UNIVERSIDADE FEDERAL DE VIÇOSA

DIOGO HENRIQUE DA SILVA

LOCALIZATION PHENOMENA AND LIMITS OF MEAN-FIELD THEORIES IN EPIDEMIC PROCESSES ON NETWORKS

VIÇOSA - MINAS GERAIS
2020

DIOGO HENRIQUE DA SILVA

LOCALIZATION PHENOMENA AND LIMITS OF MEAN-FIELD THEORIES IN EPIDEMIC PROCESSES ON NETWORKS

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

Adviser: Silvio da Costa Ferreira Junior

VIÇOSA - MINAS GERAIS
2020

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

T

S586l
2020

Silva, Diogo Henrique da, 1989-
Localization phenomena and limits of mean-field theories in epidemic spreading processes on complex networks / Diogo Henrique da Silva. - Viçosa, MG, 2020.
143 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.
Orientador: Silvio da Costa Ferreira Júnior.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 132-143.

1. Teoria das redes complexas (Física). 2. Teoria de campo médio. 3. Transformações de fase (Física estatística). I. Universidade Federal de Viçosa. Departamento de Física. Programa de Pós-Graduação em Física Aplicada. II. Título.

CDD 22. ed. 530.1595

DIOGO HENRIQUE DA SILVA

LOCALIZATION PHENOMENA AND LIMITS OF MEAN-FIELD THEORIES IN EPIDEMIC PROCESSES ON NETWORKS

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

APPROVED: December 10, 2020.

Assent:

Diogo Henrique da Silva
Author

Silvio da Costa Ferreira Junior
Adviser

To my family.

ACKNOWLEDGMENTS

I express my acknowledgment to everyone who participated of my journey to achieve this goal of my academic and personal formation. I acknowledge my relatives, in particular, my parents who have been supporting my choices during these years and my brother Brasilino who shared with me my years in this university. I acknowledge my friends represented here by Nathann, Ricardo and Ismael.

I acknowledge the research group partners, Wesley and Guilherme, for the collaborations. I acknowledge Professor Silvio (adviser) for sharing with me his knowledge and for allowing me to be part of the team led by him. I acknowledge Professors Francisco Rodrigues, Romualdo Pastor-Satorras, and Claudio Castellano for the collaborations in the articles used as basis of this thesis.

I acknowledge GISC, NISC, Universidade Federal de Viçosa and its Physics Department for providing the conditions to perform this thesis. I am grateful by the financial support from the agencies *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG), *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) and its program *Ciência sem fronteiras*. This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, (Brazil) Grant, No.0.001.

ABSTRACT

DA SILVA, Diogo Henrique, D.Sc., Universidade Federal de Viçosa, December, 2020. **Localization phenomena and limits of mean-field theories in epidemic processes on networks**. Adviser: Silvio da Costa Ferreira Junior.

Complex networks have been applied to represent many real systems and investigations of dynamical processes on their top are of interest, in particular, the epidemic model susceptible-infected-susceptible (SIS). In this model, theoretical descriptions of the transition from a disease-free to an endemic phase at the epidemic threshold are usually performed by means of mean-field theories. The quenched mean-field theory (QMF) takes into account the network structure regarding dynamical correlations. The dynamical correlations are added to QMF theory in a pairwise level in the pair-quenched mean-field theory (PQMF). We verify that, as in the QMF case, the PQMF theory can be described by the spectral properties of a Jacobian matrix which emerges within this theory. The absence of degree correlations, which allows to simplify these theories, has been considered in many studies. We analyze the effects of degree correlations on the performance of mean-field theories in determining the epidemic threshold and prevalence of the SIS model on real and synthetic networks. We investigate if there is a relation between this performance and structural and spectral properties of the network and matrices associated with the respective theories. Usually, localization in dynamical processes is investigated through eigenvectors of matrices associated to theoretical approaches near to the transition. We study this problem introducing a normalized activity vector determined by the nodal activities. We construct basis for interpreting the localization inherent to epidemic processes at the threshold and the onset of a delocalized endemic phase just above it. The method is generic and applicable to theories, stochastic simulations, real data and any network.

