions (i.e., the intensity of their attitude to the impeachment issue), can have quite different spreading capabilities.

Our method to quantify echo chambers is built upon two main ingredients: i) we reconstruct a communication network based on mentions, in which people can actually discuss and exchange information related to ongoing political events, and ii) we define a continuous political position measure, by classifying hashtags as expressing sentiments in favor or against the impeachment, which is independent of the network's reconstruction. We then observe that anti- and pro-impeachment sentiments clearly separate into different communities in the PC network. It is important to remark that, while it is well known that networks formed by RTs can have a strong partisan structure, since RTs generally imply endorsement, this observation is new for mention networks, in which users characterized by opposite sentiments can easily interact [149].

These two clusters of users sharing similar opinions, or echo chambers, can be characterized by looking at the correlations between the in-flow and out-flow of sentiments, as well as between the average sentiments of an individual and her/his nearest neighbors. The topologies of the two echo chambers, however, are not exactly equivalent. Users expressing anti-impeachment sentiments tend to lean towards the extreme, achieving a position $P \simeq +1$, while users with pro-impeachment sentiments show smoother tendencies, reflected into the presence of medium-sized communities with an overall negative position.

The differences in the topological structure of the two communities can be related to the political context under study: while users characterized by anti-impeachment sentiments refer to a more homogeneous political area (*Partido dos Trabalhadores* and small left-wing parties), pro-impeachment users share different political views, including center and right-wing positions, and show different levels of sympathy in favor to the impeachment. Another possible and important source of asymmetry is the constant release of content from other sources, in particular from the traditional media, broadcasting mostly contents that stimulate pro-impeachment sentiments, possibly reinforcing their dissemination to a more diversified audience.

By running simple models of information spreading, we observed that, on average, users are more likely to receive information from peers sharing the same average sentiments. We then see that people with predominantly pro-impeachment sentiments are able to broadcast their message to a potentially larger audience than other users. Furthermore, such audiences are also characterized by a greater diversity of opinions, indicating that pro-impeachment sentiments can spread to users expressing both pro- and anti-impeachment attitudes, a signature that echo chambers can be broken.

Chapter 7

Metapopulation structures with heterogeneous subpopulations

The emergence and accessibility of large datasets, especially of social behavior, is helping the study of the influence of contact patterns in epidemic models. In particular, the availability of data about the mobility of individuals in different levels [51, 188], from global to urban, is being essential to refine dynamical models with compartmental dynamics [59]. Data-driven models are developed to improve the spatial-temporal accuracy of predictions of recent epidemic outbreaks [53, 62–64, 189, 190] using large amount of real data as inputs. However, this refinement makes the analytical treatment of these models more complicated, and theoretical frameworks are needed to gather mobility data and allow at the same time a theoretical analysis. Recently, a new formalism able to overcome some of these difficulties [61, 191] was developed, being applied to situations in which data about urban demography and mobility patterns are available. In this chapter, we define an extension of the model developed in Ref. [61] including different social contact patterns in the populations. We perform an analytical study in synthetic metapopulation networks and discuss the effects of contact heterogeneities in the epidemic threshold.

7.1 Epidemics in metapopulations with recurrent mobility

One possible way to represent mathematically mobility patterns is by using metapopulations. In this case, individuals are considered to live in a set of subpopulations (or patches) and flows of individuals happen between these patches. For the epidemic modeling, spreading

processes are characterized by local reactions inside each patch [192, 193] and global diffusion processes or mobility between them. Usually, these local interactions are assumed to happen homogeneously, without the explicit introduction of different contact patterns as those considered in PL networks studied in the previous chapters. In the last decade, the use of metapopulation models incorporating real data of flows of individuals between different geographical areas has been made [52, 60, 194, 195] to balance the refinement capability and the ability to perform analytical predictions from these models. However, these techniques are commonly based on assumptions that limit the direct application in real situations. As an example, a natural hypothesis of these models is to assume simplified mobility patterns such as random diffusion of agents [196, 197] or continuous models of commuting flows [198–200]. In this perspective, many interesting analytical results have been found for the influence of the mobility pattern on the epidemic threshold [60].

New formalisms were developed [201–204], in which the random patterns of individuals are put aside to include the recurrent nature of most human displacements as from home to work and vice-versa in big metropolitan areas [52]. Yet, this refinement was done at the expense of continuing to use very simple attributes of subpopulations by assuming a homogeneous mixing of individuals. A particular framework was proposed in Ref. [61], allowing to compute analytically the epidemic threshold and to extract useful information to tackle real epidemics when data are available. Recently, this model was adapted to study the COVID-19 situation in Spain [62, 64], and to observe the outbreak variability in distinct geographical levels in Brazil [63]. In Sec. 7.3 we include one more important ingredient, which is the intrinsic heterogeneous nature of human contacts. This is done by a generalization of the model defined in Ref. [61] described in the next section.

7.2 Markov-chain approach in networks with recurrent mobility

Consider a metapopulation network with $i = 1, \dots, \Omega$ patches (vertices) populated by $N = \sum_i n_i$ individuals. Each individual is associated to its residence (one of the patches) and can travel to another patch following mobility flows. The flow of individuals from patches i to j can be described by a directed and weighted network of patches, in which $A_{ij} = 1$ means that at least one individual of i travels to j , and the weight W_{ij} is the number of individuals from i that commute to j daily; see Fig. 7.1. The matrix W_{ij} is also called origin-destination (OD) matrix. It

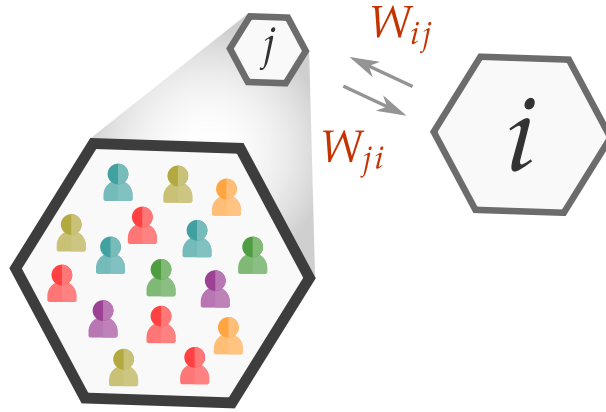


Figure 7.1: Example of a simple metapopulation with two patches i and j . The weights W_{ij} and W_{ji} count the number of individuals that travel from patch i to j and vice-versa, respectively. Inside each patch there are individuals that interact according to a compartmental dynamics.

is useful to define the probability that once an individual living in i decides to move, it goes to patch j with probability

$$R_{ij} = \frac{W_{ij}}{\sum_{l=1}^N W_{il}}, \quad (7.1)$$

which is the normalized weight. After each time step, all individuals return to their residence.

We can describe the recurrent dynamics as follows. Starting the dynamics at time t and evolving it to $t + 1$, we have:

- (i) Each individual residing in i choose to travel to a patch j with probability p , proportionally to R_{ij} ;
- (ii) Residents of i interact with individuals in patches j to where they moved. The evolution of the fraction ρ_i of infected individuals residing in i is computed;
- (iii) After one time step $\Delta t = 1$, individuals return to their resident patch.

The recurrent mobility hypothesis guarantees that all individuals living in i will continue to have this place as origin for their movement in each time step. Since we will consider that all individuals have a same probability p of moving to a different patch, the number of individuals resident in j that travel to i is given by

$$n_{j \rightarrow i} = [(1 - p) \delta_{ij} + p R_{ji}] n_j, \quad (7.2)$$

where we assume that all individuals of j have the same probability R_{ji} to travel to i . The first term represents individuals that stay in their residence, while the second term computes the

probability of moving to a patch i . This way we can compute the total number of individuals traveling to patch i as

$$\tilde{n}_i = \sum_j n_{j \rightarrow i} = \sum_j [(1-p) \delta_{ij} + p R_{ji}] n_j, \quad (7.3)$$

where tilde variables represent quantities computed after step (i) takes place.

Let us consider a discrete-time SIS dynamics and compute the time evolution of the fraction of infected individuals resident in patch i at time t . The healing probability will be represented by $\bar{\mu}$ and the infection probability by $\bar{\lambda}$. In this case, the number of infected individuals traveling from j to i is $I_{j \rightarrow i} = [(1-p) \delta_{ij} + p R_{ji}] I_j$, where I_j is the number of infected individuals that live in patch j , and the total number of infected individuals in i is given by $\tilde{I}_i = \sum_j I_{j \rightarrow i}$. We can express the fraction of infected individuals that lives in i as $\rho_i = \frac{I_i}{n_i}$, while

$$\tilde{\rho}_i = \frac{\tilde{I}_i}{\tilde{n}_i} = \frac{\sum_j I_{j \rightarrow i}}{\sum_j n_{j \rightarrow i}} = \frac{\sum_j [(1-p) \delta_{ij} + p R_{ji}] \rho_j n_j}{\sum_j [(1-p) \delta_{ij} + p R_{ji}] n_j} \quad (7.4)$$

is the fraction of the population which is currently in patch i . Assuming that all individuals have the same contact patterns, interacting uniformly with everyone in the patches they visit, the temporal evolution of the fraction of infected individuals resident in i in a SIS dynamics will be given by

$$\rho_i(t+1) = (1 - \bar{\mu}) \rho_i(t) + [1 - \rho_i(t)] \Pi_i(t). \quad (7.5)$$

The first term in the r.h.s. of Eq. (7.5) corresponds to the probability that individuals do not recover, which happens with probability $1 - \bar{\mu}$ for each individual resident in i . The second term represents the probability that susceptible individuals become infected at time t that will depends on $\Pi_i(t)$, which is the probability that a healthy individual of i becomes infected at time t and is computed as

$$\Pi_i(t) = (1-p) \pi_i(t) + p \sum_{j=1}^N R_{ij} \pi_j(t), \quad (7.6)$$

where $\pi_i(t)$ is the probability that a healthy individual currently in patch i is infected at time t . Note that this probability depends on all individuals that are in i , not necessarily resident of i . In the well-mixed population approximation we can compute the complementary probability of not being infected by any of the \tilde{n}_i individuals in i , which happens with probability $\bar{\lambda} \tilde{\rho}_i(t)$,

as

$$\pi_i(t) = 1 - [1 - \bar{\lambda}\tilde{\rho}_i(t)]^{\tilde{n}_i}, \quad (7.7)$$

in which \tilde{n}_i is the total number of individuals which are in patch i and $\tilde{\rho}_i(t)$ is the fraction of infected individuals currently in i at time t . The latter quantity is given by Eq. (7.4), where ρ_i is now a function of the time t .

7.2.1 Critical regime and linearization

Proceeding to obtain the epidemic threshold, we look at the steady state regime of Eq. (7.5). That will happen when $\rho_i(t+1) - \rho_i(t) \rightarrow 0$, as $t \rightarrow \infty$ for all patches i . Close to the epidemic threshold, the fractions ρ_i are very small and represented by $\rho_i^* \equiv \epsilon_i \ll 1$. After a linearization of Eqs. (7.5) and (7.7) we have [61]

$$\bar{\mu}\epsilon_i \approx \bar{\lambda} \sum_j M_{ij}\epsilon_j, \quad (7.8)$$

where

$$M_{ij} = \left\{ (1-p)^2 \delta_{ij} + p(1-p) [R_{ji} + R_{ij}] + p^2 \sum_l R_{il}R_{jl} \right\} n_j. \quad (7.9)$$

Equation (7.8) defines an eigenvalue problem and the epidemic threshold will be given by

$$\bar{\lambda}_c = \frac{\bar{\mu}}{\Lambda_{\max}(M)}, \quad (7.10)$$

with Λ_{\max} being the leading eigenvalue of the matrix M . Equation (7.10) can be solved numerically for any metapopulation structure with demographic n_i and mobility data R_{ij} .

7.2.2 Perturbation analysis for the homogeneous mixing case

Equation (7.9) can be written as

$$M_{ij} = \left[\delta_{ij} + p (R_{ij} + R_{ji} - 2\delta_{ij}) + p^2 \left(\delta_{ij} - R_{ij} - R_{ji} + \sum_l R_{il}R_{jl} \right) \right] n_j, \quad (7.11)$$

and a perturbation analysis is made close to the static case $p = 0$ in Ref. [61] resulting in Ω eigenvalues corresponding to each patch i , given by

$$\Lambda_i(M) \approx n_i + 2pn_i(R_{ii} - 1) + p^2 \sum_{j \neq i} \frac{n_i n_j (R_{ij} + R_{ji})^2}{n_i - n_j}. \quad (7.12)$$

The first term in the r.h.s. corresponds to the eigenvalue of the unperturbed system, with $p = 0$, resulting in a threshold $\bar{\lambda}_c = \bar{\mu}n_{\max}^{-1}$, where n_{\max} is the number of individuals residing in the most populated patch. The second term is the first-order correction, and is always negative since $R_{ii} \leq 1$. So, close to $p = 0$ there is always an increase of the epidemic threshold: as people start to move, the disease will have more difficulties to spread due to homogenization of populations, reducing the difference between effective populations \tilde{n}_i . Finally, the second-order correction is always positive, resulting in a reduction of the epidemic threshold due to the compensation of recovery over contagion processes [61]. This is a counter-intuitive result explained by the existence of this trade-off between the homogenization of the population and the correlations in the mobility patterns that change the critical behavior of the system [61].

7.2.3 Different critical scenarios and the detriment effect

In Ref. [61], authors suggest a synthetic star-like (or wheel-like) metapopulation network to investigate different critical regimes. In this case, a central patch (hub) contains κ leaves. The hub h has a population of $n_h = n_{\max}$ individuals, while leaves l have a fraction $\alpha \in [0,1]$ of the hub population, $n_l = \alpha n_{\max}$. The mobility of individuals living in the hub to the leaves is uniform, given by

$$R_{hl} = \frac{1}{\kappa}, \quad (7.13)$$

while the mobility between leaves, hub and other leaves is controlled by the parameter δ . The flow from one leaf to the hub is

$$R_{lh} = \delta, \quad (7.14)$$

and between adjacent leaves happens in only one way, given by

$$R_{l,l+1} = 1 - \delta, \quad (7.15)$$

as shown in Fig. 7.2.

By considering a star-like metapopulation, three different critical regimes were found [61]. Applying the star-like structure in Eq. (7.12), an approximation for the leading eigenvalue is [61]:

$$\Lambda_{\max}(M) = n_{\max} - 2pn_{\max} + p^2\alpha\kappa \frac{\left(\delta + \frac{1}{\kappa}\right)^2}{1 - \alpha} n_{\max}. \quad (7.16)$$

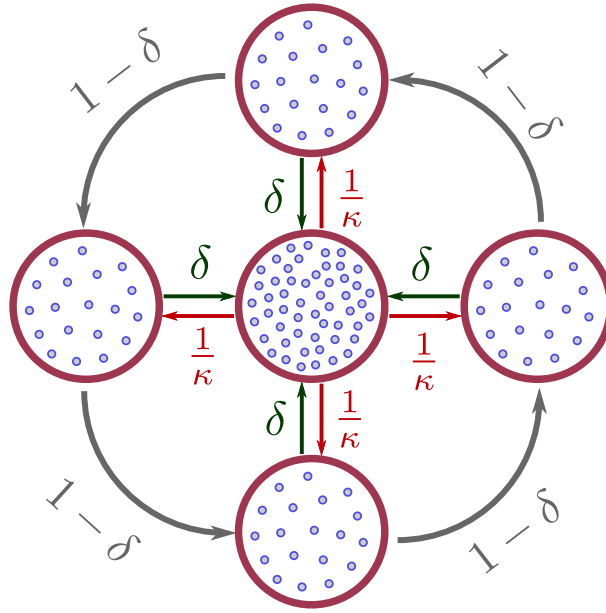


Figure 7.2: The star-like metapopulation structure consists of $\kappa + 1$ patches: a hub and its κ leaves. Flows of individuals from the hub are equal to each leaf, with probability κ^{-1} , while the flow from a leaf to the center happens with probability δ . Individuals residing in the leaves can also travel to the next adjacent leaf with probability $1 - \delta$. The hub has a population of $n_h = n_{\max}$ individuals and leaves have $n_l = \alpha n_{\max}$ with $\alpha < 1$. The individuals are represented by the smaller circles inside the patches. Based on Fig. 1(c) of [61].

Evaluating the value p^* for which the eigenvalue has an extreme value

$$\left. \frac{d\Lambda_{\max}}{dp} \right|_{p=p^*} = 0,$$

resulting in

$$p^* = \frac{1 - \alpha}{\alpha \kappa \left(\delta + \frac{1}{\kappa} \right)^2}. \quad (7.17)$$

Three different regimes are possible. The first happens if $p^* < n_{\max}^{-1}$, for which in average no movement is possible, since $n_i p^* < 1$ for all patches. This minimum will vanish in the limit $n_{\max} \gg 1$, and will result in an epidemic threshold always decreasing with p . An intermediate case will happen if $n_{\max}^{-1} < p^* < 1$, for which the maximum eigenvalue of M will have a minimum in this region and then the epidemic threshold will increase with p , reaching a maximum and then decreasing again. A non-physical probability is if $p^* > 1$, meaning that the leading eigenvalue will always decrease, therefore the epidemic threshold will increase for all values of p . These three types, denoted by I, II, and III, respectively, are schematically shown in Fig. 7.3. For further discussion see Ref. [61].

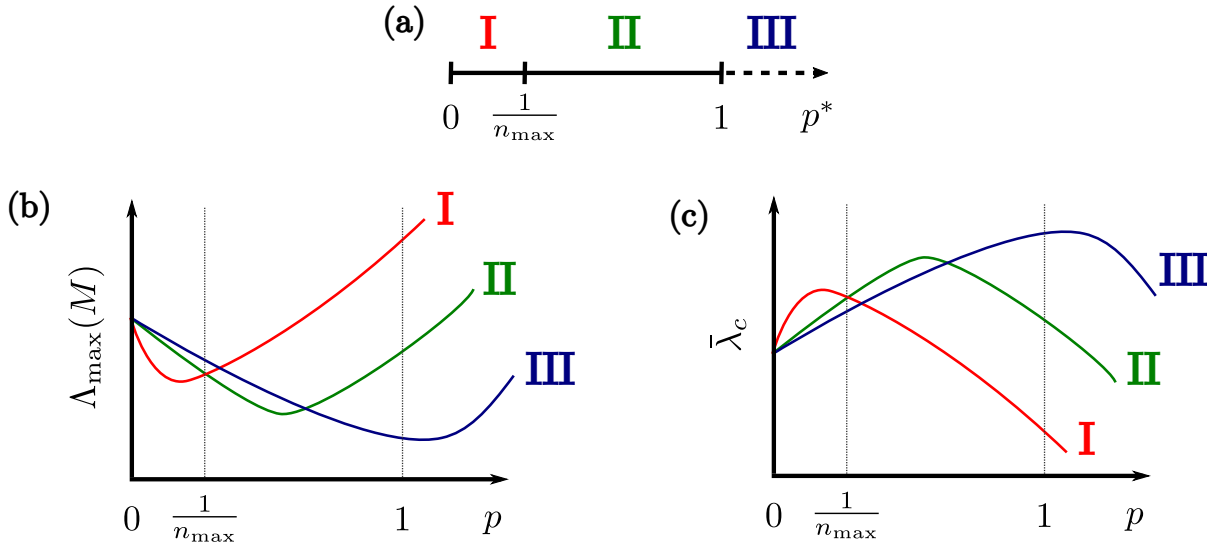


Figure 7.3: Schematic representation of three critical scenarios in networks with recurrent mobility. (a) Definition of three regions by the possible values of p^* : type I, given by $0 < p^* < n_{\max}^{-1}$, type II, $n_{\max}^{-1} < p^* < 1$, and type III, $p^* > 1$. (b) Qualitative behavior of the leading eigenvalue, and (c) of the epidemic threshold as a function of p for three different scenarios.

7.3 Markovian equations for heterogeneous subpopulations

Now we propose a modification to consider different levels of heterogeneities inside each patch. In Ref. [61], all individuals inside a patch interact with all others with the same probability following a homogeneous mixing. Here we propose a model in which individuals have a different social degree or *connectivity* k . Each patch i has $n_i^{[k]}$ individuals with connectivity k . Then, the population of patch i is

$$n_i = \sum_k n_i^{[k]} = \sum_k n_i P_i(k), \quad (7.18)$$

where $P_i(k)$ is the probability that i has an individual of connectivity k ,

$$P_i(k) = \frac{n_i^{[k]}}{n_i}. \quad (7.19)$$

We assume a heterogeneous mixing similar to the one used in the HMF theory, in an annealed regime. One hypothesis is that individuals with connectivity k will preserve this value when traveling to another patch: socially active individuals in their residence are assumed to be also very social when visiting others. After traveling, they interact with k others following the probability to connect to an individual of connectivity k' given by $P(k'|k)$. For the sake of simplicity, we will consider the lack of connectivity correlations and Eq. (2.18) holds. The total

number of individuals in the metapopulation is $N = \sum_i n_i$.

We denote the number of infected individuals residing in i that have connectivity k as $I_i^{[k]}$, implying that the total of infected individuals resident in i is $I_i = \sum_k I_i^{[k]}$. The fraction of infected individuals of patch i with connectivity k is

$$\rho_i^{[k]} = \frac{I_i^{[k]}}{n_i^{[k]}}, \quad (7.20)$$

while the fraction of infected individuals of i is $\rho_i = \sum_k P_i(k) \rho_i^{[k]}$. An equation similar to Eq. (7.2) is written as

$$n_{j \rightarrow i}^{[k]} = [(1 - p_k) \delta_{ij} + p_k R_{ji}] n_j^{[k]}, \quad (7.21)$$

that is the number of individuals with connectivity k that travel from patch j to i and the effective population of these individuals in patch i is

$$\tilde{n}_i^{[k]} = \sum_j n_{j \rightarrow i}^{[k]} = \sum_j [(1 - p_k) \delta_{ij} + p_k R_{ji}] n_j P_j(k), \quad (7.22)$$

while the total effective population of i is $\tilde{n}_i = \sum_k \tilde{n}_i^{[k]}$. Note that if p_k is constant, independent of k , we have the same Eq. (7.2) since $\sum_k P_i(k) = 1$.

Another quantity that can be evaluated is the effective connectivity distribution $\tilde{P}_i(k)$, defined as the probability of finding an individual of connectivity k currently in patch i , given by

$$\tilde{P}_i(k) = \frac{\tilde{n}_i^{[k]}}{\tilde{n}_i} = \frac{\sum_j n_{j \rightarrow i}^{[k]}}{\sum_j n_{j \rightarrow i}} = \frac{\sum_j [(1 - p_k) \delta_{ij} + p_k R_{ji}] n_j P_j(k)}{\sum_j \sum_{k'} [(1 - p_{k'}) \delta_{ij} + p_{k'} R_{ji}] n_j P_j(k')}. \quad (7.23)$$

We can define the effective moments as $\langle \widetilde{k^n} \rangle_i = \sum_k k^n \tilde{P}_i(k)$. We will show later that a simple analysis of these quantities is not able to predict the epidemic threshold as a function of p . To derive the Markovian equations we will use a similar approach to the HMF theory by computing the time evolution of the fraction of individuals with connectivity k that live in i as

$$\rho_i^{[k]}(t+1) = (1 - \bar{\mu}) \rho_i^{[k]}(t) + \left[1 - \rho_i^{[k]} \right] \Pi_i^{[k]}(t), \quad (7.24)$$

where

$$\Pi_i^{[k]}(t) = (1 - p_k) \pi_i^{[k]}(t) + p_k \sum_{j=1}^N R_{ij} \pi_j^{[k]}(t) \quad (7.25)$$

is the probability that a healthy individual with connectivity k belonging to patch i becomes infected at time t . $\pi_i^{[k]}(t)$ is the probability that an individual in patch i of connectivity k becomes infected at time t , which can be computed in terms of the probability that no infection comes from the k selected contacts using Eqs. (7.28) and (7.29) as

$$\pi_i^{[k]}(t) = 1 - \prod_{k'} \left[1 - \bar{\lambda} \tilde{P}_i(k'|k) \tilde{\rho}_i^{[k']}(t) \right]^k, \quad (7.26)$$

in which

$$\tilde{\rho}_i^{[k]} = \frac{\tilde{I}_i^{[k]}}{\tilde{n}_i^{[k]}} = \frac{\sum_j I_{j \rightarrow i}^{[k]}}{\sum_j n_{j \rightarrow i}^{[k]}} \quad (7.27)$$

is the effective fraction of infected individuals with connectivity k in patch i and $\tilde{P}_i(k'|k)$ is the effective probability that an individual with connectivity k is connected to another with connectivity k' . We consider the uncorrelated case for the sake of simplicity, which translates into

$$\tilde{P}_i(k'|k) = \frac{k' \tilde{n}_i^{[k']}}{\sum_{k''} k'' \tilde{n}_i^{[k'']}} = \frac{k' \tilde{P}_i(k')}{\langle k \rangle_i}, \quad (7.28)$$

which is the probability of selecting an edge from a vertex with connectivity k' that is in patch i ; see Eq. (2.18).

Finally, the effective fraction of infected individuals with connectivity k in patch i is given by

$$\tilde{\rho}_i^{[k]}(t) = \frac{\sum_j [(1 - p_k) \delta_{ij} + p_k R_{ji}] n_j P_j(k) \rho_j^{[k]}(t)}{\sum_j [(1 - p_k) \delta_{ij} + p_k R_{ji}] n_j P_j(k)}, \quad (7.29)$$

where the denominator is given by Eq. (7.22) and the numerator is the number of infected individuals that are in patch i . Note that even if the population of each patch n_j is the same, the terms $n_j^{[k]}$ or $P_j(k)$ do not cancel out because they depend on the connectivity distribution for each patch j , which are not necessarily the same.

In Eq. (7.26) we need the product $\tilde{P}_i(k'|k) \tilde{\rho}_i^{[k']}(t)$. Using the Eq. (7.28) and Eq. (7.27) we get

$$\tilde{P}_i(k'|k) \tilde{\rho}_i^{[k']}(t) = \frac{k'}{\tilde{n}_i \langle k \rangle_i} \sum_j I_{j \rightarrow i}^{[k']}(t),$$

resulting in

$$\mathcal{P}_i^{[k']}(t) \equiv \tilde{P}_i(k'|k) \tilde{\rho}_i^{[k']}(t) = \frac{k' \sum_j [(1 - p_{k'}) \delta_{ij} + p_{k'} R_{ji}] n_j P_j(k') \rho_j^{[k']}(t)}{\sum_{k''} k'' \sum_j [(1 - p_{k''}) \delta_{ij} + p_{k''} R_{ji}] n_j P_j(k'')}, \quad (7.30)$$

where we denoted the product as $\mathcal{P}_i^{[k']}(t)$ since it is independent of the connectivity k .

To find the steady state equation close to the epidemic threshold we will consider only terms up to first order on $\rho_i^{[k]}(t \rightarrow \infty) = \bar{\epsilon}_{ik} \ll 1$ and Eq. (7.26) becomes

$$\pi_i^{[k]}(t \rightarrow \infty) \equiv \pi_i^{*[k]} \approx 1 - \prod_{k'} [1 - \bar{\lambda} k \mathcal{P}_i^{*[k']}] \approx \bar{\lambda} k \sum_{k'} \mathcal{P}_i^{*[k']},$$

in which $\mathcal{P}_i^{*[k]} \equiv \mathcal{P}_i^{[k]}(t \rightarrow \infty)$, resulting in

$$\pi_i^{*[k]} = \frac{\bar{\lambda} k}{Q_i} \sum_{k'} k' \sum_j [(1 - p_{k'}) \delta_{ij} + p_{k'} R_{ji}] n_j P_j(k') \bar{\epsilon}_{jk'}, \quad (7.31)$$

where $Q_i \equiv \sum_k k \sum_j [(1 - p_k) \delta_{ij} + p_k R_{ji}] n_j P_j(k)$ is the effective number of edges in patch i . Note that $\sum_i Q_i = \sum_k \sum_j k P_j(k) n_j$ is the total number of edges in the system, a conserved quantity. The summation in Eq. (7.31) is basically the number of edges emanating from infected individuals currently at patch i . The steady state of the dynamics is obtained from Eq. (7.24), resulting in

$$\bar{\mu} \bar{\epsilon}_{ik} = (1 - \bar{\epsilon}_{ik}) \Pi_i^{*[k]}, \quad (7.32)$$

and now we have to linearize $\Pi_i^{*[k]}$ from Eq. (7.25), which is done by using the results of Eq. (7.31). After some algebra and neglecting terms of order $\mathcal{O}(\bar{\epsilon}_{ik}^2)$, we have

$$\Pi_i^{*[k]} = \bar{\lambda} \sum_j \sum_{k'} \bar{M}_{ik}^{jk'} \bar{\epsilon}_{jk'}, \quad (7.33)$$

where

$$\bar{M}_{ik}^{jk'} = k k' P_j(k') \left[\underbrace{(1 - p_k)(1 - p_{k'}) \frac{\delta_{ij}}{Q_i}}_{\textcircled{1}} + \underbrace{(1 - p_k) p_{k'} \frac{R_{ji}}{Q_i}}_{\textcircled{2}} + \underbrace{p_k (1 - p_{k'}) \frac{R_{ij}}{Q_j}}_{\textcircled{3}} + \underbrace{p_k p_{k'} \sum_l \frac{R_{il} R_{jl}}{Q_l}}_{\textcircled{4}} \right] n_j. \quad (7.34)$$

Substituting (7.34) in Eq. (7.32), and keeping only linear terms, since $\bar{\epsilon}_{ik} \ll 1 \forall (i, k)$, we end up with

$$\bar{\mu}\bar{\epsilon}_{ik} = \bar{\lambda} \sum_j \sum_{k'} \bar{M}_{ik}^{jk'} \bar{\epsilon}_{jk'}, \quad (7.35)$$

that defines an eigenvalue problem and an epidemic threshold

$$\bar{\lambda}_c = \frac{\bar{\mu}}{\Lambda_{\max}(M)}. \quad (7.36)$$

The elements of matrix M given by Eq. (7.34) represent four types of interactions in the metapopulation. Basically, the element $\bar{M}_{ik}^{jk'}$ represents the probability that a resident of patch i with connectivity k is in contact with another individual of patch j and connectivity k' . The first term (1) accounts for interactions of residents of the patch, that do not move. In term (2), an individual of i stays and interacts with a traveler from patch j in patch i , that arrived with probability $p_{k'}R_{ji}$. A similar event happens in term (3), in which an individual of i travels to patch j and interact there with a resident of j with probability p_kR_{ij} . Finally, in term (4), both individuals of patches i and j travel to a patch l , arriving there with probability $p_kp_{k'}R_{il}R_{jl}$.

The interpretation for the expression in Eq. (7.9) for the original model is similar, with terms of interactions between residents with probability $(1-p)^2$, of resident and travelers with $p(1-p)[R_{ij} + R_{ji}]$ and of travelers with $p^2 \sum_l R_{il}R_{jl}$. However, in that case, all individuals could interact with each other homogeneously, without a limitation on the number of contacts, so the probability M_{ij} only depended on the demography terms n_j . In Eq. (7.34), these interactions also depend on the connectivity of both individuals i and j , by terms in the form $kk'P_j(k')n_j/Q_l$, related to the number of edges available to make the connections, in which l is the patch where the interaction happens.

Before going ahead, it is useful to make the transformation $\bar{\epsilon}_{ik} \mapsto k\epsilon_{ik}$ in Eq. (7.35), resulting in

$$\bar{\mu}\epsilon_{ik} = \bar{\lambda} \sum_j \sum_{k'} M_{ik}^{jk'} \epsilon_{jk'}, \quad (7.37)$$

where the new matrix M will be given by the elements

$$M_{ik}^{jk'} = k'^2 P_j(k') \left[(1-p_k)(1-p_{k'}) \frac{\delta_{ij}}{Q_i} + (1-p_k)p_{k'} \frac{R_{ji}}{Q_i} + p_k(1-p_{k'}) \frac{R_{ij}}{Q_j} + p_k p_{k'} \sum_l \frac{R_{il}R_{jl}}{Q_l} \right] n_j. \quad (7.38)$$

If the movement is independent of the connectivity k , i.e., $p_k = p$, Eq. (7.37) becomes

$$\bar{\mu}\epsilon_i = \bar{\lambda} \sum_j M_{ij}\epsilon_j, \quad (7.39)$$

in which the elements ϵ_i of the eigenvectors are now independent of k . A new matrix M is defined, whose elements are

$$M_{ij} = \langle k^2 \rangle_j \left[(1-p)^2 \frac{\delta_{ij}}{Q_i} + p(1-p) \left(\frac{R_{ji}}{Q_i} + \frac{R_{ij}}{Q_j} \right) + p^2 \sum_l \frac{R_{il}R_{jl}}{Q_l} \right] n_j, \quad (7.40)$$

where the effective number of edges Q_i is

$$Q_i = \sum_j \langle k \rangle_j \left[(1-p)\delta_{ij} + pR_{ji} \right] n_j. \quad (7.41)$$

In the following sections we will deal only with this case of movement independent of k .

To check the consistency of these equations, let us consider the static case in which all individuals stay in their patches and do not move: $p_k = 0 \forall k$. So, Eq. (7.40) becomes

$$M_{ij}|_{p_k=0} = \langle k^2 \rangle_j \frac{\delta_{ij}}{Q_i|_{p_k=0}} n_j,$$

where $Q_i|_{p_k=0} = n_i \langle k \rangle_i$, that after being used in Eq. (7.39) results in $\bar{\mu}\epsilon_i = \bar{\lambda} \frac{\langle k^2 \rangle_i}{\langle k \rangle_i} \epsilon_i$. This case consists of isolated subpopulations in an annealed regime in which the epidemic threshold will be given by the first subpopulation in the active state, if its population is not so small compared to other patches. Indeed, the usual epidemic threshold known in the HMF theory is obtained,

$$\bar{\lambda}_c = \bar{\mu} \frac{\langle k \rangle_i}{\langle k^2 \rangle_i} \Big|_{i=\max}, \quad (7.42)$$

where $i = \max$ refers to the patch with the maximum value of $\frac{\langle k^2 \rangle_i}{\langle k \rangle_i}$, responsible for the leading eigenvalue of M when n_{\max} represents a significant fraction of the total population.

7.3.1 Synthetic structures with heterogeneous subpopulations

To investigate the effects of the heterogeneities in connectivity on different synthetic networks, comparing exact and numerical results for the epidemic thresholds with approximations given by effective connectivity distribution and the lower bound, we define two kinds of net-

works. The first one is a metapopulation of only two patches, to isolate the effect of different levels of heterogeneities between two single subpopulations. Next, we define a star-like network such as the one presented in Sec. 7.2.3 but including contact patterns on the individuals living in each patch.

Metapopulation of two patches

This synthetic metapopulation consists of $\Omega = 2$ patches, in which the fluxes are symmetrical: $R_{12} = R_{21} = 1$, and $R_{ii} = 0$ for $i = 1, 2$. We can consider that the first patch is the most heterogeneous one, leading the epidemic threshold of the system, without loss of generality. As an example, we can consider that this first patch contains a fraction α of individuals of connectivity 1, and $1 - \alpha$ of connectivity k_{\max} , while the second has only individuals of connectivity $k_2 = \langle k \rangle_1$, predetermined. So, the connectivity distribution is bimodal and given by $P_1(k) = \alpha \delta_{k1} + (1 - \alpha) \delta_{kk_{\max}}$, with $\alpha = \frac{k_{\max} - \langle k \rangle_1}{k_{\max} - 1}$ and $P_2(k) = \delta_{k\langle k \rangle_1}$. Figure 7.4 shows a schematic example of this connectivity configuration for $k_{\max} = 5$.

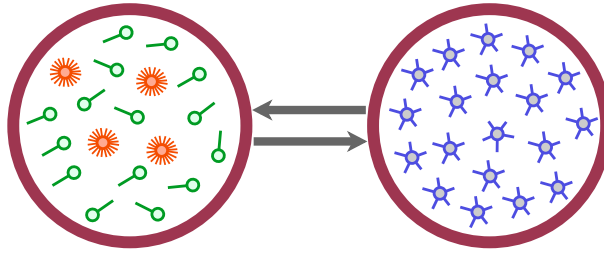


Figure 7.4: Example of a metapopulation with two patches. The first is a heterogeneous patch with individuals of connectivity 1 with probability α , or $k_{\max} = 20$ with complementary probability $1 - \alpha$, and the second is a homogeneous patch with all individuals with the same connectivity $k_2 = \langle k \rangle_1 = 5$. Individuals can travel from one patch to the other with probability p , always having the same destination.

Star-like metapopulation

In this metapopulation network, a hub has κ leaves, whose populations are n_h and n_l for the hub and leaves, respectively. The mobility from leaves to the hub, and between leaves depends on the parameter δ , and the elements R_{ij} of the mobility matrix are the same defined in Eqs. (7.13) to (7.15). Figure 7.5 shows an example of this metapopulation structure with heterogeneous individuals in the hub, while individuals have the same connectivity in the leaves.

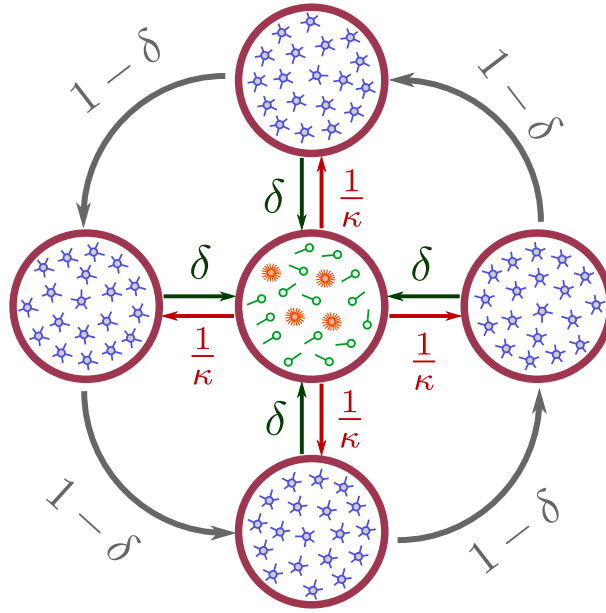


Figure 7.5: Example of a star-like metapopulation network with $\kappa + 1$ patches. In this example, the center and leaves have the same number of individuals, $n_h = n_l = n$, while the hub is a heterogeneous patch with individuals of connectivity 1 with probability α , or $k_{\max} = 20$ with complementary probability $1 - \alpha$, and each leaf is a homogeneous patch with individuals of same connectivity $k_l = \langle k \rangle_h = 5$. The flow from hub to a leaf happens with probability $R_{hl} = \kappa^{-1}$, from leaves to hub with $R_{lh} = \delta$, and between adjacent leaves with $R_{l,l+1} = 1 - \delta$.

7.3.2 Agreement with Monte Carlo simulations

Before proceeding, it is important to check the validity of the Markovian equations. To this aim, we define a Monte Carlo (MC) algorithm for the stochastic simulation of the SIS model on top of the metapopulation with heterogeneous contact patterns. We expect that the average quantities agree with the Markovian equations for the order parameters, such as the fraction of infected individuals. The proposed process is also a discrete-time dynamics. In each time step, each individual is tested to move with probability p and, if accepted, it moves to a patch j with probability R_{ij} . Then, all individuals with connectivity k choose k individuals in the patch they currently are at random, following $\tilde{P}(k'|k)$ and are infected with probability $\bar{\lambda}$ if the contacted individual is infected. Healing happens with probability $\bar{\mu}$ for each infected individual. The simulation procedure is the following:

1. A fraction ρ_{ini} of the population is infected as the initial condition.
2. For each time step $\Delta t = 1$ and each patch i , each individual resident in i is tested to move with probability p , choosing a patch j proportionally to R_{ij} .
3. Each infected individual heals with probability $\bar{\mu}$.

4. Each susceptible individual with connectivity k selects at random at maximum k contacts in patch i . For each attempt, it can be infected with probability

$$\bar{\lambda} \frac{\sum_k k \tilde{\rho}_i^{[k]}(t) \tilde{n}_i^{[k]}}{\sum_k k \tilde{n}_i^{[k]}},$$

or remain susceptible with complementary probability. The attempts stop if the individual becomes infected.

5. All individuals return to their residences and the time is updated from t to $t + 1$.

In step 4, the connectivity of the susceptible individual is k , and it selects k edges at random. If the selected edge comes from an infected individual, it becomes infected with probability $\bar{\lambda}$. The infection is proportional to the number of edges coming from infected individuals *currently* in patch i .

To avoid the absorbing-state, we infect a small fraction ρ_{pump} of individuals at random when this state is reached [95]. This keeps the dynamics always active. The comparison between MC and Markovian equations was done for the steady-state considering two types of metapopulations, in Figs. 7.6 and 7.7. From now on, we assume that all patches contain the same number of individuals $n_i = n$ to isolate the effect of contact heterogeneity.

Comparing the MC simulations with the temporal evolution of the Markovian equations, we have an excellent agreement. The simulations were done by starting with all individuals infected, $\rho_{\text{pump}} = 2 \times 10^{-4}$ and 10 dynamical samples in a metapopulation of one patch with

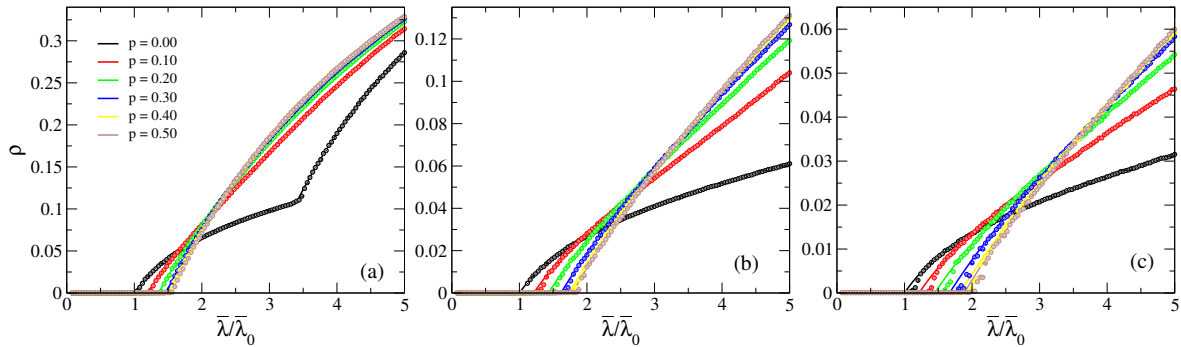


Figure 7.6: Equilibrium regimes of the Markovian equations (lines) and MC simulations (symbols) for a two-patches metapopulation with $n_1 = n_2 = n = 10^4$. One patch follows a bimodal connectivity distribution with values 1 and (a) $k_{\text{max}} = 20$, (b) 50, and (c) 100, while in the second all individuals have the same connectivity $k_2 = \langle k_1 \rangle = 5$. A fraction $\rho_{\text{pump}} = 2 \times 10^{-4}$ was used in 10 dynamical samples.

bimodal distribution with $k_{\max} = 20, 50$, and 100 , and another patch with $k_2 = \langle k \rangle_1 = 5$. The results are shown in Fig. 7.6. As p increases, the heterogeneities of the patches become smaller due to the homogenization of the population, increasing the epidemic threshold. The highest epidemic threshold happens for $p = 0.5$ in all cases, and decreases symmetrically for $p > 0.5$. It is interesting to note that in the curve for $p = 0$ in Fig. 7.6 we see an inflection at $\bar{\lambda}/\bar{\lambda}_0 \approx 3.4$ that corresponds to the activation of the most homogeneous patch.

We also perform stochastic simulations for the star-like metapopulation by considering different levels of mobility patterns. In Fig. 7.7 we compare the results from the Markovian equations with MC simulations for $n_i = 10^4$ individuals with $\kappa = 10$ leaves, but different levels of contact heterogeneity. First, we consider a heterogeneity only between patches, in which the hub has individuals with the same connectivity $k_h = 100$, while leaves have $k_l = 10$. We observe an increase of the epidemic threshold when individuals start to move. The agreement with MC simulations is also remarkable.

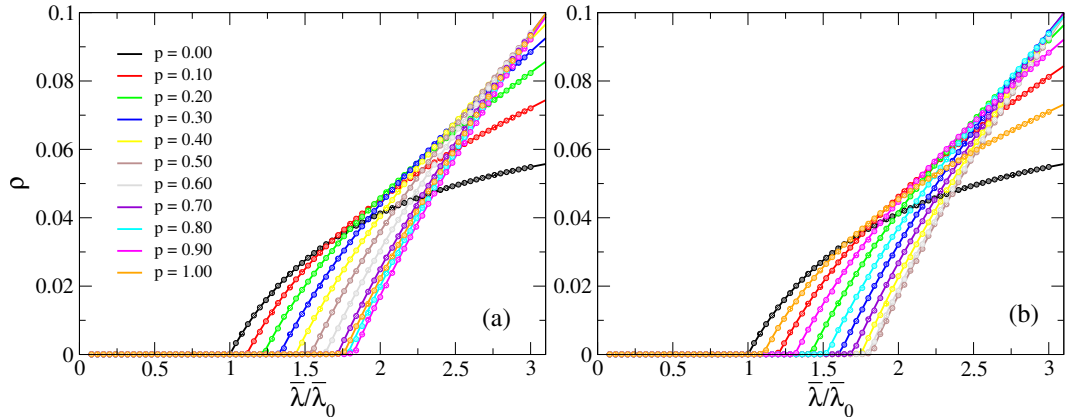


Figure 7.7: Equilibrium regimes of the Markovian equations (lines) and MC simulations (symbols) for a star-like metapopulation with $n_h = n_l = n = 10^4$ and $\kappa = 10$ leaves. The hub contains individuals with connectivity $k_h = 100$, and the leaves $k_l = 10$. The mobility pattern is given by (a) $\delta = 0.1$ and (b) 0.9 . A fraction $\rho_{\text{pump}} = 2 \times 10^{-4}$ was used in 10 dynamical samples.