Keywords: Complex networks. Spreading dynamics. Mean-field theories.

RESUMO

DA SILVA, Diogo Henrique, D.Sc., Universidade Federal de Viçosa, dezembro de 2020. **Fenômenos de localização e limites de teorias de campo médio para processos epidêmicos em redes**. Orientador: Silvio da Costa Ferreira Júnior.

Sistemas reais podem ser representados em redes complexas e é de interesse investigar processos dinâmicos nessas redes, dentre os quais destacamos o modelo epidêmico suscetível-infectado-suscetível (SIS). Nesse modelo, as descrições teóricas da transição de um estado livre de doença para um endêmico a uma dada taxa de infecção, chamada de limiar epidêmico, são comumente realizadas a partir das teorias de campo médio. Destacamos a teoria QMF do inglês *quenched mean-field*, que considera toda a estrutura da rede desprezando as correlações dinâmicas, e a teoria PQMF do inglês *pair-quenched mean-field*, que adiciona correlações dinâmicas à teoria QMF em um nível de pares. Verificamos que, assim como na teoria QMF, a teoria PQMF pode ser descrita pelas propriedades espectrais de uma matriz Jacobiana associada à ela. Em geral, essas teorias são estudadas na ausência de correlação de grau permitindo simplificações. Analisamos os efeitos da correlação de grau, tanto em redes sintéticas quanto em redes reais, na precisão das teorias de campo médio em prever o limiar epidêmico e a prevalência epidêmica. Além disso, verificamos se há relação da precisão das teorias com as propriedades estruturais das redes e espectrais das matrizes associadas às teorias. A localização em espalhamentos dinâmicos é usualmente estudada explorando propriedades espectrais de matrizes associadas a descrições teóricas próximas a transição. Analisamos a localização a partir do vetor de atividade normalizado definido em termos da atividade dos nós da rede. Construímos bases para interpretação da localização inerente aos processos epidêmicos no limiar e em uma fase endêmica não localizada logo acima dele. O método é aplicável as simulações estocásticas, teorias, dados reais e redes distintas.

Palavras-chave: Redes complexas. Dinâmica de espalhamento. Campo médio.

LIST OF ILLUSTRATION

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

CM − Configuration model

CP − Contact process

HMF − Heterogeneous mean-field

IPR − Inverse participation ratio

LEV − Largest eigenvalue

MF − Mean-field

NAV − Normalized activity vector

NAVCPD − Normalized activity vector component probability distribution

PEV − Principal eigenvector

PHMF− Pair-homogeneous mean-field

PLA − Preferential linking attachment

PQMF − Pair-quenched mean-field

QMF − Quenched mean-field

QS − Quasistationary

RRN − Random regular network

SEIR − Susceptible-exposed-infected-recovered

SEIRS − Susceptible-exposed-infected-recovered-susceptible

SI − Susceptible-infected

SIR − Susceptible-infected-recovered

SIRS − Susceptible-infected-recovered-susceptible

SIS − Susceptible-infected-susceptible

UCM − Uncorrelated configuration model

WPCM − Weber and Porto configuration model

LIST OF PUBLICATIONS

- Spectral properties and the accuracy of mean-field approaches for epidemics on correlated power-law networks

  Authors: **Diogo H. Silva**, Silvio C. Ferreira, Wesley Cota, Romualdo Pastor-Satorras and Claudio Castellano

  Physical Review Research 1, 033024 (2019); DOI: 10.1103/PhysRevResearch.1.033024

  *Chapter 5 is based on this reference.

- High prevalence regimes in the pair-quenched mean-field theory for the susceptible-infected-susceptible model on networks

  Authors: **Diogo H. Silva**, Francisco A. Rodrigues and Silvio C. Ferreira

  Physical Review E 102, 012313 (2020); DOI: 10.1103/PhysRevE.102.012313

  *Chapter 6 is based on this reference.