7.3.3 Analytical evaluation of the epidemic threshold

We now evaluate analytically the epidemic threshold as a function of the mobility parameter p . Using the Gershgorin circle theorem [205] it is possible to show that an upper bound for the eigenvalue of the matrix M is given by

$$\Lambda_{\max} \leq \max_i \left(\sum_j M_{ij} \right). \quad (7.43)$$

See Appendix C.1 for more details. With that we can get a lower bound for the epidemic threshold described in Eq. (7.40) as

$$\bar{\lambda}_c \geq \frac{\bar{\mu}}{\max_i \sum_j \langle k^2 \rangle_j \left[(1-p)^2 \frac{\delta_{ij}}{Q_i} + p(1-p) \left(\frac{R_{ji}}{Q_i} + \frac{R_{ij}}{Q_j} \right) + p^2 \sum_l \frac{R_{il}R_{jl}}{Q_l} \right] n_j}. \quad (7.44)$$

In Fig. 7.8 we compare the numerical threshold obtained by the leading eigenvalue of Eq. (7.40) and the r.h.s. of Eq. (7.44), confirming the lower bound for the epidemic threshold. The absolute relative deviation can be high for intermediate values of p and depends on the structure of the mobility network.

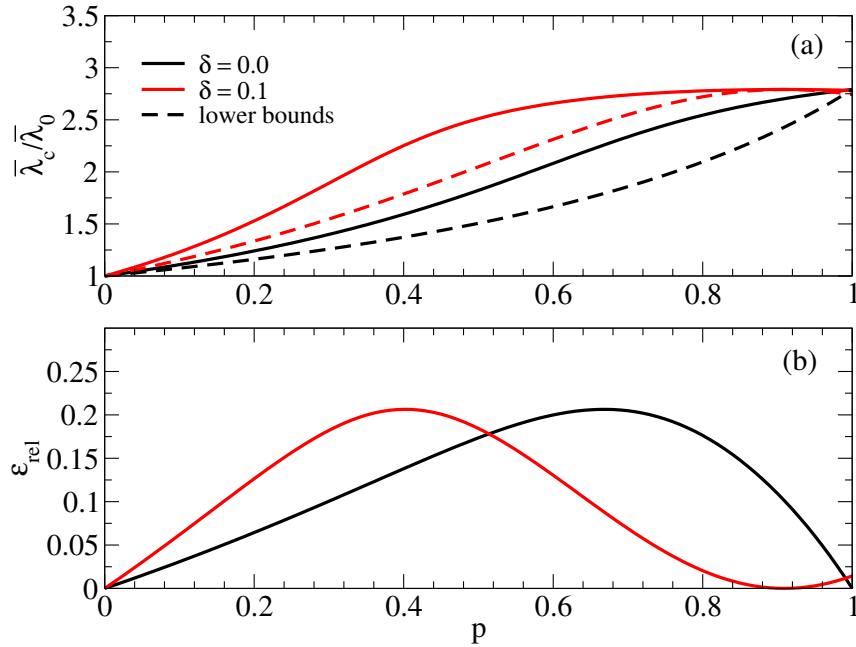


Figure 7.8: Comparison between the numerical estimation of the threshold via calculation of the leading eigenvalue of the matrix M_{ij} defined by Eq. (7.40) and the estimate given by Eq. (7.44). The relative deviation ϵ_{rel} is also calculated. The metapopulation consists of a star structure of $\kappa = 10$ leaves and $n_i = n = 10^4$ individuals in each patch. The parameter δ , shown in the legends, tune the mobility between patches; see Sec. 7.3.1. The center has individuals with connectivity either 1 or 20, and individuals residing in the leaves have connectivity $k_2 = \langle k \rangle_1 = 5$.

An approximation for the epidemic threshold would be to use the effective connectivity distribution to evaluate

$$\bar{\lambda}_c \approx \bar{\mu} \min_i \left[\frac{\widetilde{\langle k \rangle}_i}{\langle k^2 \rangle_i} \right], \quad (7.45)$$

where i is the patch in which this value is minimal, corresponding to the one that will be activated first. This expression describe only qualitatively the epidemic threshold, since we are neglecting dynamical correlations.

We investigate the precision of the epidemic thresholds obtained via effective connectivity distribution (dotted lines), the upper bound limit (dashed lines), exact result (solid lines) and numerical threshold (points) obtained by finding the leading eigenvalue of the matrix M in Fig. 7.9. We start by the metapopulation with two patches. The exact result for the leading eigenvalue as a function of p is detailed in Appendix C.3. The maximum value happens at $p^* = 0.5$ and is equal to

$$\frac{\Lambda_{\max}(p = 0.5)}{\Lambda_1^{(0)}} = \frac{a + 1}{2}, \quad (7.46)$$

where $a = \langle k^2 \rangle_2 / \langle k^2 \rangle_1 < 1$. Note that as $\langle k^2 \rangle_1$ increases with k_{\max} , $a \rightarrow 0$, this expression

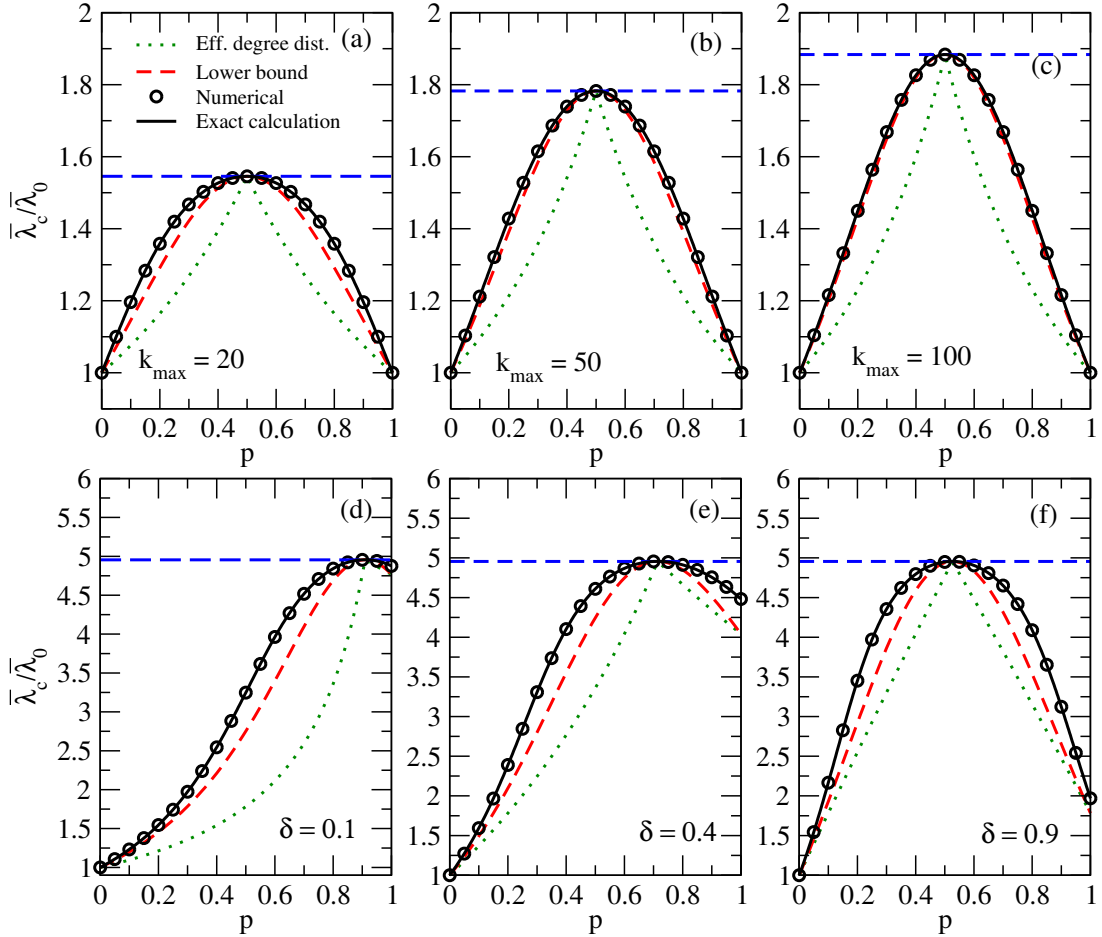


Figure 7.9: Comparison between different approximations, analytical, and numerical results for (a-c) two patches metapopulation with the same parameters of Fig. 7.6, with (a) $k_{\max} = 20$, (b) 50, and (c) 100, and (d-f) star-like metapopulation where the hub contains individuals following a bimodal distribution of connectivity with $k_{\max} = 50$, the $\kappa = 10$ leaves have individuals with connectivity $k_l = \langle k_h \rangle = 5$, and (d) $\delta = 0.1$, (e) 0.4, and (f) 0.9. All patches have $n_i = n = 10^5$ individuals. Results for the epidemic threshold obtained by the effective connectivity distribution (dotted lines), lower bound limit (dashed lines), exact expressions (solid lines), and via numerical evaluation of the leading eigenvalue of M (points) are shown. The horizontal dashed line indicates the maximum value predicted in (a-c) Appendix C.3 and (d-f) Appendix C.4, agreeing with the numerical evaluation of the leading eigenvalue of Eq. (7.40).

converges to $1/2$ so that the highest epidemic threshold happens at $\bar{\lambda}/\bar{\lambda}_0 = 2$. This analysis is in consonance with the results shown in Fig. 7.9(a-c). We can also notice that in this case the lower bound estimate given by Eq. (7.44) follows very well the behavior of the actual epidemic threshold, while the effective connectivity distribution by Eq. (7.45) only agrees on the extremes $p = 0, 1$, and in the maximum $p^* = 0.5$.

We also present a similar analysis in star-like metapopulations in Fig. 7.9(d-f) for a fixed level of heterogeneity and varying mobility patterns. The exact result for the epidemic threshold in this structure can also be evaluated and details are given in Appendix C.4. Again, exact and numerical results agree with each other, and the lower bound, Eq. (7.44), gives a good approximation for the actual epidemic threshold, while the effective connectivity distribution only agrees for $p = 0, 1$, and at $p = p^*$. It is interesting to observe that the epidemic threshold at $p = p^*$ depends on the level of heterogeneity, but do not on the mobility patterns. These maxima shift from right to left as δ increases without changing their values.

Perturbation analysis and first order correction

We proceed by making a perturbation analysis of the eigenvalues of the matrix M up to first order to quantify the effect of heterogeneity when mobility is small, with $p \ll 1$. We write this expression as a polynomial up to second order in p in Eq. (7.40), to obtain

$$M_{ij} = \langle k^2 \rangle_j \left\{ \frac{\delta_{ij}}{Q_i} + p \left[\frac{R_{ji}}{Q_i} + \frac{R_{ij}}{Q_j} - 2 \frac{\delta_{ij}}{Q_i} \right] + p^2 \left[\frac{\delta_{ij}}{Q_i} - \frac{R_{ji}}{Q_i} - \frac{R_{ij}}{Q_j} + \sum_l \frac{R_{il}R_{jl}}{Q_l} \right] \right\} n_j, \quad (7.47)$$

that has terms similar to those in Eq. (7.11). With Q_i being also a function of p , we perform a Taylor expansion up to terms of order $O(p^2)$, Appendix C.2, to obtain

$$\frac{\Lambda_i}{\Lambda_i^{(0)}} \approx 1 + \left[R_{ii} - 1 - \sum_{j \neq i} R_{ji} \frac{n_j \langle k \rangle_j}{n_i \langle k \rangle_i} \right] p = 1 + \Lambda_i^{(1)} p, \quad (7.48)$$

with the first order correction. Note that the coefficient of p in this expression is always negative since $R_{ii} < 1$ and the summation is positive, resulting in the detriment of the epidemic threshold, as in the original model with homogeneous populations defined in Sec. 7.2. Another interesting result is that $\Lambda_i^{(1)}/\Lambda_i^{(0)}$ does not depend on the second moment of the connectivity distribution, but only on the ratio n_j/n_i and of average connectivity $\langle k \rangle_j/\langle k \rangle_i$ of different patches i and j .

In Fig. 7.10, we check the accuracy of Eq. (7.48) in star-like metapopulations for different

parameters δ and κ , in which individuals of the hub have connectivity 1 or $k_{\max} = 50$ and leaves have $k_l = \langle k \rangle_1 = 5$. At first order in p , the situation is always detrimental due to the homogenization of populations. Here, since all patches have the same number of individuals, this change is due to differences between the average connectivity of patches. First, by fixing the number of leaves as $\kappa = 10$ and varying the mobility δ from leaves to hubs, we see that the detrimental effect increases proportionally to δ and κ , as predicted by Eq. (7.48). In all cases, individuals residing in the hub always travel to the leaves irrespective of the values of δ . For small values of δ , individuals living in leaves do not change the effective connectivity distribution of the hub, since the movement mostly happens from hub to leaves. As δ increases, more individuals start to move to the hub, increasing the homogenization of the population, changing the contagion processes. But one can see that the maximum value does not depend on δ , as already seen in Figs. 7.9(d-f), since it will only change for which value of p the homogenization will happen. Similarly, the greater the number of leaves κ , the greater the homogenization effect. As the number of individuals in all patches are equal, there are κn_i individuals to mix with those residing in the hub. When mobility starts to work, leaving $p = 0$, several individuals of small connectivity $k_l = \langle k_h \rangle = 5$ cause the effective heterogeneity to drastically drop since the number of individuals with connectivity k_{\max} is small and distributed for more patches.

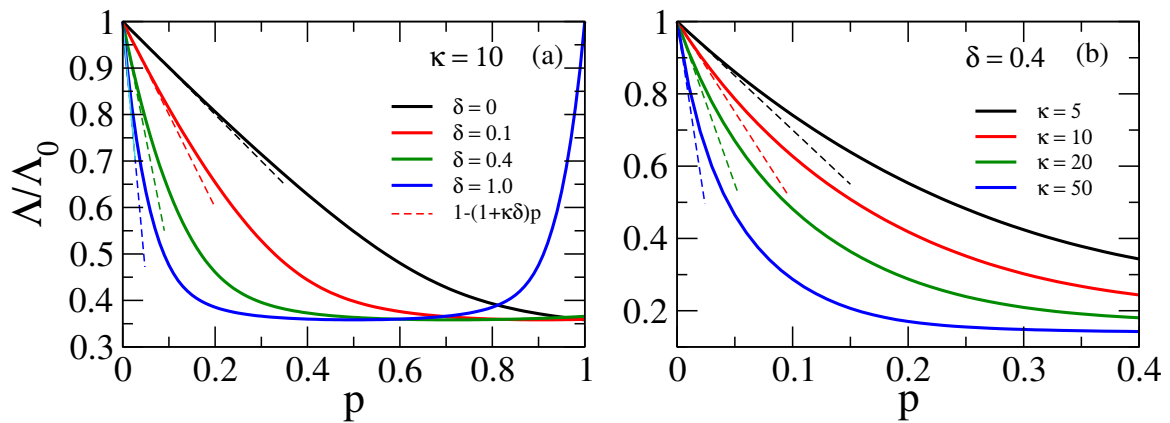


Figure 7.10: Comparison of the first order approximation in the perturbation analysis, Eq. (7.47), and numerical evaluation of the leading eigenvalue of the matrix M for star-like metapopulation structures. In both cases, individuals in the hub have connectivity 1 or $k_{\max} = 50$, while leaves have connectivity $k_l = \langle k_h \rangle = 5$. In (a) there are $\kappa = 10$ leaves and different values of δ , indicated in the legend, while in (b) the parameter $\delta = 0.4$ is fixed and the number of leaves κ varies, indicated in the legend. Dashed lines are given by $1 - (1 + \kappa\delta)p$, from Eq. (7.47).

7.4 Summary and perspectives

With the advance of data mining techniques in mobility and social patterns [188, 206–208], from individual to whole population level, the representation through metapopulations has several challenges to increase their degree of realism [209]. The new framework proposed in Ref. [61] has been continuously improved to incorporate more ingredients, such as high-order mobility patterns [191], different temporal scales of travels [210] and interactions between humans and vectors of diseases [211, 212].

In this work we adapted the model of Ref. [61] to account for the heterogeneity in the number of contacts made by individuals, that is a well-known property in many social networks. We described a complete set of Markovian equations for a discrete-time SIS dynamics on subpopulations with recurrent mobility. Performing a stability analysis close to the epidemic threshold we evaluated analytically its dependence on demographic, mobility, and social data. The Markovian equations were validated by extensive stochastic simulations. To investigate the role of heterogeneities, we defined two synthetic metapopulation networks and calculated the exact expressions for the epidemic thresholds. These results were compared with analytical approximations and perturbation analysis.

The first investigated case was a metapopulation with only two patches with different levels of heterogeneity and mobility between them. Starting with $p = 0$, the epidemic threshold increases up to a maximum at $p^* = 0.5$, and then decreases symmetrically, with the same epidemic threshold for $p = 0$ and 1. As shown by an exact expression, the maximum value of the epidemic threshold converges to a constant value. Another structure considers a hub with κ leaves, with mobility tuned by a parameter δ . Considering the same heterogeneity levels, the maximum epidemic threshold happens for smaller values of p^* as the mobility from leaves to hub increases, accelerating the homogenization process. However, the maximum value of the epidemic threshold does not change, depending only on the different levels of heterogeneity.

We did not find configurations for which the epidemic threshold for $p > 0$ would be smaller than the original threshold when compared to the static case $p = 0$, corresponding to type I of Ref. [61]. Future works will be made analyzing the effects of a k -dependent mobility, including correlations in mobility patterns and breaking the homogenization mechanism, since some individuals will have a higher probability to travel between patches.

Although we only analyzed synthetic metapopulation networks, this model can be inves-

tigated using a data-driven approach in the sense that one can easily accommodate real data of demographics, mobility, and contact patterns to describe more realistic situations. Indeed, the model proposed in Ref. [61] was also adapted to study the impact of mobility and social distancing interventions for the spreading of COVID-19 in Spain [62, 64]. Another case was the model proposed in Ref. [63] to measure the outbreak diversity of epidemic waves in different geographical scales in Brazil. However, these models still assumed a homogeneous mixing instead of the heterogeneous one described here, and the effect of contact heterogeneity deserves to be investigated in these realistic situations, such as in super-spreading events [213].

Part IV

Concluding remarks

Chapter 8

Summary and prospects

Complex networks can describe different natural systems while providing an adequate mathematical representation to deal with problems of many interacting elements. Using ideas and basic physical principles such as local rules of interactions between agents, empirical results can be reproduced both by models of networks and of dynamical processes. By transforming problems that might have previously been considered impossible to solve, it is possible to find similarities between these systems through the graph theory formalism, with universal behaviors that seem to be independent of which systems we are describing.

In particular, basic epidemic models can lead to absorbing-state phase transitions that strongly depend on the network structure. Analytical mean-field approximations such as heterogeneous and quenched mean-field theories are used and their results must be compared to statistical computational results from simulations. Since real networks are very large and heterogeneous, in Sec. 3.3.2 and Ref. [34] we presented an optimized Gillespie algorithm (OGA) for the simulation of generic epidemic models. Indeed, our algorithm was recently modified to be even more efficient using a composition and rejection algorithm [214].

We also performed an investigation about some specific properties of the susceptible-infected-susceptible (SIS) model on large networks [35]. First, in Chapter 4, we proposed two modifications of the standard SIS model preserving its central properties of spontaneous healing and infection capacity increasing unlimitedly with the degree. We then found that even small modifications in the local rules lead to conflicting results among them. While in the standard SIS model we concluded that localized regions of the network were always responsible for the epidemic activation on PL networks, it could not be the case for the modified dynamics. Specifically, our results suggest that the metastable, localized, and active states of the standard

SIS model, which can sustain the endemic state for $\gamma > 3$, are not universal. Our results relight the discussion of the choice of suitable theoretical approaches and the conception of epidemic modeling to describe real systems.

In Chapter 5 we tested the validity of an existent hypothesis that both finite dimension and hierarchy are necessary for the emergence of Griffiths phases (GPs). First, we investigated the dynamical behavior of the SIS model evolving on independent heterogeneous random networks [43], showing that Griffiths's effects were possible and explained by an optimal fluctuation theory. However, these effects were wiped out in the thermodynamic limit. Motivated by the fact that many real networks have a modular structure, such as the brain, and that several of them can be non-hierarchical and can have infinite dimension, we also investigated the SIS model on single weakly connected modular networks. In this case, we observed that loosely coupled modules act as effective rare regions slowing down the inactivation and leading to GPs [44]. As a consequence, we find extended regions of control parameter with continuously changing dynamical critical exponents for single network realizations in the thermodynamic limit, as in a real GP, while the avalanche size distributions of spreading events exhibit robust PL tails. Our findings relax the requirements of finite dimension and hierarchy, which can help to rationalize the criticality of modular systems in the framework of GPs.

We also focused on different applications of spreading dynamics on complex networks. Instead of using only an undirected and unweighted representation of networks, as done in the chapters in Part II, we considered directed and weighted networks in Part III.

In Chapter 6 we made an extensive study of the structural and dynamical properties of a political communication network extracted from real data in Twitter, proposing a method to quantify echo chamber effects on these networks [49]. Mining 12 million Twitter messages, we reconstructed a network in which users interchanged opinions related to the impeachment of the former Brazilian President Dilma Rousseff and observed the presence of echo chambers in the strongly connected component of the network, reflected in two well-separated communities of similar sizes with opposite views of the impeachment process. We found strong correlations between the political positions of users and their spreading capacity by using a spreading model such as the SIS and SIR models. Users expressing pro-impeachment sentiments were capable to transmit information, on average, to a larger audience than users expressing anti-impeachment sentiments, breaking these echo chambers. This method was recently used [215] to investigate a theoretical model of echo chambers and polarization dynamics on populations.

Finally, in Chapter 7 we proposed a modification of an existing model for epidemic spreading [61] on metapopulations with recurrent mobility to accommodate social interaction data. By using synthetic networks, we made an analytical study of the effects of demography, recurrent mobility, and social patterns on the epidemic threshold of the SIS dynamics on top of these networks. In all investigated cases, we did not find any configuration for which the epidemic threshold would be smaller than in situations in which the mobility between patches does not happen. This can have important impacts on general metapopulations models when applied to real situations. In fact, the model presented here can be easily adapted to handle specific situations, in a data-driven fashion, such as in the study of super-spreading events by introducing different values of infection probabilities $\bar{\lambda}_i$, depending on the patches, and by mobility probabilities dependent on the connectivity of individuals.

We expect that the results of this thesis will bifurcate in other contributions. Indeed, other works have been done with our contributions. For example, the accuracy of mean-fields theories was recently compared on correlated real networks in Ref. [73] using spectral analysis of the leading eigenvalues of the adjacency matrix. Within more practical applications, we contributed to the modeling of COVID-19 considering recurrent mobility in Spain [62, 64], with the data-driven Markov-chain approach presented in Chapter 7, and with stochastic simulations for the situation in Brazil [63], provided by the algorithm [34] proposed in Chapter 3. In the case of Spain, the formalism presented in Ref. [61] was adapted to consider specific compartments, such as exposed, presymptomatic, hospitalized, recovered, and deceased individuals, and age-dependent parameters to estimate the required number of intensive care units. Time-dependent rules of mobility and confinement were also introduced to evaluate the effect of non-pharmaceutical interventions. The same model was used to estimate the impact of interventions in Bogotá and Medellín, in Colombia. For Brazil, we also adapted the model to have exposed, presymptomatic, and recovered compartments, but we focused on the spatio-temporal spreading through distinct geographical scales, observing a great dispersion of incidence curves, meaning that the disease spreads differently over regions due to the heterogeneity of the country. For further discussion, see the articles published in Physical Review Research [63] and Physical Review X [64]. Both works contributed to the guidance of public authorities in Spain, Colombia, and Brazil to mitigate the impacts of the COVID-19 pandemic, being just a few examples of how the mathematical and simulation modeling of spreading processes can help to combat and to control real epidemics, going beyond theoretical interests.

Appendix A

Computer implementations of SIS epidemic models

To build the computer implementations used in Chapter 4, all involved rates are reckoned using statistically exact prescriptions based on the Gillespie algorithms [84]. We consider phantom processes that do nothing but counting for time increments. Below we present recipes for the models investigated in the present work.

A.1 SIS-S

The SIS-S dynamics in a network of size N with infection and healing rates λ and μ can be efficiently simulated as follows. A list with all infected vertices, their number N_{inf} , and the number of edges N_e emanating from them are recorded and constantly updated. Each time step involves the following procedures. (i) With probability

$$p = \frac{\mu N_{\text{inf}}}{\mu N_{\text{inf}} + \lambda N_e}, \quad (\text{A.1})$$

an infected vertex is selected with an equal chance and healed. (ii) With complementary probability $1 - p$, an infected vertex is selected with probability proportional to its degree. A neighbor of the selected vertex is chosen with an equal chance and, if susceptible, it becomes infected. Otherwise, no change of state is implemented (it is a phantom process). (iii) The time is incremented by

$$\tau = \frac{-\ln(u)}{\mu N_{\text{inf}} + \lambda N_e}, \quad (\text{A.2})$$

where u is a pseudo-random number uniformly distributed in the interval $(0,1)$ and the simulation runs to the next step.

A.2 SIS- \mathcal{A}

This model implementation is very similar to the contact process [25]. A list with the infected vertices and their number N_{inf} is built and constantly updated. At each time step, the rules are the following. (i) With probability

$$p = \frac{\mu}{\mu + \lambda}, \quad (\text{A.3})$$

an infected vertex is randomly chosen and healed. (ii) With complementary probability $1 - p$, all susceptible neighbors of a randomly chosen infected vertex are infected at once. (iii) The time is incremented by

$$\tau = \frac{-\ln(u)}{(\mu + \lambda)N_{\text{inf}}}. \quad (\text{A.4})$$

A.3 SIS- \mathcal{T}

As in SIS- \mathcal{S} and \mathcal{A} , a list containing the infected vertices and their number N_{inf} is built and constantly updated. We have also to maintain an auxiliary list including the number of infected neighbors n_i of each vertex i and the total number of susceptible vertices N_{SI} that have at least one infected neighbor. At each time step, the rules are the following. (i) With probability

$$p = \frac{\mu N_{\text{inf}}}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}, \quad (\text{A.5})$$

an infected vertex is selected with an equal chance and healed. (ii) With complementary probability $1 - p$, an infected vertex is selected with probability proportional to its degree and one of its neighbors is randomly chosen. If the selected neighbor i is susceptible it is accepted and infected with probability $1/n_i$. The procedure of choosing a susceptible vertex is repeated until one of them is found. (iii) The time is incremented by

$$\tau = \frac{-\ln(u)}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}. \quad (\text{A.6})$$

A.4 Simulation on uncorrelated annealed networks

On uncorrelated annealed networks, the unique difference in SIS- \mathcal{S} and SIS- \mathcal{A} with respect to the quenched case is that the choice of the neighbors to be infected is done by selecting any vertex of the network with probability proportional to its degree.

For SIS- \mathcal{T} , however, the algorithm becomes trickier and, consequently, very slow. The probability that a susceptible vertex j becomes infected is given by

$$P_j = 1 - (1 - \Theta)^{k_j}, \quad (\text{A.7})$$

where k_j is the degree of vertex j and $\Theta = N_e/(N\langle k \rangle)$ is the probability that a randomly selected neighbor (at the other side of the edge) is infected in the annealed network. Let us define a total rate that one tries to infect a susceptible vertex as $L = \lambda(N - N_{\text{inf}})$, which is larger than the real one since only the susceptible vertices that have at least one infected neighbor can actually be infected and this happens with probability $P_j < 1$. The total rate that a vertex is healed is $M = \mu N_{\text{inf}}$. The algorithm becomes the following. (i) An infected vertex is randomly chosen and healed with probability $p = M/(L + M)$. (ii) With probability $1 - p$, a susceptible vertex is randomly chosen and infected with probability P_j . (iii) The time is incremented by $\tau = -\ln(u)/(L + M)$.

The exactness of these algorithms is confirmed in Fig. 4.2 where simulations on annealed networks are compared with the integration of the HMF equations.

Appendix B

Supplementary tables and results for Chapter 6

In this Appendix, we show supplemental information for Chapter 6. First, in Secs. B.1 and B.2, the full list of hashtags obtained by the manual classification described in Sec. 6.2.1 is shown for pro-, anti-impeachment, neutral and divergent sentiments. In Secs. B.3, B.4 and B.5, we present a series of robustness analysis of structure and dynamics on these networks. In Sec. B.3 some results obtained for the 20-neutral network are presented to another network reconstructed considering the 52 divergent hashtags as neutral, defining the 72-neutral network, and some of its structural properties are compared to the ones of 20-neutral network. In Sec. B.4, we show that conclusions drawn from the SIS dynamics are similar when we consider a SIR-like dynamics instead. Finally, in Sec. B.5, we show results for a complex contagion model, showing that the results are robust as well concerning different spreading dynamics.

B.1 List of the 443 hashtags with partial and full agreement

In this section, we present supplementary tables of hashtags used in the reconstruction of the political communication networks considered in Chapter 6. Those for which an agreement was reached are shown in Tables B.1 to B.3, while the 52 with divergent classification are presented in Table B.4.

Table B.1: List of all the 184 hashtags classified as pro-impeachment leaning. For each hashtag, the opinion O_i of each volunteer i is reported. Four choices were possible: $s = \{-1, 0, +1, \times\}$.

Hashtag	O ₁	O ₂	O ₃	O ₄	Hashtag	O ₁	O ₂	O ₃	O ₄	Hashtag	O ₁	O ₂	O ₃	O ₄
final classification: -1														
1 13marco2016brasilnasruas	-1	-1	-1	-1	51 euapoiodeltan	-1	-1	-1	-1	102 ptdesmoronando	-1	-1	-1	-1
2 13mareuvou	-1	-1	-1	-1	52 eugritomoro	-1	-1	-1	-1	103 ptexit	-1	-1	-1	-1
3 31julho	-1	-1	-1	-1	53 eulutopelomoro	-1	-1	-1	-1	104 ptnuncamais	-1	-1	-1	-1
4 31julhoconfirmado	-1	-1	-1	-1	54 euquerolulapreso	-1	-1	-1	-1	105 pttraidoresdobrasil	-1	-1	-1	-1
5 31julhodireitafirme	-1	-1	-1	-1	55 eusomoro	-1	-1	-1	-1	106 ptvergonhanacional	-1	-1	-1	-1
6 31julhoouvou	-1	-1	-1	-1	56 extincadopt	-1	-1	-1	-1	107 quedadoplanalto	-1	-1	-1	-1
7 31julhoforcatal	-1	-1	-1	-1	57 ficatemer	-1	-1	-1	-1	108 quedaplanalto	-1	-1	-1	-1
8 31julhopelobrasil	-1	-1	-1	-1	58 fimdopt	-1	-1	-1	-1	109 renunciadilha	-1	-1	-1	-1
9 31jullavajato	-1	-1	-1	-1	59 foracomunismo	-1	-1	-1	-1	110 renunciadilha	-1	-1	-1	-1
10 31julvamos	-1	-1	-1	-1	60 foradilha	-1	-1	-1	-1	111 repudionomeacaoalula	-1	-1	-1	-1
11 4dezvempraru	-1	-1	-1	-1	61 foradilmalula	-1	-1	-1	-1	112 senadovotesim	-1	-1	-1	-1
12 acaboudilha	-1	-1	-1	-1	62 foralula	-1	-1	-1	-1	113 setembrosemilha	-1	-1	-1	-1
13 aceleramoro	-1	-1	-1	-1	63 forapt	-1	-1	-1	-1	114 simpeloimpeachment	-1	-1	-1	-1
14 acelerasnado	-1	-1	-1	-1	64 herancamaldita	-1	-1	-1	-1	115 somosoro	-1	-1	-1	-1
15 acelerastf	-1	-1	-1	-1	65 impeachmentdilmaja	-1	-1	-1	-1	116 somostodosjanaina	-1	-1	-1	-1
16 agostotchauquerida	-1	-1	-1	-1	66 impeachmentsim	-1	-1	-1	-1	117 somostodosoro	-1	-1	-1	-1
17 antagonistasnasruas	-1	-1	-1	-1	67 independenciasempt	-1	-1	-1	-1	118 souoro	-1	-1	-1	-1
18 apoiamostemer	-1	-1	-1	-1	68 lavajato	-1	-1	-1	-1	119 tchaumaldita	-1	-1	-1	-1
19 aragaopetralhao	-1	-1	-1	-1	69 lavajatoeuapoio	-1	-1	-1	-1	120 tchaupt	-1	-1	-1	-1
20 atopelofimdopt	-1	-1	-1	-1	70 lulacovar	-1	-1	-1	-1	121 tchauquerida	-1	-1	-1	-1
21 avantelavajato	-1	-1	-1	-1	71 luladenunciado	-1	-1	-1	-1	122 tchauqueridaday	-1	-1	-1	-1
22 avantetemer	-1	-1	-1	-1	72 luladenunciadonalavajato	-1	-1	-1	-1	123 temermelhorquept	-1	-1	-1	-1
23 brasilapoiajuliomcarcelo	-1	-1	-1	-1	73 lulaedilmanacadeia	-1	-1	-1	-1	124 todoapoioalavajato	-1	-1	-1	-1
24 brasilapoiatemer	-1	-1	-1	-1	74 lulagolpista	-1	-1	-1	-1	125 todoscontratoffoli	-1	-1	-1	-1
25 brasililivredopt	-1	-1	-1	-1	75 lulaladrao	-1	-1	-1	-1	126 todosnasruas31julho	-1	-1	-1	-1
26 brasilquerlulapreso	-1	-1	-1	-1	76 lulalixomundial	-1	-1	-1	-1	127 ultimopanelaco	-1	-1	-1	-1
27 brasilreprovadilha	-1	-1	-1	-1	77 lulamente	-1	-1	-1	-1	128 vaiadilha	-1	-1	-1	-1
28 brasilsemilha	-1	-1	-1	-1	78 lulananacadeia	-1	-1	-1	-1	129 vaiaterimpeachment	-1	-1	-1	-1
29 brasilsemt	-1	-1	-1	-1	79 lulananacadeiaja	-1	-1	-1	-1	130 vazostirarbrasilsovermelho	-1	-1	-1	-1
30 brazilnocorrupt	-1	-1	-1	-1	80 lulannuncamais	-1	-1	-1	-1	131 vazavacaloca	-1	-1	-1	-1
31 cadeia2ainstancia	-1	-1	-1	-1	81 lulapajaula	-1	-1	-1	-1	132 vempelomoro	-1	-1	-1	-1
32 cadeialula	-1	-1	-1	-1	82 lulapresoja	-1	-1	-1	-1	133 vempredatencao	-1	-1	-1	-1
33 camarasemt	-1	-1	-1	-1	83 lulareu	-1	-1	-1	-1	134 vempriefesp	-1	-1	-1	-1
34 cassacaodosdireitospoliticossim	-1	-1	-1	-1	84 lulavaipromoro	-1	-1	-1	-1	135 vempraruai3mar	-1	-1	-1	-1
35 decidamsim	-1	-1	-1	-1	85 lulavergonhanacional	-1	-1	-1	-1	136 vempraruai3marco	-1	-1	-1	-1
36 desculpasdopt	-1	-1	-1	-1	86 maranhaoempregadodopt	-1	-1	-1	-1	137 vempraruai7abril	-1	-1	-1	-1
37 dia13mareuvou	-1	-1	-1	-1	87 mexeucomomoroexeucomigo	-1	-1	-1	-1	138 vempraruai31jul	-1	-1	-1	-1
38 dia31juleuvou	-1	-1	-1	-1	88 morobrasilteapoia	-1	-1	-1	-1	139 vempraruai31julho	-1	-1	-1	-1
39 dilmabandida	-1	-1	-1	-1	89 morolidermundial	-1	-1	-1	-1	final classification: -1*				
40 dilmacaixa2	-1	-1	-1	-1	90 moropresidente	-1	-1	-1	-1	1 04dezvempraru	-1	-1	-1	x
41 dilmaculpada	-1	-1	-1	-1	91 novaseleicoesnao	-1	-1	-1	-1	2 13marco	-1	x	-1	-1
42 dilmaculpadaoea	-1	-1	-1	-1	92 oabdopt	-1	-1	-1	-1	3 13marcoouvou	-1	-1	-1	-1
43 dilmafungo	-1	-1	-1	-1	93 oabrepte92	-1	-1	-1	-1	4 31juleuvou	-1	-1	-1	-1
44 dilmagolpista	-1	-1	-1	-1	94 obrigadampf	-1	-1	-1	-1	5 31julhoavantebrasil	x	-1	-1	-1
45 dilmajaera	-1	-1	-1	-1	95 obrigadompf	-1	-1	-1	-1	6 4dezeuvou	-1	-1	-1	-1
46 dilmamente	-1	-1	-1	-1	96 ocupabh	-1	-1	-1	-1	7 avantepf	-1	-1	-1	0
47 dilmamentirosa	-1	-1	-1	-1	97 olimpiadassemilha	-1	-1	-1	-1	8 boratemer	-1	-1	-1	0
48 dilmnanacadeia	-1	-1	-1	-1	98 ouvocevaiouelafica	-1	-1	-1	-1	9 brasilapoialavajato	-1	-1	-1	x
49 dilmnanomerepresenta	-1	-1	-1	-1	99 passadilha	-1	-1	-1	-1	10 brasilnasruas	-1	-1	-1	x
50 dilmare	-1	-1	-1	-1	100 preندهojemoro	-1	-1	-1	-1	11 brasilnasruas20nov	-1	-1	-1	x
					101 ptcabou	-1	-1	-1	-1	12 crimeabandeirabr	-1	-1	-1	+1

Table B.2: List of all the 200 hashtags classified as anti-impeachment leaning. For each hashtag, the opinion O_i of each volunteer i is reported. Four choices were possible: $s = \{-1, 0, +1, \times\}$.

Hashtag	O ₁	O ₂	O ₃	O ₄	Hashtag	O ₁	O ₂	O ₃	O ₄	Hashtag	O ₁	O ₂	O ₃	O ₄	Hashtag	O ₁	O ₂	O ₃	O ₄				
final classification: +1					51	foratemerficahaddad	+1	+1	+1	+1	102	moralistassemoral	+1	+1	+1	+1	153	tonaruaforatemer	+1	+1	+1	+1	
1	180diasdegolpe	+1	+1	+1	+1	52	foratemer golpista	+1	+1	+1	+1	103	moroamigodecunha	+1	+1	+1	+1	154	vaiteforatemersim	+1	+1	+1	+1
2	54milhoesdedilmas	+1	+1	+1	+1	53	foratemerladrao	+1	+1	+1	+1	104	moroexonerado	+1	+1	+1	+1	155	vaite rlula	+1	+1	+1	+1
3	acaradogolpe	+1	+1	+1	+1	54	foratemerolimpico	+1	+1	+1	+1	105	moronacadeia	+1	+1	+1	+1	156	vaite rluta	+1	+1	+1	+1
4	aeciogolpista	+1	+1	+1	+1	55	foratemerrio2016	+1	+1	+1	+1	106	mulherescontratemer	+1	+1	+1	+1	157	vosbarrarosgolpistas	+1	+1	+1	+1
5	agoraerua	+1	+1	+1	+1	56	forcadilha	+1	+1	+1	+1	107	naoagolpe	+1	+1	+1	+1	158	vemprademocracia	+1	+1	+1	+1
6	alutacomecou	+1	+1	+1	+1	57	forcalula	+1	+1	+1	+1	108	naovaitergolpe	+1	+1	+1	+1	159	volta_querida_democracia	+1	+1	+1	+1
7	amarsemtemer	+1	+1	+1	+1	58	forcaquerida	+1	+1	+1	+1	109	naovoupagaropacto	+1	+1	+1	+1	160	voltadilha	+1	+1	+1	+1
8	anulamaranhao	+1	+1	+1	+1	59	getouttemer	+1	+1	+1	+1	110	ocupaminc	+1	+1	+1	+1	161	voltadilmapresidenta	+1	+1	+1	+1
9	anulastf	+1	+1	+1	+1	60	globogolpista	+1	+1	+1	+1	111	ocupaoimpiada	+1	+1	+1	+1	162	voltalula	+1	+1	+1	+1
10	anulatusdosupremo	+1	+1	+1	+1	61	golpeaquinoapassa	+1	+1	+1	+1	112	ocuparedeesgoto	+1	+1	+1	+1	163	voltaquerida	+1	+1	+1	+1
11	apoioalula	+1	+1	+1	+1	62	golpeday	+1	+1	+1	+1	113	ocupasenado	+1	+1	+1	+1	final classification: +1*					
12	avantetemerpraca deia	+1	+1	+1	+1	63	golpedeestado	+1	+1	+1	+1	114	ocupatudo	+1	+1	+1	+1	1	anulateori	+1	+1	+1	0
13	baralhodogolpe	+1	+1	+1	+1	64	golpeemachista	+1	+1	+1	+1	115	ocupatudocontraogolpe	+1	+1	+1	+1	2	brasiljusto	-1	+1	+1	+1
14	barulhacoforatemer	+1	+1	+1	+1	65	golpenao	+1	+1	+1	+1	116	ogolpeefichasuja	+1	+1	+1	+1	3	censuranuncamais	+1	+1	x	+1
15	blogueiroscondilma	+1	+1	+1	+1	66	golpenuncamais	+1	+1	+1	+1	117	opovodecide	+1	+1	+1	+1	4	coxinhaco	-1	+1	+1	+1
16	brasilcontraogolpe	+1	+1	+1	+1	67	golpismodamia	+1	+1	+1	+1	118	opovoquerdemocracia	+1	+1	+1	+1	5	cunhaetemer	x	+1	+1	+1
17	byedemocracyday	+1	+1	+1	+1	68	golpista	+1	+1	+1	+1	119	parabenspresidentadilha	+1	+1	+1	+1	6	cunhanacadeia	+1	+1	0	+1
18	cinogolpista	+1	+1	+1	+1	69	golpistas	+1	+1	+1	+1	120	pelademocracia	+1	+1	+1	+1	7	cunhasnacadeia	0	+1	+1	+1
19	comlulaporlula	+1	+1	+1	+1	70	golpistasday	+1	+1	+1	+1	121	pmdbgolpista	+1	+1	+1	+1	8	democraciaja	x	+1	+1	+1
20	coupinbrazil	+1	+1	+1	+1	71	gritocontraogolpe	+1	+1	+1	+1	122	povocomlula	+1	+1	+1	+1	9	desapegadolulastf	+1	+1	-1	+1
21	culturapelademocracia	+1	+1	+1	+1	72	gritodosexcluidos	+1	+1	+1	+1	123	psdbteupassadotecondena	+1	+1	+1	+1	10	desligaogolpe	x	+1	+1	+1
22	cunhagolpista	+1	+1	+1	+1	73	impeachmentsemcrimeegolpe	+1	+1	+1	+1	124	queremcalaraune	+1	+1	+1	+1	11	desligatv	+1	x	+1	+1
23	decidapelademocracia	+1	+1	+1	+1	74	jatempluta	+1	+1	+1	+1	125	ralatemer	+1	+1	+1	+1	12	dilmanovamente	+1	0	+1	+1
24	derrubargolpenasruas	+1	+1	+1	+1	75	jucagolpista	+1	+1	+1	+1	126	reforma naorenunciassim	+1	+1	+1	+1	13	eleicaoja	0	+1	+1	+1
25	devolverenan	+1	+1	+1	+1	76	lulaestamoscomvoce	+1	+1	+1	+1	127	renangolpista	+1	+1	+1	+1	14	eleicoesja	+1	0	+1	+1
26	dia31vaisermaior	+1	+1	+1	+1	77	lulaestamoscontigo	+1	+1	+1	+1	128	renunciatemer	+1	+1	+1	+1	15	ficalula	-1	+1	+1	+1
27	dilmacoracaovalente	+1	+1	+1	+1	78	lulaeterno	+1	+1	+1	+1	129	respeiteasurnas	+1	+1	+1	+1	16	ficaquerida	+1	+1	+1	0
28	dilmaeiocente	+1	+1	+1	+1	79	lulaeuconfio	+1	+1	+1	+1	130	ripdemocracia	+1	+1	+1	+1	17	foraserra	0	+1	+1	+1
29	dilmafica	+1	+1	+1	+1	80	lulaeudefendo	+1	+1	+1	+1	131	sendemocraciasempaz	+1	+1	+1	+1	18	fueratemer	+1	+1	0	+1
30	dilmaficagolpesai	+1	+1	+1	+1	81	lulaeurespeito	+1	+1	+1	+1	132	senadovotenoa	+1	+1	+1	+1	19	golpe	+1	+1	x	+1
31	dilmanatvbrasil	+1	+1	+1	+1	82	lulafica	+1	+1	+1	+1	133	somostodoslula	+1	+1	+1	+1	20	grevegeral	+1	x	+1	+1
32	dilmavolta	+1	+1	+1	+1	83	lulaisworththefight	+1	+1	+1	+1	134	somostodospt	+1	+1	+1	+1	21	joapaulo13comlula	x	+1	+1	+1
33	diretasja	+1	+1	+1	+1	84	lulalidermundial	+1	+1	+1	+1	135	soscoupinbrazil	+1	+1	+1	+1	22	libertemzedirceu	+1	x	+1	+1
34	ditaduratemer	+1	+1	+1	+1	85	lulaminstroja	+1	+1	+1	+1	136	soupelademocracia	+1	+1	+1	+1	23	lula2018	+1	0	+1	+1
35	eduardocunhagolpista	+1	+1	+1	+1	86	lulaperseguidopolitico	+1	+1	+1	+1	137	soupt	+1	+1	+1	+1	24	lulacasacivil	+1	+1	+1	0
36	egolpe	+1	+1	+1	+1	87	lulapresidente	+1	+1	+1	+1	138	standwithlula	+1	+1	+1	+1	25	lulaladenovo	+1	0	+1	+1
37	egolpesim	+1	+1	+1	+1	88	lulario2016	+1	+1	+1	+1	139	standwithlula	+1	+1	+1	+1	26	lularesiste	-1	+1	+1	+1
38	elmundocondilma	+1	+1	+1	+1	89	lulavalealuta	+1	+1	+1	+1	140	stopcoupinbrazil	+1	+1	+1	+1	27	lutareumdireito	+1	+1	-1	+1
39	emdefesadademocracia	+1	+1	+1	+1	90	lulavolta	+1	+1	+1	+1	141	teinadilha	+1	+1	+1	+1	28	naotemarrego	+1	x	+1	+1
40	entrouparaolixodahistoria	+1	+1	+1	+1	91	lutapelademocracia	+1	+1	+1	+1	142	temercaradepau	+1	+1	+1	+1	29	natalesemtemer	0	+1	+1	+1
41	esquentagrevegeral	+1	+1	+1	+1	92	lutarsempre	+1	+1	+1	+1	143	temerecunha	+1	+1	+1	+1	30	nenhumdireitoa menos	+1	+1	+1	0
42	estamoscomlula	+1	+1	+1	+1	93	lutopelademocracia	+1	+1	+1	+1	144	temereglobounidosnogolpe	+1	+1	+1	+1	31	ocupabrasilia	+1	+1	-1	+1
43	estamos todoscomlula	+1	+1	+1	+1	94	marchadascoxinhas	+1	+1	+1	+1	145	temergolpista	+1	+1	+1	+1	32	renunciacunha	0	+1	+1	+1
44	estoucomlula	+1	+1	+1	+1	95	marchadoscorruptos	+1	+1	+1	+1	146	temergolpistafrouxo	+1	+1	+1	+1	33	stfacovardado	+1	0	+1	+1
45	fairplayparadilha	+1	+1	+1	+1	96	marchadoscoxinhas	+1	+1	+1	+1	147	temerjamais	+1	+1	+1	+1	34	temersilveriodosreis	+1	-1	+1	+1
46	ficadilha	+1	+1	+1	+1	97	marchadospatinhospamonhas	+1	+1	+1	+1	148	temerout	+1	+1	+1	+1	35	traidores dopovo	+1	+1	-1	+1
47	foracoxinhas	+1	+1	+1	+1	98	mblgolpista	+1	+1	+1	+1	149	tocomdilha	+1	+1	+1	+1	36	vaitervai	+1	+1	-1	+1
48	foragilmar	+1	+1	+1	+1	99	mentiraenaglobo	+1	+1	+1	+1	150	tocomlula	+1	+1	+1	+1	37	vazatemer	0	+1	+1	+1
49	foragolpista	+1	+1	+1	+1	100	mexeucomlulamexeucomigo	+1	+1	+1	+1	151	todoscomdilha	+1	+1	+1	+1						
50	foragolpistas	+1	+1	+1	+1	101	mobilizacaototal	+1	+1	+1	+1	152	todoscomlula	+1	+1	+1	+1						

Table B.3: List of all the 20 hashtags classified as neutral leaning ($s = 0$) and all the 39 hashtags classified as not related ($s = \times$). For each hashtag, the opinion O_i of each volunteer i is reported. Four choices were possible: $s = \{-1, 0, +1, \times\}$.

Hashtag	O_1	O_2	O_3	O_4	Hashtag	O_1	O_2	O_3	O_4	Hashtag	O_1	O_2	O_3	O_4
final classification: 0					final classification: \times					final classification: \times				
1 foracorrutos	0	0	0	0	1 16ago	\times	\times	\times	\times	5 bolsonaro	\times	\times	-1	\times
2 impeachment	0	0	0	0	2 18marco	\times	\times	\times	\times	6 bolsonaro2018	\times	\times	-1	\times
3 tchaucunha	0	0	0	0	3 brasilnaonu	\times	\times	\times	\times	7 bolsonaropresidente	\times	\times	-1	\times
final classification: 0*					4 camara	\times	\times	\times	\times	8 contraogolpeedia18	\times	+1	\times	\times
1 delatacunha	0	+1	0	0	5 constituicao	\times	\times	\times	\times	9 corrupcao	0	\times	\times	\times
2 dilma	-1	0	0	0	6 dia18	\times	\times	\times	\times	10 cunha	\times	\times	\times	0
3 dilmanosbt	+1	0	0	0	7 dia18_03	\times	\times	\times	\times	11 delcidiotemrazao	\times	\times	\times	-1
4 dilmarousseff	0	0	0	-1	8 golacosdadilma	\times	\times	\times	\times	12 deputados	\times	\times	\times	0
5 diretasja2018	0	0	-1	0	9 justica	\times	\times	\times	\times	13 dilmacaradarenuncia	\times	+1	\times	\times
6 eduardocunha	0	0	\times	0	10 listatripliceagu	\times	\times	\times	\times	14 eleicoes2016	\times	\times	0	\times
7 ficamedina	\times	0	0	0	11 mandato	\times	\times	\times	\times	15 listafechadanao	\times	\times	0	\times
8 forabandidos	0	0	-1	0	12 moroaguarda	\times	\times	\times	\times	16 martraira	\times	\times	\times	-1
9 foraminc	0	0	0	+1	13 mpf	\times	\times	\times	\times	17 ocupario	\times	\times	-1	\times
10 forastf	0	0	0	-1	14 ocupabrazil	\times	\times	\times	\times	18 petrobras	0	\times	\times	\times
11 foratodosratos	0	0	\times	0	15 senadores	\times	\times	\times	\times	19 planalto	\times	\times	\times	0
12 janotgolpista	0	0	0	-1	16 timepetrobras	\times	\times	\times	\times	20 politica	+1	\times	\times	\times
13 seeufosseadilma	-1	0	0	0	final classification: \times^*					21 presaledopovo	+1	\times	\times	\times
14 sessaodoimpeachment	-1	0	0	0	1 10medidassemgolpe	\times	\times	\times	0	22 senado	+1	\times	\times	\times
15 stfvergonhanacional	0	0	0	\times	2 31mar	\times	\times	\times	+1	23 stf	\times	\times	\times	0
16 vempraruua	0	0	0	-1	3 anistiacaixa2nao	\times	\times	0	\times					
17 votacaoimpeachment	-1	0	0	0	4 anistiarcaixa2egolpe	\times	\times	0	\times					

B.2 List of the 52 hashtags with divergent opinions

In Table B.4 we show the list of hashtags with divergent opinions of the four volunteers. These hashtags were considered as neutral to reconstruct the 72-neutral network.

Table B.4: List of the 52 hashtags for which an agreement was not achieved. For each hashtag, the opinion O_i of each volunteer i is reported. Four choices were possible: $s = \{-1, 0, +1, \times\}$.

Hashtag	O_1	O_2	O_3	O_4	Hashtag	O_1	O_2	O_3	O_4
final classification: 0?					27 lulanorecife	+1	0	0	\times
1 2ainstanciadeia	\times	-1	-1	\times	28 mapadoimpeachment	-1	0	-1	0
2 aceleralavajatostf	\times	\times	-1	-1	29 mastenhoconviccao	\times	0	+1	+1
3 acordabrasil	-1	0	-1	\times	30 micheltemer	\times	0	\times	0
4 adeuscunha	\times	+1	0	+1	31 mudabrasil	-1	0	\times	0
5 brasilcontrastf	-1	+1	0	\times	32 ocupabrasil	0	+1	-1	0
6 comandantelula	\times	0	0	-1	33 ocupacopacabana	\times	0	+1	+1
7 cunhacaiu	-1	0	+1	+1	34 ocupapaulista	-1	+1	-1	0
8 desejoproteemer	+1	-1	0	0	35 ocuparj	-1	\times	+1	\times
9 eavezdasmulheres	\times	\times	+1	+1	36 ocupasp	-1	0	-1	\times
10 fimforoprivilegiado	0	0	\times	\times	37 ocupastf	+1	0	0	+1
11 foracunha	-1	+1	+1	0	38 olimpeachment	+1	-1	-1	0
12 forajuca	0	+1	+1	-1	39 panelaco	\times	\times	-1	+1
13 foraladrao	0	-1	\times	+1	40 possedadesonra	\times	+1	+1	0
14 foraoab	\times	\times	+1	+1	41 renanpreso	0	\times	0	\times
15 forapmdb	0	0	+1	+1	42 renanreu	\times	-1	0	\times
16 forarenan	\times	0	+1	+1	43 renantemealavajato	-1	-1	0	\times
17 forarodrigomaia	0	-1	0	+1	44 renunciaja	0	-1	-1	+1
18 foratemer	+1	0	0	+1	45 salvealavajato	\times	-1	-1	\times
19 foratodos	0	+1	0	\times	46 sergiomoro	-1	0	-1	0
20 impeachmentbrazil	-1	0	0	+1	47 somostodosgolpistas	-1	-1	+1	+1
21 impeachmentday	-1	0	+1	0	48 souptpq	\times	+1	+1	0
22 impeachmentja	\times	-1	-1	0	49 tchauquerido	\times	+1	-1	\times
23 jucanacadeia	+1	+1	0	0	50 temer	\times	0	\times	0
24 lula	+1	0	\times	0	51 teorigolpista	-1	\times	+1	+1
25 lulala	-1	\times	-1	+1	52 vergonhacongressobr	\times	+1	+1	\times
26 lulaministro	+1	0	0	+1					

B.3 Effects of the hashtag classification on the reconstructed network

Figure B.1 reproduces Fig. 6.5 for the activity frequency of tweets in the SCC of the 72-neutral network for different sentiments. As expected, we note that the fraction of tweets containing neutral tweets increases as a natural consequence of the addition of 52 neutral hashtags for the network reconstruction.

In Fig. B.2, we repeat the analysis of distributions of activity and burstiness to the 72-neutral network, showing results almost indistinguishable from those of Fig. 6.6 for the 20-neutral

network.

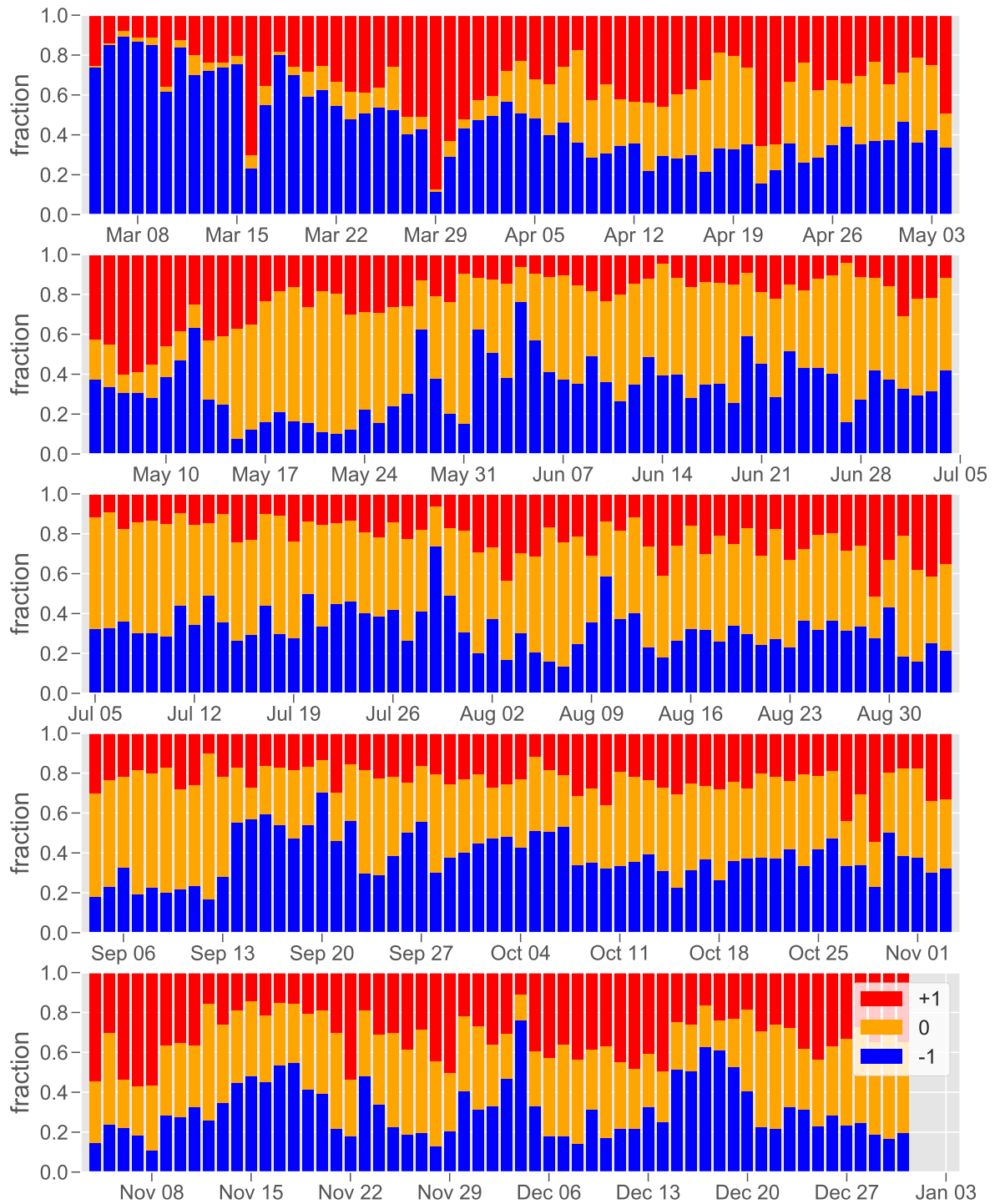


Figure B.1: Activity frequency of tweets for the SCC of the 72-neutral network. The legend indicates the colors corresponding to the activity for -1 , 0 and $+1$ interactions.

In Fig. B.3 we repeat Fig. 6.8 for the 72-neutral network, showing the distribution of political position, activity and community structures. Results are very similar, except for the increase of secondary anti-impeachment communities. The community analysis resulted in $Q = 0.435$ and

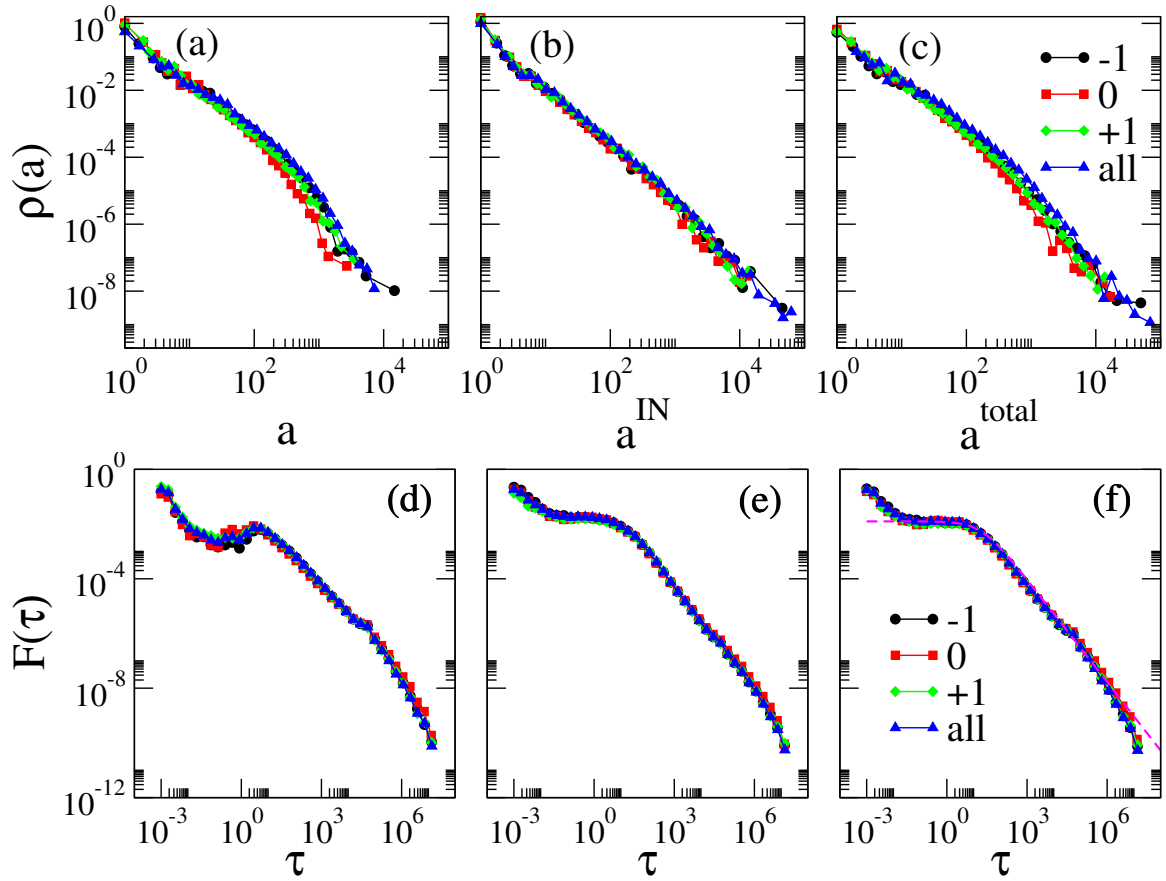


Figure B.2: Distributions of activity a and burstiness τ (a,d) of senders, (b,e) receivers, and (c,f) total, of interactions with sentiments -1 , 0 , $+1$ and all. The dashed curve in (f) is given by $F(\tau) = \kappa\alpha / (1 + \kappa\tau)^{(\alpha+1)}$, with $\kappa = 0.05$ and $\alpha = 0.25$.

0.431 for the 20-neutral and 72-neutral networks, respectively. The community structures of both networks are described in Table B.5.

The echo chambers effects in the structure of the 72-neutral network are similar to those of 20-neutral network, as shown in Fig. B.4 for the correlation between the political position of users with tweets received and of neighbors in the integrated network.

The comparison of results for the spreading capacity of users is done by using the SIS dynamics with $\bar{\lambda} = 0.5$ and $\tau = 7$ days for Fig. B.5(a), equivalent to the heat map of Fig. 6.13(e), and $\bar{\lambda} = 0.2$ and $\tau = 7$ days in Fig. B.5(b) corresponding to Fig. 6.14(a). Note that in addition to the conclusions being the same of 20-neutral network, the greater spreading capacity for pro-impeachment users is even clearer in Fig. B.5(b).

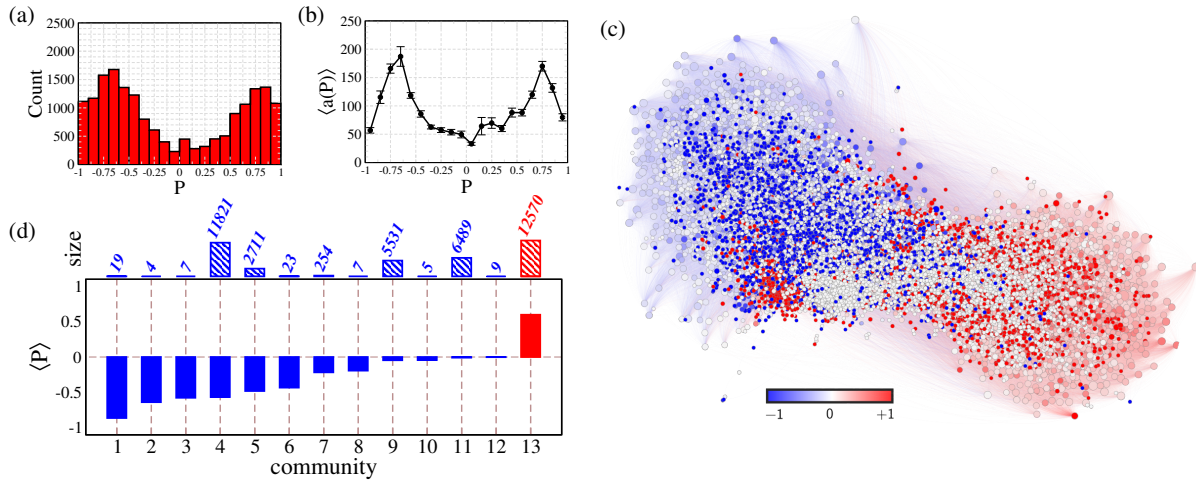


Figure B.3: Figure 6.8 of the main text for the 72-neutral network. (a) Number of users as a function of political position P . (b) Average activity as a function of P . Only users with activity $a \geq 10$ in the SCC are considered for (a) and (b). (c) Visualization of the time-aggregated representation of the PC network, formed by $N = 39\,525$ users in the SCC. The size of vertices increases (non-linearly) with their degree. Colors represent political position, as defined in the main text, blue for pro-, red for anti-impeachment, and white for neutral average leaning of users. (d) Community size and average political position of different communities identified by the Louvain algorithm.

Table B.5: Community structure of the networks 20-neutral and 72-neutral, according to the Louvain algorithm. Very small communities with only a few vertices are omitted due to the resolution limit of the modularity optimization [216].

20-neutral		72-neutral	
Size	$\langle P \rangle$	Size	$\langle P \rangle$
10 502	0.840 ± 0.437	12 570	0.598 ± 0.438
9937	-0.687 ± 0.428	11 821	-0.566 ± 0.436
4238	-0.097 ± 0.852	6489	-0.009 ± 0.654
3708	-0.011 ± 0.829	5531	-0.045 ± 0.698
2599	-0.529 ± 0.427	2711	-0.481 ± 0.393
170	-0.484 ± 0.781	254	-0.217 ± 0.764
52	-0.043 ± 0.884	23	-0.433 ± 0.592
37	-0.448 ± 0.696	19	-0.863 ± 0.217
26	-0.827 ± 0.450	9	-0.002 ± 0.623
23	0.998 ± 0.009		
18	-0.520 ± 0.617		
9	-0.459 ± 0.806		
8	-0.811 ± 0.275		

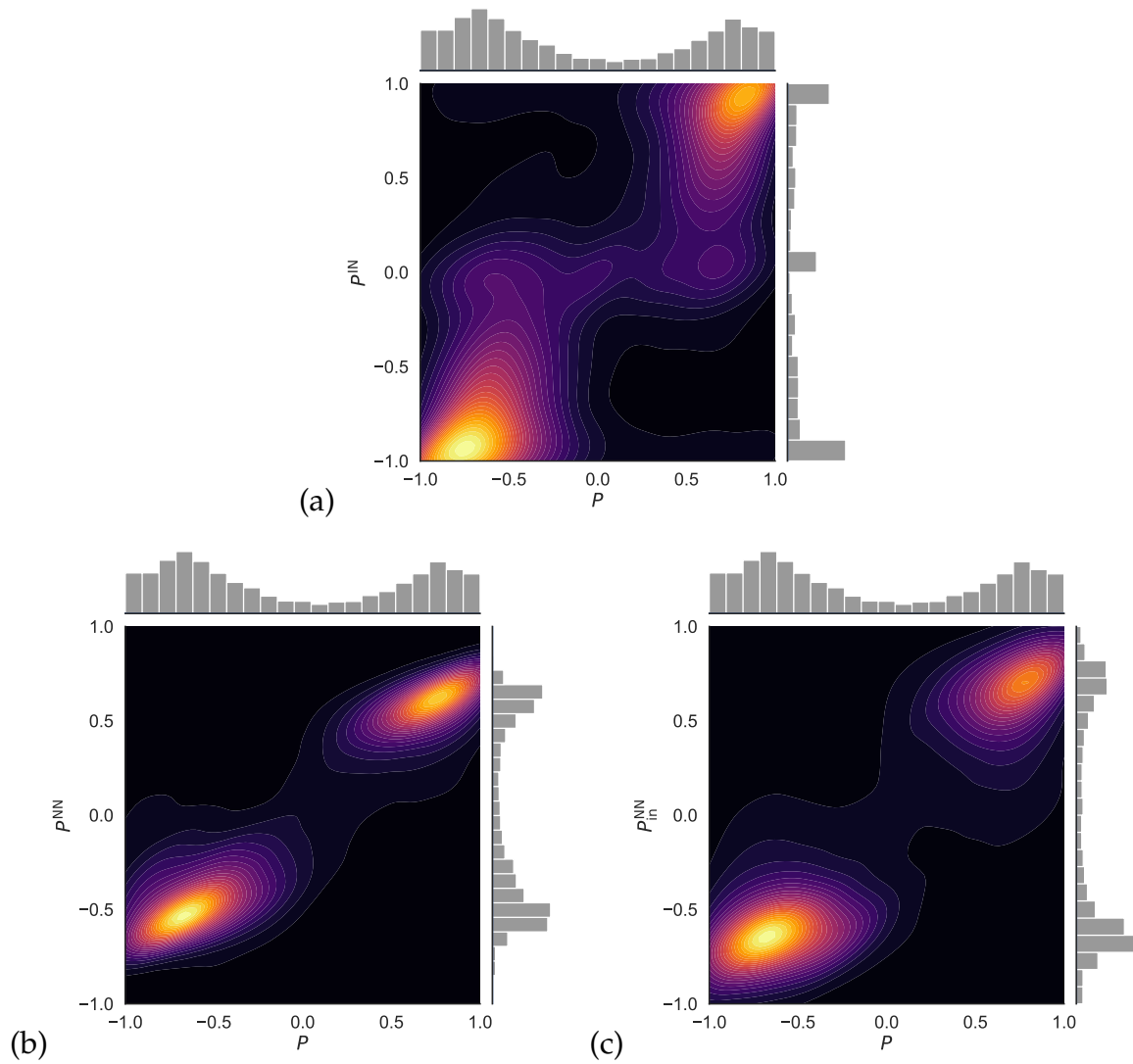


Figure B.4: Figure 6.10 of the main text for the 72-neutral network. Contour maps for average (a) political position of tweets received P^{IN} , (b) sentiment of successors P^{NN} and (c) predecessors $P^{\text{NN}}_{\text{in},i}$ against the political position P of a user for the 20-neutral network. Colors represent the density of users: the lighter the larger the number of users. Probability distribution of P , P^{NN} and $P^{\text{NN}}_{\text{in},i}$ are plotted in the axes. Only users with activity $a \geq 10$ (corresponding to 17 923 users) are considered.

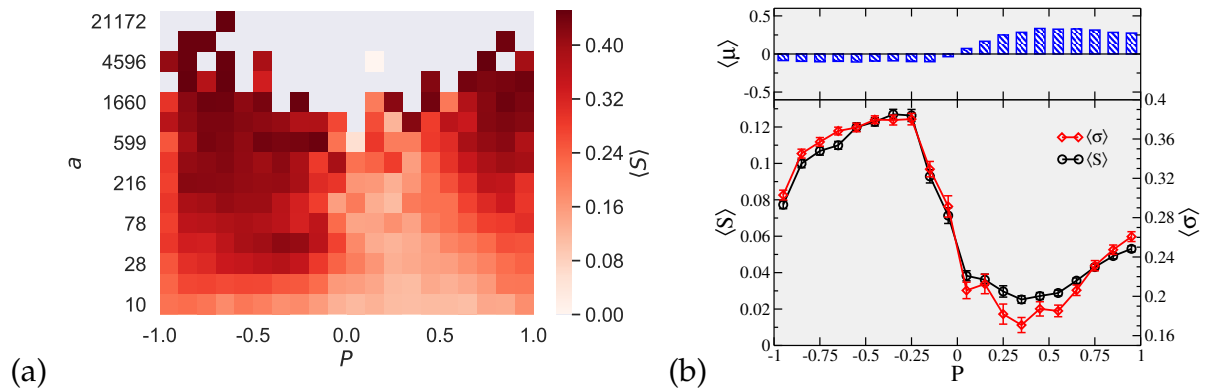


Figure B.5: Robustness of results for spreadability capacity in 72-neutral network. (a) Heat map of the average spreading capacity $\langle S \rangle$ of users, as a function of their political position P and activity a . The transmission probability of the SIS dynamics is $\bar{\lambda} = 0.5$ and $\tau = 7$ days. (b) Average spreading capacity $\langle S(P) \rangle$ (black curve, left axes), diversity $\langle \sigma(P) \rangle$ (red curve, right axes) and political position $\langle \mu(P) \rangle$ (bars, top panel) of the set of influence reached by users with political position P . Transmission probability $\bar{\lambda} = 0.2$ and $\tau = 7$ days. Only the 13556 users with activity $a \in [10, 100]$ are considered. Results are averaged over 100 runs, error bars represent the standard error.

B.4 Results for SIS and SIR dynamics with different parameters

In Figs. B.6 and B.7, analyses of the dependence with the infection rate and healing times of the average spreading capacity $\langle S \rangle$, diversity σ , and political position μ of the set of influence \mathcal{I} are shown for SIS and SIR epidemic processes, respectively for parameters $\bar{\lambda}$ and τ not covered in the main text. We can see that despite expected quantitative differences due to the nature of models, both dynamical processes exhibit similar behaviors which are also preserved as the parameters are varied.

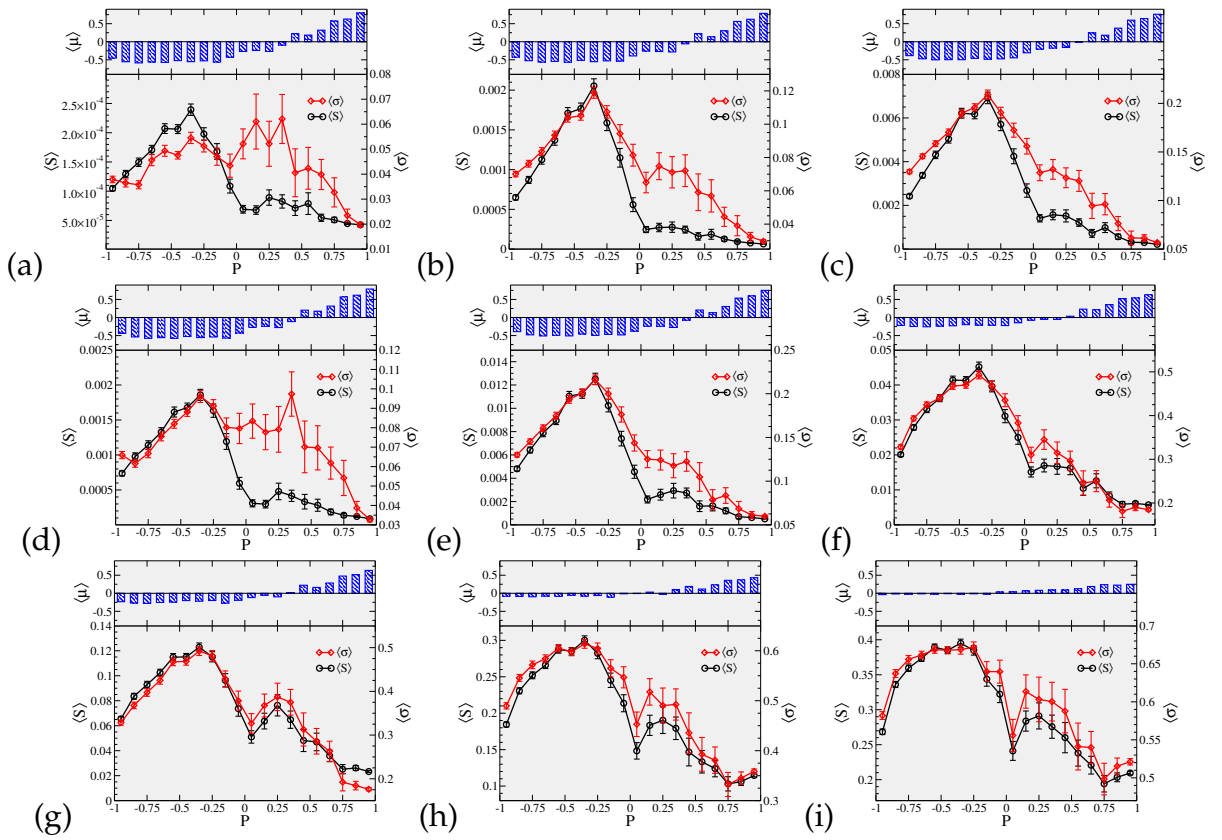


Figure B.6: Average spreading capacity $\langle S \rangle$ (black, left axes), diversity σ (red, right axes), and political position μ (top panel) of the set of influence \mathcal{I} , as a function of the political position P , for SIS model with transmission probability (a)–(c) $\bar{\lambda} = 0.05$, (d)–(f) $\bar{\lambda} = 0.10$ and (g)–(i) $\bar{\lambda} = 0.50$ for the temporal 20-neutral network. The healing times τ are (a,d,g) 1 day, (b,e,h) 3 days and (c,f,i) 7 days. Only users with activity $a \in [10,100]$ were considered. Averages were performed over 100 runs.

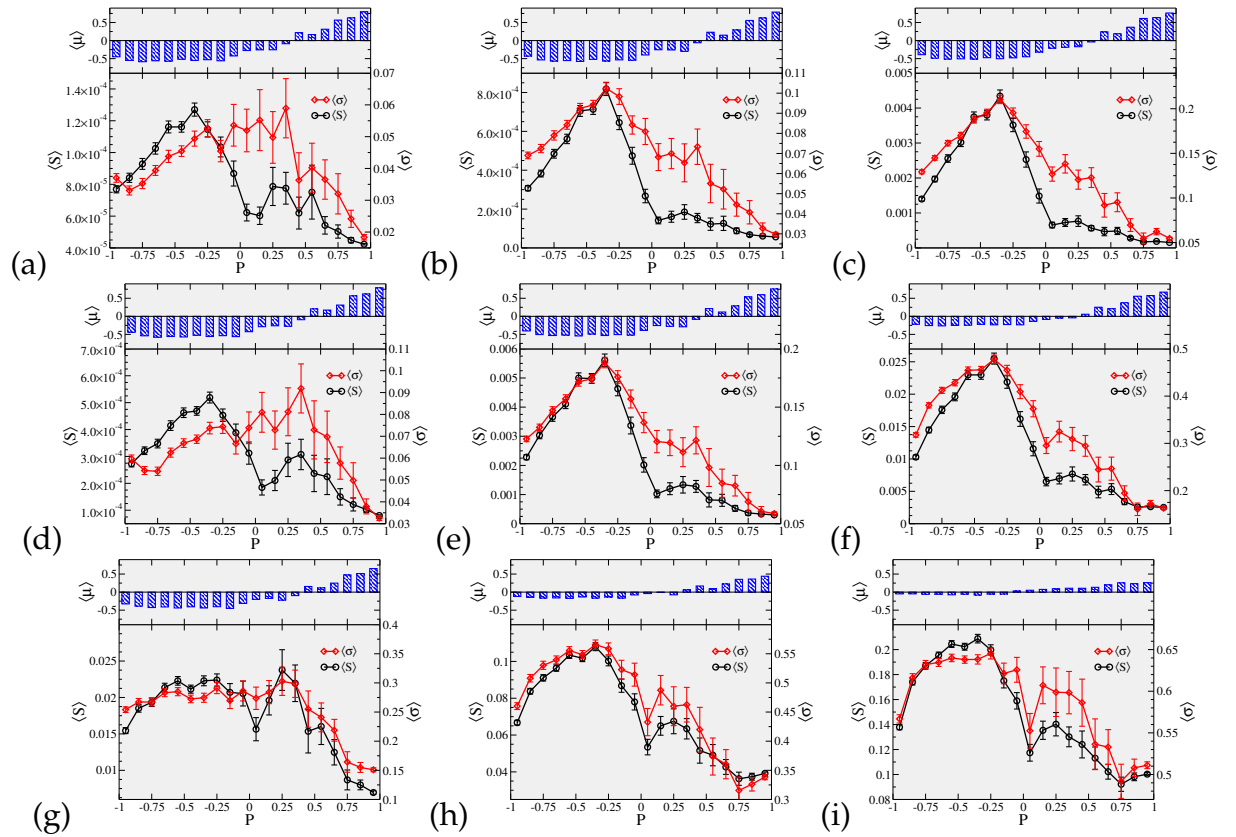


Figure B.7: Average spreading capacity $\langle S \rangle$ (black, left axes), diversity σ (red, right axes), and political position μ (top panel) of the set of influence \mathcal{I} , as a function of the political position P , for SIR model with transmission probability (a)–(c) $\bar{\lambda} = 0.05$, (d)–(f) $\bar{\lambda} = 0.10$ and (g)–(i) $\bar{\lambda} = 0.50$ for the temporal 20-neutral network. The healing times τ are (a,d,g) 1 day, (b,e,h) 3 days and (c,f,i) 7 days. Only users with activity $a \in [10,100]$ were considered. Averages were performed over 100 runs.

B.5 Results for the Watts threshold model

To check the robustness of our results on different spreading models, we have considered additionally a modification of the classic Watts threshold model of complex contagion [183]. In this model, each individual is either in state S or I , whose interpretation is akin to the one in the SIR/SIS models. We have considered the absolute-threshold version of the Watts model on temporal networks described in Ref. [187], in which each individual is endowed with a threshold value Φ . For each interaction at a time t , an individual in state S counts the total number of contacts from infected vertices to him/her within a time window $[t - \theta, t]$. If this value is larger than Φ , individual i flips to state I ; otherwise it remains in the S state. Transitions from I to S are forbidden. Starting from a single individual in state I , a cascade of transitions to state I is produced. In Fig. B.8 we show the results analogous to those for SIS and SIR models using the absolute-threshold Watts model to compute spreading capacity and diversity as a function of the political position P . As we can observe, all three models yield the same qualitative behavior.

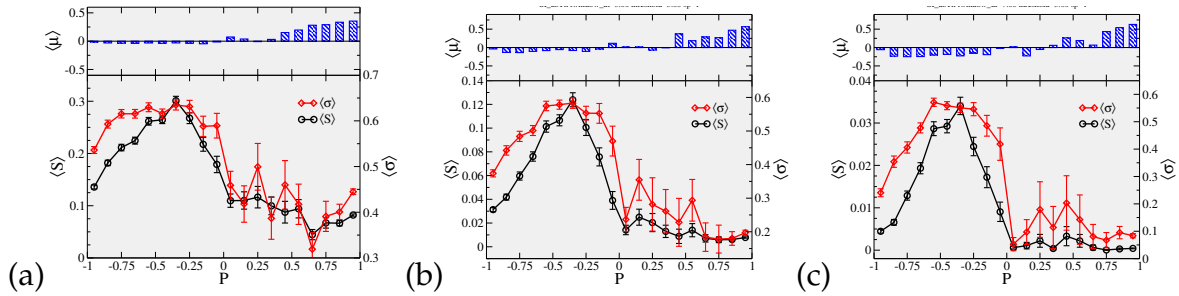


Figure B.8: Average spreading capacity $\langle S \rangle$ (black, left axes), diversity σ (red, right axes), and political position μ (top panel) of the set of influence \mathcal{I} , as a function of the political position P , for the absolute-threshold model on the 20-neutral network with (a) $\theta = 2$ days, $\Phi = 2$, (b) $\theta = 3$ days, $\Phi = 3$ and (c) $\theta = 7$ days, $\Phi = 5$. Only users with activity $a \in [10, 100]$ are considered.

Appendix C

Supplementary calculations for Chapter 7

C.1 An upper bound limit of the leading eigenvalue of a matrix

The Gershgorin circle theorem [205] states that for a matrix A with elements a_{ij} , all eigenvalues lie within at least one of the Gershgorin disks $D(a_{ii}, R_i)$, centered in a_{ii} with radius R_i defined by

$$R_i = \sum_{j \neq i} |a_{ij}|. \quad (\text{C.1})$$

To prove this, let Λ_i be the i -th eigenvalue of A and $\vec{x}_i = \{x_j\}$ the corresponding eigenvector with elements $x_i = 1$ and $|x_j| \leq 1 \forall j \neq i$. Since $A\vec{x}_i = \Lambda_i\vec{x}_i$,

$$\begin{aligned} \sum_j a_{ij}x_j &= \Lambda_i x_i = \Lambda_i, \\ \sum_{j \neq i} a_{ij}x_j + a_{ii}x_i &= \Lambda_i, \\ \implies \Lambda_i - a_{ii} &= \sum_{j \neq i} a_{ij}x_j. \end{aligned}$$

By the triangle inequality,

$$|\Lambda_i - a_{ii}| = \left| \sum_{j \neq i} a_{ij}x_j \right| \leq \sum_{j \neq i} |a_{ij}||x_j| \leq \sum_{j \neq i} |a_{ij}| = R_i.$$

For non-negative matrix, $a_{ij} \geq 0$. Thus,

$$\begin{aligned} -R_i &\leq \Lambda_i - a_{ii} \leq R_i \\ -\sum_{j \neq i} |a_{ij}| &= -\sum_{j \neq i} a_{ij} \leq \Lambda_i - a_{ii} \leq \sum_{j \neq i} |a_{ij}| = \sum_{j \neq i} a_{ij}, \end{aligned}$$

implying that

$$\Lambda_i \leq \sum_j a_{ij}, \quad (\text{C.2})$$

in which we used $\sum_{j \neq i} a_{ij} + a_{ii} = \sum_i a_{ij}$. The leading eigenvalue will be inside of the circle with the largest value of $a_{ii} + R_i$.

C.2 Perturbation analysis of Eq. (7.40)

We proceed by making a perturbation analysis of the eigenvalues of the matrix M up to second order on p . After writing Eq. (7.40) as a polynomial in p , we get Eq. (7.47). Since Q_i is also a function of p , we must perform a Taylor expansion around $p = 0$, knowing that $Q_i|_{p=0} = n_i \langle k \rangle_i$. The first derivative of Q_i is

$$\left. \frac{dQ_i}{dp} \right|_{p=0} = \sum_j \langle k \rangle_j (R_{ji} - \delta_{ij}) n_j,$$

and the second derivative evaluated in $p = 0$ is

$$\left. \frac{d^2 Q_i}{dp^2} \right|_{p=0} = 0.$$

Let us define

$$r_i \equiv \sum_j (-R_{ji} + \delta_{ij}) n_j \langle k \rangle_j,$$

so that

$$\left. \frac{d}{dp} \left(\frac{1}{Q_i} \right) \right|_{p=0} = \frac{r_i}{(n_i \langle k \rangle_i)^2},$$

and

$$\left. \frac{d^2}{dp^2} \left(\frac{1}{Q_i} \right) \right|_{p=0} = 2 \frac{r_i^2}{(n_i \langle k \rangle_i)^3}.$$

Next, by keeping only terms up to order 2, we have

$$\frac{1}{Q_i} = \frac{1}{n_i \langle k \rangle_i} + p \frac{r_i}{(n_i \langle k \rangle_i)^2} + p^2 \frac{r_i^2}{(n_i \langle k \rangle_i)^3} + O(p^3).$$

Substituting the last expression in Eq. (7.47) we get, after some algebra,

$$M_{ij} \approx \widetilde{M}_{ij}^{(0)} + p \widetilde{M}_{ij}^{(1)} + p^2 \widetilde{M}_{ij}^{(2)}, \quad (\text{C.3})$$

where

$$\widetilde{M}_{ij}^{(0)} = \delta_{ij} \frac{\langle k^2 \rangle_i}{\langle k \rangle_i}, \quad (\text{C.4a})$$

$$\widetilde{M}_{ij}^{(1)} = \left[\frac{R_{ij}}{n_j \langle k \rangle_j} + \frac{R_{ji}}{n_i \langle k \rangle_i} + \frac{\delta_{ij}}{n_i \langle k \rangle_i} \left(\frac{r_i}{n_i \langle k \rangle_i} - 2 \right) \right] n_j \langle k^2 \rangle_j, \quad (\text{C.4b})$$

$$\widetilde{M}_{ij}^{(2)} = \left[\frac{\delta_{ij}}{n_i \langle k \rangle_i} \left(\frac{\tilde{r}_i}{n_i \langle k \rangle_i} \right)^2 - \frac{R_{ij}}{n_j \langle k \rangle_j} \left(\frac{\tilde{r}_j}{n_j \langle k \rangle_j} \right) - \frac{R_{ji}}{n_i \langle k \rangle_i} \left(\frac{\tilde{r}_i}{n_i \langle k \rangle_i} \right) + \sum_l \frac{R_{il} R_{jl}}{n_l \langle k \rangle_l} \right] n_j \langle k^2 \rangle_j, \quad (\text{C.4c})$$

and

$$\tilde{r}_i = -(r_i - n_i \langle k \rangle_i) = \sum_{j \neq i} R_{ji} n_j \langle k \rangle_j.$$

From the static case, we know that there are Ω unperturbed eigenvalues $\Lambda_i^{(0)} = \langle k^2 \rangle_i / \langle k \rangle_i$, for $p = 0$, with normalized eigenvectors $\vec{\epsilon}_i = \{\epsilon_j\}$ and $\epsilon_j = \delta_{ij}$; see Eq. (7.42). Assuming that the eigenvalues are not degenerate, the new eigenvalues will be given by [217]

$$\Lambda_i \approx \Lambda_i^{(0)} + p \Lambda_i^{(1)} + p^2 \Lambda_i^{(2)}, \quad (\text{C.5})$$

where

$$\Lambda_i^{(0)} = \frac{\langle k^2 \rangle_i}{\langle k \rangle_i}, \quad (\text{C.6a})$$

$$\Lambda_i^{(1)} = \vec{\epsilon}_i \widetilde{\mathbf{M}}^{(1)} \vec{\epsilon}_i, \quad (\text{C.6b})$$

$$\Lambda_i^{(2)} = \sum_{j \neq i} \frac{\left(\vec{\epsilon}_i \widetilde{\mathbf{M}}^{(1)} \vec{\epsilon}_j \right) \left(\vec{\epsilon}_j \widetilde{\mathbf{M}}^{(1)} \vec{\epsilon}_i \right)}{\Lambda_i^{(0)} - \Lambda_j^{(0)}} + \vec{\epsilon}_i \widetilde{\mathbf{M}}^{(2)} \vec{\epsilon}_i. \quad (\text{C.6c})$$

Substituting Eq. (C.4b) in Eq. (C.6b), after some algebra we get the first correction to the eigenvalue,

$$\frac{\Lambda_i^{(1)}}{\Lambda_i^{(0)}} = R_{ii} - 1 - \sum_{j \neq i} R_{ji} \frac{n_j \langle k \rangle_j}{n_i \langle k \rangle_i}. \quad (\text{C.7})$$

To get the second order correction, we substitute Eqs. (C.4b) and (C.4c) in Eq. (C.6c), resulting in

$$\begin{aligned} \frac{\Lambda_i^{(2)}}{\Lambda_i^{(0)}} = n_i \langle k \rangle_i \left\{ \sum_{j \neq i} \frac{n_j \langle k^2 \rangle_j}{\langle k \rangle_i - \langle k \rangle_j} \left[\frac{R_{ij}}{n_j \langle k \rangle_j} + \frac{R_{ji}}{n_i \langle k \rangle_i} \right]^2 + \frac{1}{(n_i \langle k \rangle_i)^3} \left(\sum_l R_{li} n_l \langle k \rangle_l \right)^2 \right. \\ \left. - \frac{2R_{ii}}{(n_i \langle k \rangle_i)^2} \sum_l R_{li} n_l \langle k \rangle_l + \sum_l \frac{R_{il}^2}{n_l \langle k \rangle_l} \right\}. \end{aligned} \quad (\text{C.8})$$

This expression is always positive, meaning that the leading eigenvalue has a minimum for some value of p^* . In Fig. C.1 we compare the perturbation analysis with numerical, exact and lower bound estimate for the epidemic threshold. The agreement is remarkable close for $p \ll 0$ and increases with the level of heterogeneity.

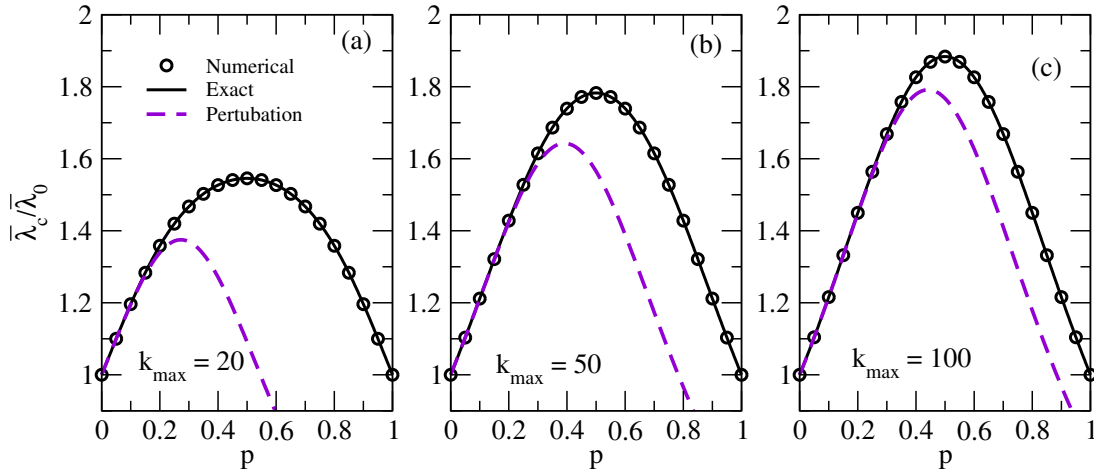


Figure C.1: Comparison of the numerical (points), exact (solid lines) and approximation by Eq. (C.5) with $\frac{\Lambda_i^{(1)}}{\Lambda_i^{(0)}}$ given by Eq. (C.8), and $\frac{\Lambda_i^{(2)}}{\Lambda_i^{(0)}}$ expressed in Eq. (C.8). The metapopulation network has two-patches in which the first has $n_i = 10^5$ individuals of sociability either 1 or (a) $k_{\max} = 20$, (b) 50 or (c) 100, and the second patch has individuals with $k_2 = \langle k \rangle_1 = 5$.

C.3 Exact evaluation of the epidemic threshold for a two-patches metapopulation

In the two-patches metapopulation, $R_{12} = R_{21} = 1$. Applying this and considering a k -independent movement, $p_k = p$, in Eq. (7.40), we have

$$M_{ii} = \langle k^2 \rangle_i \left[(1-p)^2 \frac{1}{Q_i} + p^2 \frac{1}{Q_j} \right] n_i, \quad (\text{C.9})$$

$$M_{ij} = \langle k^2 \rangle_j \left[(1-p)p \left(\frac{1}{Q_i} + \frac{1}{Q_j} \right) \right] n_j, \quad (\text{C.10})$$

where

$$Q_i = \langle k \rangle_i (1-p)n_i + p \langle k \rangle_j n_j \quad (\text{C.11})$$

for $i \neq j$. Using Eq. (7.39), we have to solve the following equation,

$$\bar{\mu} \epsilon_i = \bar{\lambda} \sum_j M_{ij} = \bar{\lambda} M_{ii} \epsilon_i + \bar{\lambda} M_{ij} \epsilon_j, \quad (\text{C.12})$$

that can be translated in the evaluation of the eigenvalue of a 2×2 matrix

$$M = \begin{pmatrix} \langle k^2 \rangle_1 \left[(1-p)^2 \frac{1}{Q_1} + p^2 \frac{1}{Q_2} \right] n_1 & \langle k^2 \rangle_2 \left[(1-p)p \left(\frac{1}{Q_1} + \frac{1}{Q_2} \right) \right] n_2 \\ \langle k^2 \rangle_1 \left[(1-p)p \left(\frac{1}{Q_1} + \frac{1}{Q_2} \right) \right] n_1 & \langle k^2 \rangle_2 \left[(1-p)^2 \frac{1}{Q_2} + p^2 \frac{1}{Q_1} \right] n_2 \end{pmatrix}. \quad (\text{C.13})$$

In this case, the leading eigenvalue will be given by

$$\Lambda_{\max} = \frac{\text{Tr } M + \sqrt{(\text{Tr } M)^2 - 4 \det M}}{2}, \quad (\text{C.14})$$

where $\text{Tr } M$ is the trace of the matrix M , and $\det M$ its determinant. Assuming that $\langle k \rangle_1 = \langle k \rangle_2$ and $\langle k^2 \rangle_2 = a \langle k^2 \rangle_1$, after solving the previous equations in a computer algebra system such as

Maxima [218], we get

$$\frac{\Lambda_{\max}(p)}{\Lambda_1^{(0)}} = \frac{\sqrt{(a+1)^2(4p^4 - 8p^3) + 8(a^2 + 1)p^2 - 4(a+1)^2p + (a-1)^2 + (a+1)(2p^2 - 2p + 1)}}{2}, \quad (\text{C.15})$$

with $a \equiv \langle k^2 \rangle_2 / \langle k^2 \rangle_1 < 1$.

C.4 Exact evaluation of the epidemic threshold for a star-like metapopulation

In this case, we have to evaluate seven different terms:

- M_{hh} : contact of two individuals resident in the hub;
- M_{lh} : contact of one resident of a leaf with another from the hub;
- M_{hl} : contact of one resident of the hub with another of a leaf;
- M_{ll} : contact of two individuals resident in the same leaf;
- $M_{l,l+1}$: contact of one resident of a leaf with another from its adjacent leaf;
- $M_{l,l-1}$: contact of one resident of the adjacent leaf with one from the other leaf;
- M_{ln} : contact of two residents of different and not adjacent leaves;

The mobility matrix elements R_{ij} are expressed in Eqs. (7.13) to (7.15). Applying these expressions in Eq. (7.40), we have

$$M_{hh} = \langle k^2 \rangle_h \left[(1-p)^2 \frac{1}{Q_h} + \frac{p^2}{\kappa} \frac{1}{Q_l} \right] n_h \quad (\text{C.16a})$$

$$M_{lh} = \langle k^2 \rangle_h \left[(1-p)p \left(\frac{1}{\kappa} \frac{1}{Q_l} + \delta \frac{1}{Q_h} \right) + p^2 \frac{(1-\delta)}{\kappa} \frac{1}{Q_l} \right] n_h \quad (\text{C.16b})$$

$$M_{hl} = \langle k^2 \rangle_l \left[(1-p)p \left(\delta \frac{1}{Q_h} + \frac{1}{\kappa} \frac{1}{Q_l} \right) + p^2 \frac{(1-\delta)}{\kappa} \frac{1}{Q_l} \right] n_l \quad (\text{C.16c})$$

$$M_{ll} = \langle k^2 \rangle_l \left[(1-p)^2 \frac{1}{Q_l} + p^2 \frac{(1-\delta)^2}{Q_l} + p^2 \frac{\delta^2}{Q_h} \right] n_l \quad (\text{C.16d})$$

$$M_{l,l+1} = \langle k^2 \rangle_l \left[(1-p)p \frac{(1-\delta)}{Q_l} + p^2 \frac{\delta^2}{Q_h} \right] n_l \quad (\text{C.16e})$$

$$M_{l,l-1} = \langle k^2 \rangle_l \left[(1-p)p \frac{(1-\delta)}{Q_l} + p^2 \frac{\delta^2}{Q_h} \right] n_l \quad (\text{C.16f})$$

$$M_{ln} = \langle k^2 \rangle_l \left(p^2 \frac{\delta^2}{Q_h} \right) n_l, \quad (\text{C.16g})$$

with $Q_h = (1 - p)\langle k \rangle_h n_h + p\kappa\delta n_l \langle k \rangle_l$ and $Q_l = (1 - p)\langle k \rangle_l n_l + p\kappa^{-1}n_h \langle k \rangle_h + p(1 - \delta)n_l \langle k \rangle_l$.

Again, by evaluating Eq. (7.39), we have, for the hub,

$$\bar{\mu}\epsilon_h = \bar{\lambda} \sum_j M_{hj} = \bar{\lambda} M_{hh}\epsilon_h + \kappa\bar{\lambda} M_{hl}\epsilon_l, \quad (\text{C.17})$$

while for a leaf we have

$$\bar{\mu}\epsilon_l = \bar{\lambda} \sum_j M_{lj} = \bar{\lambda} M_{lh}\epsilon_h + \bar{\lambda} M_{ll}\epsilon_l + \bar{\lambda} M_{l,l+1}\epsilon_l + \bar{\lambda} M_{l,l-1}\epsilon_l + \bar{\lambda}(\kappa - 3)M_{ln}\epsilon_l, \quad (\text{C.18})$$

in which the factor 3 in the last term is since there are $\kappa - 3$ other leaves not directly connected to a single leaf ($R_{ln} = 0$). The previous expressions can be translated in an eigenvalue problem to solve of the 2×2 matrix

$$M = \begin{bmatrix} M_{hh} & \kappa M_{hl} \\ M_{lh} & M_{ll} + M_{l,l+1} + M_{l,l-1} + (\kappa - 3)M_{ln} \end{bmatrix}, \quad (\text{C.19})$$

whose terms are defined in Eq. (C.16).

Like in the case for a two-patches metapopulation, the leading eigenvalue will be given by $\Lambda_{\max} = \frac{\text{Tr } M + \sqrt{(\text{Tr } M)^2 - 4 \det M}}{2}$, that was solved in the Maxima [218] software to get the results shown in the main text.

Bibliography

- [1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford New York, 2010).
- [2] A.-L. Barabási and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, 2016).
- [3] P. Sen and B. K. Chakrabarti, *Sociophysics: An Introduction* (OUP Oxford, 2013).
- [4] L. d. F. Costa, O. N. Oliveira, G. Travieso, F. A. Rodrigues, P. R. Villas Boas, L. Antiqueira, M. P. Viana, and L. E. Correa Rocha, “Analyzing and modeling real-world phenomena with complex networks: a survey of applications,” *Adv. Phys.* **60**, 329–412 (2011).
- [5] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Rev. Mod. Phys.* **74**, 47–97 (2002).
- [6] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2012).
- [7] P. W. Anderson, “More is different,” *Science* **177**, 393–396 (1972).
- [8] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Rev. Mod. Phys.* **87**, 925–979 (2015).
- [9] S. Chatterjee and R. Durrett, “Contact processes on random graphs with power law degree distributions have critical value 0,” *Ann. Probab.* **37**, 2332–2356 (2009).
- [10] R. Durrett, “Some features of the spread of epidemics and information on a random graph,” *Proc. Natl. Acad. Sci. USA* **107**, 4491–4498 (2010).
- [11] C. Castellano and R. Pastor-Satorras, “Thresholds for epidemic spreading in networks,” *Phys. Rev. Lett.* **105**, 218701 (2010).

- [12] J. P. Gleeson, “Binary-state dynamics on complex networks: Pair approximation and beyond,” *Phys. Rev. X* **3**, 021004 (2013).
- [13] 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.* **6**, 888–893 (2010).
- [14] M. E. J. Newman, “The spread of epidemic disease on networks,” *Phys. Rev. E* **66**, 016128 (2002).
- [15] M. Boguñá, C. Castellano, and R. Pastor-Satorras, “Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks,” *Phys. Rev. Lett.* **111**, 068701 (2013).
- [16] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras, “Collective versus hub activation of epidemic phases on networks,” *Phys. Rev. E* **93**, 032314 (2016).
- [17] A. S. Mata and S. C. Ferreira, “Multiple transitions of the susceptible-infected-susceptible epidemic model on complex networks,” *Phys. Rev. E* **91**, 012816 (2015).
- [18] G. F. de Arruda, E. Cozzo, T. P. Peixoto, F. A. Rodrigues, and Y. Moreno, “Disease localization in multilayer networks,” *Phys. Rev. X* **7**, 011014 (2017).
- [19] G. St-Onge, J.-G. Young, E. Laurence, C. Murphy, and L. J. Dubé, “Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks,” *Phys. Rev. E* **97**, 022305 (2018).
- [20] C.-R. Cai, Z.-X. Wu, M. Z. Q. Chen, P. Holme, and J.-Y. Guan, “Solving the dynamic correlation problem of the susceptible-infected-susceptible model on networks,” *Phys. Rev. Lett.* **116**, 258301 (2016).
- [21] Z.-W. Wei and B.-H. Wang, “Susceptible-infected-susceptible model on networks with eigenvector localization,” *Phys. Rev. E* **101**, 042310 (2020).
- [22] C. Castellano and R. Pastor-Satorras, “Relating topological determinants of complex networks to their spectral properties: Structural and dynamical effects,” *Phys. Rev. X* **7**, 041024 (2017).
- [23] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, “The spread of low-credibility content by social bots,” *Nat. Commun.* **9**, 4787 (2018).

- [24] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Critical phenomena in complex networks," [Rev. Mod. Phys. 80, 1275–1335 \(2008\)](#).
- [25] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*, Collection Alea-Saclay: Monographs and Texts in Statistical Physics (Cambridge University Press, Cambridge, 1999).
- [26] 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 86, 041125 \(2012\)](#).
- [27] H. K. Lee, P.-S. Shim, and J. D. Noh, "Epidemic threshold of the susceptible-infected-susceptible model on complex networks," [Phys. Rev. E 87, 062812 \(2013\)](#).
- [28] P. Shu, W. Wang, M. Tang, and Y. Do, "Numerical identification of epidemic thresholds for susceptible-infected-recovered model on finite-size networks," [Chaos 25, 063104 \(2015\)](#).
- [29] C. Castellano and R. Pastor-Satorras, "Non-mean-field behavior of the contact process on scale-free networks," [Phys. Rev. Lett. 96, 038701 \(2006\)](#).
- [30] H. Hong, M. Ha, and H. Park, "Finite-size scaling in complex networks," [Phys. Rev. Lett. 98, 258701 \(2007\)](#).
- [31] C. Castellano and R. Pastor-Satorras, "Routes to thermodynamic limit on scale-free networks," [Phys. Rev. Lett. 100, 148701 \(2008\)](#).
- [32] S. C. Ferreira, R. S. Ferreira, C. Castellano, and R. Pastor-Satorras, "Quasistationary simulations of the contact process on quenched networks," [Phys. Rev. E 84, 1–7 \(2011\)](#).
- [33] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, "Heterogeneous pair-approximation for the contact process on complex networks," [New J. Phys. 16, 053006 \(2014\)](#).
- [34] W. Cota and S. C. Ferreira, "Optimized Gillespie algorithms for the simulation of markovian epidemic processes on large and heterogeneous networks," [Comput. Phys. Commun. 219, 303–312 \(2017\)](#).
- [35] W. Cota, A. S. Mata, and S. C. Ferreira, "Robustness and fragility of the susceptible-infected-susceptible epidemic models on complex networks," [Phys. Rev. E 98, 012310 \(2018\)](#).

- [36] D. Sornet, *Critical Phenomena in Natural Sciences*, Springer Series in Synergetics (Springer-Verlag, Berlin/Heidelberg, 2006).
- [37] U. C. Tauber, *Critical Dynamics* (Cambridge University Press, Cambridge, 2014) p. 511.
- [38] D. R. Chialvo, "Emergent complex neural dynamics," *Nat. Phys.* **6**, 744–750 (2010).
- [39] O. Kinouchi and M. Copelli, "Optimal dynamical range of excitable networks at criticality," *Nat. Phys.* **2**, 348–351 (2006).
- [40] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality: An explanation of the $1/f$ noise," *Phys. Rev. Lett.* **59**, 381–384 (1987).
- [41] J. Hidalgo, J. Grilli, S. Suweis, M. A. Muñoz, J. R. Banavar, and A. Maritan, "Information-based fitness and the emergence of criticality in living systems," *Proc. Natl. Acad. Sci. USA* **111**, 10095–10100 (2014).
- [42] R. B. Griffiths, "Nonanalytic behavior above the critical point in a random ising ferromagnet," *Phys. Rev. Lett.* **23**, 17–19 (1969).
- [43] 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* **93**, 032322 (2016).
- [44] W. Cota, G. Ódor, and S. C. Ferreira, "Griffiths phases in infinite-dimensional, non-hierarchical modular networks," *Sci. Rep.* **8**, 9144 (2018).
- [45] O. Sporns, *Networks of the Brain* (MIT Press, 2010).
- [46] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Phys. Rev. E* **66**, 035103 (2002).
- [47] G. Palla, I. J. Farkas, P. Pollner, I. Derenyi, and T. Vicsek, "Directed network modules," *New J. Phys.* **9**, 186 (2007).
- [48] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S.-M. Kim, and D. Eisenberg, "Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions," *Nucleic Acids Res.* **30**, 303–305 (2002).

- [49] 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. 8, 35 \(2019\)](#).
- [50] R. K. Garrett, "Echo chambers online?: Politically motivated selective exposure among internet news users," [J. Comput. Mediat. Commun. 14, 265–285 \(2009\)](#).
- [51] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," [Nature 453, 779 \(2008\)](#).
- [52] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," [Proc. Natl. Acad. Sci. USA 106, 21484–21489 \(2009\)](#).
- [53] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, *et al.*, "Seasonal transmission potential and activity peaks of the new influenza a (H1N1): a Monte Carlo likelihood analysis based on human mobility," [BMC Med. 7, 45 \(2009\)](#).
- [54] P. Bajardi, C. Poletto, J. J. Ramasco, M. Tizzoni, V. Colizza, and A. Vespignani, "Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic," [PLOS ONE 6, e16591 \(2011\)](#).
- [55] C. Poletto, M. F. Gomes, A. P. y Piontti, L. Rossi, L. Bioglio, D. L. Chao, I. M. L. Jr, M. E. Halloran, V. Colizza, and A. Vespignani, "Assessing the impact of travel restrictions on international spread of the 2014 west african Ebola epidemic," [Eurosurveillance 19, 20936 \(2014\)](#).
- [56] C. L. Althaus, "Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in west africa," [PLOS Curr. 6 \(2014\), 10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288](#).
- [57] J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, Q. Wu, Y. Liang, X. Wang, J. Yang, K. Sun, I. M. Longini, M. E. Halloran, P. Wu, B. J. Cowling, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, and H. Yu, "Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei

- province, China: a descriptive and modelling study," [Lancet Infect. Dis. 20](#), 793–802 (2020).
- [58] D. S. Candido, I. M. Claro, J. G. de Jesus, W. M. Souza, F. R. R. Moreira, S. Dellicour, T. A. Mellan, L. du Plessis, R. H. M. Pereira, F. C. S. Sales, E. R. Manuli, J. Thézé, L. Almeida, M. T. Menezes, C. M. Voloch, M. J. Fumagalli, T. M. Coletti, C. A. M. da Silva, M. S. Ramundo, M. R. Amorim, H. H. Hoeltgebaum, S. Mishra, M. S. Gill, L. M. Carvalho, L. F. Buss, C. A. Prete, J. Ashworth, H. I. Nakaya, P. S. Peixoto, O. J. Brady, S. M. Nicholls, A. Tanuri, Á. D. Rossi, C. K. V. Braga, A. L. Gerber, A. P. de C. Guimarães, N. Gaburo, C. S. Alencar, A. C. S. Ferreira, C. X. Lima, J. E. Levi, C. Granato, G. M. Ferreira, R. S. Francisco, F. Granja, M. T. Garcia, M. L. Moretti, M. W. Perroud, T. M. P. P. Castiñeiras, C. S. Lazari, S. C. Hill, A. A. de Souza Santos, C. L. Simeoni, J. Forato, A. C. Sposito, A. Z. Schreiber, M. N. N. Santos, C. Z. de Sá, R. P. Souza, L. C. Resende-Moreira, M. M. Teixeira, J. Hubner, P. A. F. Leme, R. G. Moreira, M. L. Nogueira, N. M. Ferguson, S. F. Costa, J. L. Proenca-Modena, A. T. R. Vasconcelos, S. Bhatt, P. Lemey, C.-H. Wu, A. Rambaut, N. J. Loman, R. S. Aguiar, O. G. Pybus, E. C. Sabino, and N. R. Faria, "Evolution and epidemic spread of SARS-CoV-2 in Brazil," [Science 369](#), 1255–1260 (2020).
- [59] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," [Nature 429](#), 180 (2004).
- [60] V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction–diffusion processes and metapopulation models in heterogeneous networks," [Nat. Phys. 3](#), 276 (2007).
- [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. 14](#), 391–395 (2017).
- [62] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Panos, and B. Steinegger, "A mathematical model for the spatiotemporal epidemic spreading of COVID19," [medRxiv](#) , 2020.03.21.20040022 (2020).
- [63] G. S. Costa, W. Cota, and S. C. Ferreira, "Outbreak diversity in epidemic waves propagating through distinct geographical scales," [Phys. Rev. Research 2](#), 043306 (2020).

- [64] 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* **10**, 041055 (2020).
- [65] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Phys. Rep.* **424**, 175–308 (2006).
- [66] G. Bianconi, *Multilayer Networks: Structure and Function* (OUP Oxford, 2018).
- [67] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, “ k -core organization of complex networks,” *Phys. Rev. Lett.* **96**, 040601 (2006).
- [68] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature* **393**, 440–442 (1998).
- [69] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
- [70] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, “Cut-offs and finite size effects in scale-free networks,” *Eur. Phys. J. B - Condens. Matter* **38**, 205–209 (2004).
- [71] 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* **79**, 036110 (2009).
- [72] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys. Rev. Lett.* **86**, 3200–3203 (2001), 0010317 .
- [73] D. H. Silva, S. C. Ferreira, W. Cota, R. Pastor-Satorras, and C. Castellano, “Spectral properties and the accuracy of mean-field approaches for epidemics on correlated power-law networks,” *Phys. Rev. Research* **1**, 033024 (2019).
- [74] E. Ravasz, “Hierarchical organization of modularity in metabolic networks,” *Science* **297**, 1551–1555 (2002).
- [75] E. Ravasz and A.-L. Barabási, “Hierarchical organization in complex networks,” *Phys. Rev. E* **67**, 026112 (2003).
- [76] J. L. Moreno and H. H. Jennings, “Statistics of social configurations,” *Sociometry* , 342–374 (1938).

- [77] P. Erdős and A. Rényi, "On random graphs, I," *Publ. Math.* **6**, 290–297 (1959).
- [78] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science* **286**, 509–512 (1999).
- [79] M. E. J. Newman, *Handbook of Graphs and Networks: From the Genome to the Internet* (Wiley New York, 2003).
- [80] A. Békéssy, "Asymptotic enumeration of regular matrices," *Stud. Sci. Math. Hungar.* **7**, 343–353 (1972).
- [81] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. Algorithms* **6**, 161–180 (1995).
- [82] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, "Generation of uncorrelated random scale-free networks," *Phys. Rev. E* **71**, 027103 (2005).
- [83] D. Landau, *A guide to Monte Carlo simulations in statistical physics* (Cambridge University Press, Cambridge, 2015).
- [84] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.* **22**, 403–434 (1976).
- [85] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem. Lett.* **81**, 2340–2361 (1977).
- [86] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Vol. 1 (Elsevier, 1992).
- [87] R. Pastor-Satorras, "Epidemic dynamics in finite size scale-free networks," *Phys. Rev. E* **65**, 035108 (2002).
- [88] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Dynamics and Control (OUP Oxford, 1992).
- [89] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*, Vol. 15 (Cambridge University Press, 2001).
- [90] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2011).

- [91] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press, 2004).
- [92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, New York, NY, USA, 2007).
- [93] G. Ódor, *Universality in Nonequilibrium Lattice Systems: Theoretical Foundations* (World Scientific, Singapore, 2007).
- [94] M. Henkel, H. Hinrichsen, and L. S. Oenning, *Non-Equilibrium Phase Transitions. Volume 1: Absorbing Phase Transitions*, Theoretical and Mathematical Physics (Springer Netherlands, Dordrecht, 2008).
- [95] 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* **94**, 042308 (2016).
- [96] T. Vojta, "Rare region effects at classical, quantum and nonequilibrium phase transitions," *J. Phys. A: Math. Gen.* **39**, R143–R205 (2006).
- [97] A. S. Mata and S. C. Ferreira, "Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks," *EPL (Europhys Lett.)* **103**, 48003 (2013).
- [98] P. G. Fennell, S. Melnik, and J. P. Gleeson, "Limitations of discrete-time approaches to continuous-time contagion dynamics," *Phys. Rev. E* **94**, 052125 (2016).
- [99] M. M. de Oliveira and R. Dickman, "How to simulate the quasistationary state," *Phys. Rev. E* **71**, 016129 (2005).
- [100] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, "Localization and spreading of diseases in complex networks," *Phys. Rev. Lett.* **109**, 128702 (2012).
- [101] R. Dickman and M. A. Burschka, "Nonequilibrium critical poisoning in a single-species model," *Phys. Lett. A* **127**, 132–137 (1988).
- [102] L. Böttcher, H. J. Herrmann, and M. Henkel, "Dynamical universality of the contact process," *J. Phys. A: Math. Theor.* **51**, 125003 (2018).

- [103] R. S. Sander, M. M. de Oliveira, and S. C. Ferreira, "Quasi-stationary simulations of the directed percolation universality class in $d = 3$ dimensions," *J. Stat. Mech. Theory Exp.* **2009**, P08011 (2009).
- [104] M. M. de Oliveira, S. G. Alves, and S. C. Ferreira, "Continuous and discontinuous absorbing-state phase transitions on voronoi-delaunay random lattices," *Phys. Rev. E* **93**, 012110 (2016).
- [105] T. E. Harris, "Contact interactions on a lattice," *Ann. Probab.* **2**, 969–988 (1974).
- [106] S. H. Strogatz, *Nonlinear Dynamics And Chaos*, Studies in nonlinearity (Sarat Book House, 2007).
- [107] M. Boguñá and R. Pastor-Satorras, "Epidemic spreading in correlated complex networks," *Phys. Rev. E* **66**, 047104 (2002).
- [108] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Dynamical and correlation properties of the internet," *Phys. Rev. Lett.* **87**, 258701 (2001).
- [109] J. A. Hołyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki, "Universal scaling of distances in complex networks," *Phys. Rev. E* **72**, 026108 (2005).
- [110] C. Castellano and R. Pastor-Satorras, "Competing activation mechanisms in epidemics on networks," *Sci. Rep.* **2**, 371 (2012).
- [111] F. Radicchi and C. Castellano, "Breaking of the site-bond percolation universality in networks," *Nat. Commun.* **6**, 10196 (2015).
- [112] J. M. Beggs and D. Plenz, "Neuronal avalanches in neocortical circuits." *J. Neurosci.* **23**, 11167–77 (2003).
- [113] A. Haimovici, E. Tagliazucchi, P. Balenzuela, and D. R. Chialvo, "Brain organization into resting state networks emerges at criticality on a model of the human connectome," *Phys. Rev. Lett.* **110**, 178101 (2013).
- [114] D. Plenz, E. Niebur, and H. G. Schuster, *Criticality in Neural Systems*, Annual Reviews of Nonlinear Dynamics and Complexity (VCH) (Wiley, 2014).
- [115] M. A. Muñoz, R. Juhász, C. Castellano, and G. Ódor, "Griffiths phases on complex networks," *Phys. Rev. Lett.* **105**, 128701 (2010).

- [116] R. Juhász, G. Ódor, C. Castellano, and M. Muñoz, “Rare-region effects in the contact process on networks,” [Phys. Rev. E **85**, 1–13 \(2012\)](#).
- [117] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, “Epidemic spreading in real networks: an eigenvalue viewpoint,” in [22nd Int. Symp. Reliab. Distrib. Syst. 2003. Proceedings.](#) (IEEE Comput. Soc, Los Alamitos, CA, USA, 2003) pp. 25–34.
- [118] G. Ódor, “Localization transition, Lifschitz tails, and rare-region effects in network models,” [Phys. Rev. E **90**, 032110 \(2014\)](#).
- [119] G. Ódor and R. Pastor-Satorras, “Slow dynamics and rare-region effects in the contact process on weighted tree networks,” [Phys. Rev. E **86**, 026117 \(2012\)](#).
- [120] G. Ódor, “Rare regions of the susceptible-infected-susceptible model on Barabasi-Albert networks,” [Phys. Rev. E **87**, 042132 \(2013\)](#).
- [121] G. Ódor, “Spectral analysis and slow spreading dynamics on complex networks,” [Phys. Rev. E **88**, 032109 \(2013\)](#).
- [122] C. Buono, F. Vazquez, P. A. Macri, and L. A. Braunstein, “Slow epidemic extinction in populations with heterogeneous infection rates,” [Phys. Rev. E **88**, 022813 \(2013\)](#).
- [123] R. Pastor-Satorras and C. Castellano, “Distinct types of eigenvector localization in networks,” [Sci. Rep. **6**, 18847 \(2016\)](#).
- [124] R. Dickman, “Critical exponents for the restricted sandpile,” [Phys. Rev. E **73**, 036131 \(2006\)](#).
- [125] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” [Nat. Rev. Neurosci. **10**, 312–312 \(2009\)](#).
- [126] D. Meunier, R. Lambiotte, and E. T. Bullmore, “Modular and hierarchically modular organization of brain networks,” [Front. Neurosci. **4**, 1–11 \(2010\)](#).
- [127] P. Moretti and M. A. Muñoz, “Griffiths phases and the stretching of criticality in brain networks,” [Nat. Commun. **4**, 2521 \(2013\)](#).
- [128] G. Ódor, R. Dickman, and G. Ódor, “Griffiths phases and localization in hierarchical modular networks,” [Sci. Rep. **5**, 14451 \(2015\)](#).

- [129] S. Li, “Griffiths phase on hierarchical modular networks with small-world edges,” [Phys. Rev. E](#) **95**, 032306 (2017).
- [130] C. C. Hilgetag and A. Goulas, “Is the brain really a small-world network?” [Brain Struct. Func.](#) **221**, 2361–2366 (2016).
- [131] M. E. J. Newman and D. J. Watts, “Renormalization group analysis of the small-world network model,” [Phys. Lett. A](#) **263**, 341–346 (1999).
- [132] A. Barrat and M. Weigt, “On the properties of small-world network models,” [Eur. Phys. J. B](#) **13**, 547–560 (2000).
- [133] A. Lancichinetti, S. Fortunato, and F. Radicchi, “Benchmark graphs for testing community detection algorithms,” [Phys. Rev. E](#) **78**, 046110 (2008).
- [134] M. Bastian, S. Heymann, M. Jacomy, *et al.*, “Gephi: an open source software for exploring and manipulating networks.” [ICWSM](#) **8**, 361–362 (2009).
- [135] R. S. Ferreira and S. C. Ferreira, “Critical behavior of the contact process on small-world networks,” [Eur. Phys. J. B](#) **86**, 462 (2013).
- [136] G. Ódor, “Universality classes in nonequilibrium lattice systems,” [Rev. Mod. Phys.](#) **76**, 663–724 (2004).
- [137] A. G. Moreira and R. Dickman, “Critical dynamics of the contact process with quenched disorder,” [Phys. Rev. E](#) **54**, R3090–R3093 (1996).
- [138] T. Vojta, J. Igo, and J. a. Hoyos, “Rare regions and griffiths singularities at a clean critical point: The five-dimensional disordered contact process,” [Phys. Rev. E](#) **90**, 012139 (2014).
- [139] M. A. Muñoz, R. Dickman, A. Vespignani, and S. Zapperi, “Avalanche and spreading exponents in systems with absorbing states,” [Phys. Rev. E](#) **59**, 6175 (1999).
- [140] J. M. Beggs and N. Timme, “Being critical of criticality in the brain,” [Front. Physiol.](#) **3**, 1–14 (2012).
- [141] A. J. Fontenele, N. A. P. de Vasconcelos, T. Feliciano, L. A. A. Aguiar, C. Soares-Cunha, B. Coimbra, L. Dalla Porta, S. Ribeiro, A. J. a. Rodrigues, N. Sousa, P. V. Carelli, and M. Copelli, “Criticality between cortical states,” [Phys. Rev. Lett.](#) **122**, 208101 (2019).

- [142] M. T. Gastner and G. Ódor, "The topology of large open connectome networks for the human brain," *Sci. Rep.* **6**, 27249 (2016).
- [143] H. Masum, C. Newmark, and M. Tovey, *The Reputation Society: How Online Opinions are Reshaping the Offline World*, Information society series (MIT Press, Cambridge, MA, 2011).
- [144] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, 1994).
- [145] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, and Others, "Life in the network: the coming age of computational social science," *Science (New York, NY)* **323**, 721 (2009).
- [146] J. K. Lee, J. Choi, C. Kim, and Y. Kim, "Social media, network heterogeneity, and opinion polarization," *J. Commun.* **64**, 702–722 (2014).
- [147] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature* **489**, 295–298 (2012).
- [148] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proc. Natl. Acad. Sci. USA* **106**, 21544–21549 (2009).
- [149] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer, "Political polarization on Twitter," in *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2011).
- [150] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, "Partisan asymmetries in online political activity," *EPJ Data Sci.* **1**, 6 (2012).
- [151] A. Hanna, C. Wells, P. Maurer, L. Friedland, D. Shah, and J. Matthes, "Partisan alignments and political polarization online: A computational approach to understanding the French and US presidential elections," in *Proceedings of the 2nd Workshop on Politics, Elections and Data*, PLEAD '13 (ACM, New York, NY, USA, 2013) pp. 15–22.
- [152] I. Weber, V. R. K. Garimella, and A. Batayneh, "Secular vs. islamist polarization in Egypt on Twitter," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in*

- Social Networks Analysis and Mining - ASONAM '13* (ACM Press, New York, New York, USA, 2013) pp. 290–297.
- [153] J. Borge-Holthoefer, W. Magdy, K. Darwish, and I. Weber, “Content and network dynamics behind egyptian political polarization on Twitter,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15 (ACM, New York, NY, USA, 2015) pp. 700–711.
- [154] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, “The dynamics of protest recruitment through an online network,” *Sci. Rep.* **1**, 197 (2011).
- [155] M. T. Bastos, D. Mercea, and A. Baronchelli, “The geographic embedding of online echo chambers: Evidence from the Brexit campaign,” *PLOS ONE* **13**, 1–16 (2018).
- [156] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, “The spreading of misinformation online,” *Proc. Natl. Acad. Sci. USA* **113**, 554–559 (2016).
- [157] M. D. Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, “Echo chambers: Emotional contagion and group polarization on Facebook,” *Sci. Rep.* **6**, 37825 (2016).
- [158] M. Wojcieszak, “‘don’t talk to me’: Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism,” *New Media Soc.* **12**, 637–655 (2010).
- [159] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, “Tweeting from left to right: Is online political communication more than an echo chamber?” *Psychol. Sci.* **26**, 1531–1542 (2015), pMID: 26297377.
- [160] E. Dubois and G. Blank, “The echo chamber is overstated: the moderating effect of political interest and diverse media,” *Inf. Commun. Soc.* **21**, 729–745 (2018).
- [161] Twitter Inc., “Introduction to tweet json - retweets and quote tweets,” <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/intro-to-tweet-json.html> (2018), [Online; accessed 01-August-2018].
- [162] P. Cooper, “25 Twitter stats all marketers need to know in 2020,” <https://blog.hootsuite.com/twitter-statistics/> (2019), [Online; accessed 10-September-2020].

- [163] Wikipedia, “Impeachment of Dilma Rousseff,” https://pt.wikipedia.org/wiki/Processo_de_impeachment_de_Dilma_Rousseff (2018), [Online; accessed 01-October-2018].
- [164] R. McGrath, “Twython 3.4.0,” <https://github.com/ryanmcgrath/twython> (2016).
- [165] M. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, “Predicting the political alignment of Twitter users,” in *Proceedings of 3rd IEEE Conference on Social Computing (SocialCom)* (2011) pp. 192–199.
- [166] D. M. Romero, B. Meeder, and J. Kleinberg, “Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter,” in *Proceedings of the 20th International Conference on World Wide Web, WWW ’11* (ACM, New York, NY, USA, 2011) pp. 695–704.
- [167] A. Bessi, F. Zollo, M. Del Vicario, M. Puliga, A. Scala, G. Caldarelli, B. Uzzi, and W. Quattrociocchi, “Users polarization on Facebook and Youtube,” *PLOS ONE* **11**, 1–24 (2016).
- [168] P. A. Grabowicz, J. J. Ramasco, E. Moro, J. M. Pujol, and V. M. Eguiluz, “Social features of online networks: The strength of intermediary ties in online social media,” *PLOS ONE* **7**, 1–9 (2012).
- [169] V. Nicosia, J. Tang, C. Mascolo, M. Musolesi, G. Russo, and V. Latora, “Graph metrics for temporal networks,” in *Temporal Networks* (Springer, 2013) pp. 15–40.
- [170] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, “The rise of social bots,” *Commun. ACM* **59**, 96–104 (2016).
- [171] M. Stella, E. Ferrara, and M. De Domenico, “Bots increase exposure to negative and inflammatory content in online social systems,” *Proc. Natl. Acad. Sci. USA* **115**, 12435–12440 (2018).
- [172] R. Tarjan, “Depth-first search and linear graph algorithms,” *SIAM J. Comput.* **1**, 146–160 (1972).
- [173] E. Nuutila and E. Soisalon-Soininen, “On finding the strongly connected components in a directed graph,” *Inf. Process. Lett.* **49**, 9–14 (1994).
- [174] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, “Forceatlas2, a continuous graph

- layout algorithm for handy network visualization designed for the Gephi software," *PLOS ONE* **9**, e98679 (2014).
- [175] S. Fortunato, "Community detection in graphs," *Phys. Rep.* **486**, 75–174 (2010).
- [176] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer, "Outtweeting the Twitterers - predicting information cascades in microblogs," in *Proceedings of the 3rd Wconference on Online Social Networks*, WOSN'10 (USENIX Association, Berkeley, CA, USA, 2010) pp. 3–3.
- [177] M. Jenders, G. Kasneci, and F. Naumann, "Analyzing and predicting viral tweets," in *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion* (ACM Press, New York, New York, USA, 2013) pp. 657–664.
- [178] J. Ratkiewicz, M. D. Conover, M. Meiss, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in *In Proceedings of the 5th AAAI International Conference on Weblogs and Social Media (ICWSM'11)* (2011).
- [179] L. Zhao, H. Cui, X. Qiu, X. Wang, and J. Wang, "SIR rumor spreading model in the new media age," *Physica A* **392**, 995–1003 (2013).
- [180] C. Granell, S. Gómez, and A. Arenas, "Dynamical interplay between awareness and epidemic spreading in multiplex networks," *Phys. Rev. Lett.* **111**, 128701 (2013).
- [181] M. Karsai, G. Iniguez, K. Kaski, and J. Kertesz, "Complex contagion process in spreading of online innovation," *J. R. Soc. Interface* **11**, 20140694 (2014).
- [182] P. Holme, "Network reachability of real-world contact sequences," *Phys. Rev. E* **71**, 46119 (2005).
- [183] D. J. Watts, "A simple model of global cascades on random networks," *Proc. Natl. Acad. Sci. USA* **99**, 5766–5771 (2002).
- [184] D. Centola and M. Macy, "Complex contagions and the weakness of long ties," *Am. J. Sociol.* **113**, 702–734 (2007).
- [185] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," in *Proceedings of the 12th International Conference on*

- Knowledge-Based Intelligent Information and Engineering Systems, Part III*, KES '08 (Springer-Verlag, Berlin, Heidelberg, 2008) pp. 67–75.
- [186] A. Borodin, Y. Filmus, and J. Oren, “Threshold models for competitive influence in social networks,” in *Proceedings of the 6th International Conference on Internet and Network Economics*, WINE'10 (Springer-Verlag, Berlin, Heidelberg, 2010) pp. 539–550.
- [187] F. Karimi and P. Holme, “Threshold model of cascades in empirical temporal networks,” *Physica A* **392**, 3476–3483 (2013).
- [188] R. Guimera, S. Mossa, A. Turtleschi, and L. N. Amaral, “The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles,” *Proc. Natl. Acad. Sci. USA* **102**, 7794–7799 (2005).
- [189] S. Bansal, G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud, “Big data for infectious disease surveillance and modeling,” *J. Infect. Dis.* **214**, S375–S379 (2016).
- [190] M. E. Halloran, A. Vespignani, N. Bharti, L. R. Feldstein, K. A. Alexander, M. Ferrari, J. Shaman, J. M. Drake, T. Porco, J. N. S. Eisenberg, *et al.*, “Ebola: mobility data,” *Science* **346**, 433–433 (2014).
- [191] D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, “Spreading processes in multiplex metapopulations containing different mobility networks,” *Phys. Rev. X* **8**, 031039 (2018).
- [192] L. Sattenspiel and K. Dietz, “A structured epidemic model incorporating geographic mobility among regions,” *Math. Biosci.* **128**, 71–91 (1995).
- [193] B. Grenfell and J. Harwood, “(meta) population dynamics of infectious diseases,” *Trends Ecol. Evol.* **12**, 395–399 (1997).
- [194] V. Colizza and A. Vespignani, “Invasion threshold in heterogeneous metapopulation networks,” *Phys. Rev. Lett.* **99**, 148701 (2007).
- [195] V. Colizza and A. Vespignani, “Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations,” *J. Theor. Biol.* **251**, 450–467 (2008).

- [196] A. S. Mata, S. C. Ferreira, and R. Pastor-Satorras, “Effects of local population structure in a reaction-diffusion model of a contact process on metapopulation networks,” *Phys. Rev. E* **88**, 042820 (2013).
- [197] D. H. Silva and S. C. Ferreira, “Activation thresholds in epidemic spreading with motile infectious agents on scale-free networks,” *Chaos* **28**, 123112 (2018).
- [198] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, “A universal model for mobility and migration patterns,” *Nature* **484**, 96–100 (2012).
- [199] F. Simini, A. Maritan, and Z. Nédá, “Human mobility in a continuum approach,” *PLOS ONE* **8**, e60069 (2013).
- [200] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, “Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows,” *Phys. Rev. E* **88**, 022812 (2013).
- [201] D. Balcan and A. Vespignani, “Phase transitions in contagion processes mediated by recurrent mobility patterns,” *Nat. Phys.* **7**, 581 (2011).
- [202] V. Belik, T. Geisel, and D. Brockmann, “Natural human mobility patterns and spatial spread of infectious diseases,” *Phys. Rev. X* **1**, 011001 (2011).
- [203] D. Balcan and A. Vespignani, “Invasion threshold in structured populations with recurrent mobility patterns,” *J. Theor. Biol.* **293**, 87–100 (2012).
- [204] V. Belik, T. Geisel, and D. Brockmann, “Recurrent host mobility in spatial epidemics: beyond reaction-diffusion,” *Eur. Phys. J. B* **84**, 579–587 (2011).
- [205] E. W. Weisstein, “Gershgorin circle theorem,” <https://mathworld.wolfram.com/GershgorinCircleTheorem.html> (2003), [Online; accessed 01-February-2020].
- [206] G. Chowell, J. M. Hyman, S. Eubank, and C. Castillo-Chavez, “Scaling laws for the movement of people between locations in a large city,” *Phys. Rev. E* **68**, 066102 (2003).
- [207] R. Guimera, S. Mossa, A. Turtshi, and L. A. N. Amaral, “The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles,” *Proc. Natl. Acad. Sci. USA* **102**, 7794–7799 (2005).

- [208] R. Patuelli, A. Reggiani, S. P. Gorman, P. Nijkamp, and F.-J. Bade, "Network analysis of commuting flows: A comparative static approach to german data," *Netw. Spat. Econ.* **7**, 315–331 (2007).
- [209] F. Ball, T. Britton, T. House, V. Isham, D. Mollison, L. Pellis, and G. S. Tomba, "Seven challenges for metapopulation models of epidemics, including households models," *Epidemics* **10**, 63–67 (2015).
- [210] D. Soriano-Paños, G. Ghoshal, A. Arenas, and J. Gómez-Gardeñes, "Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics," *J. Stat. Mech: Theory Exp.* **2020**, 024006 (2020).
- [211] D. Soriano-Paños, H. Arias-Castro, F. Naranjo-Mayorga, and J. Gómez-Gardeñes, "Impact of human-human contagions in the spread of vector-borne diseases," *Eur. Phys. J. Spec. Top.* **227**, 661–672 (2018).
- [212] D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni, and J. Gómez-Gardeñes, "Vector-borne epidemics driven by human mobility," *Phys. Rev. Research* **2**, 013312 (2020).
- [213] B. M. Althouse, E. A. Wenger, J. C. Miller, S. V. Scarpino, A. Allard, L. Hébert-Dufresne, and H. Hu, "Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control," *PLOS Biology* **18**, 1–13 (2020).
- [214] G. St-Onge, J.-G. Young, L. Hébert-Dufresne, and L. J. Dubé, "Efficient sampling of spreading processes on complex networks using a composition and rejection algorithm," *Comput. Phys. Commun.* **240**, 30–37 (2019).
- [215] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, "Modeling echo chambers and polarization dynamics in social networks," *Phys. Rev. Lett.* **124**, 048301 (2020).
- [216] S. Fortunato and M. Barthélemy, "Resolution limit in community detection," *Proc. Natl. Acad. Sci. USA* **104**, 36–41 (2007).
- [217] R. A. Marcus, "Brief comments on perturbation theory of a nonsymmetric matrix: The GF matrix," *J. Phys. Chem. A* **105**, 2612–2616 (2001).
- [218] K. Žáková, "Maxima - an open alternative for engineering education," in *2011 IEEE Global Engineering Education Conference (EDUCON)* (2011) pp. 1022–1025.