- Dissecting localization phenomena of dynamical processes on networks

  Authors: **Diogo H. Silva** and Silvio C. Ferreira

  J. Phys. Complex; DOI:10.1088/2632-072X/abdd98

  *Chapter 7 is based on this reference.

- Activation thresholds in epidemic spreading with motile infectious agents on scale-free networks

  Authors: **Diogo H. Silva** and Silvio C. Ferreira

  Chaos: An Interdisciplinary Journal of Nonlinear Science 28, 123105 (2018); DOI:10.1063/1.5053911

  *This study was started during my master degree concluded in 2016 and written, revised and concluded in my Phd.

SUMMARY

Chapter 1

Introduction

Technological advances have provided tools required to study problems presenting a large number of interacting elements. Some of them can present a self-organized structure which can evolve in time, resisting to random attacks, propagating fluctuations and heterogeneity at all scales, being a complex system [1]. Many complex systems in distinct fields have been mapped onto complex networks considering that each component of the system is a node of the network and an edge connecting two nodes represents the interaction between them. In biology, protein-protein interaction forms a network which allows to represent the yeast protein-protein interactome [2] and provides an environment to analyze new treatments which could be employed in medicine [3]. In the infrastructure, the power-grid network is capable of rearrange the transmission lines across large distances, allowing to protect the system against failures [4,5], indicating for example, points where maintenance would be done, improving the applications of resources. The analysis of the face-to-face interaction applying radio frequency identification (RFID) [6] or the interaction among members of the social microblogging Twitter [7–10] by networks highlight their relevance in describing the social interactions in physical and virtual environments.

The mapping of the Internet [11] into a network demands the introduction of some quantities which allow its characterization. These quantities allow to establish a connection between network structure and system features. Nodes of a complex network can show different number of neighbors (degree of a node) and the degree

distribution depicts this characteristic. The Internet presents a power-law degree distribution [12], $P(k) \sim k^{-\gamma}$. This property is shared by a large number of networks representing real systems [1]. Usually, the formation of an edge connecting two nodes in real networks is biased. Technological networks, such as the Internet, show in general disassortative mixing [12,13], i.e., nodes of large degree tend to be connected with those of small degree, and vice-versa. Assortative mixing occurs in social networks, in which connections preferentially occur among nodes exhibiting similar degree. Uncorrelated networks are typical benchmarks for the investigation of dynamical processes on networks and they were the structure considered in many studies [1,14,15]. The shortest path length and the clustering coefficient are quantities capable of reproducing features of many social networks [16]. A path is a route connecting nodes passing through edges of the network and the shortest one is remarkable, since it allows to compute the average shortest distance that increases logarithmically with the size of the network and is associated with the small-world effects [1]. The average clustering coefficient is associated with closed routes, in particular, the formation of triangles [1]. The relevance of a node is determined by some criterion which defines a centrality. For example, in degree centrality the relevance of a node is determined by its degree. The role of structural properties in spreading processes running on the top of complex networks has been target of several investigations [17–19]. A controlled framework is required and network models play a key role in these investigations. The preferential-attachment model reproduces the power-law degree distribution exploring the evolutionary behavior observed in the internet [20]. The relation between average clustering coefficient and average shortest path in social networks is captured by the Watts-Strogatz model [16]. A brief review of network properties and some models used in this thesis to generate synthetic networks are presented in chapters 2 and 3.

Among dynamical processes running on the top of complex networks, the epidemic spreading has been investigated considering models [1,21–23], in which an individual assumes states corresponding to a disease stage and transitions between states occur with a rate which can be estimated, for example, considering databases provided by health agencies [21–23]. Epidemic models allow to study effects of specified features

of a disease such as the dynamics of the vectors *Aedes aegypti* and *Aedes albopictus*, seasonal effects in Zika virus spreading in America [22], or the Ebola virus transmission in funerals [21].

In this thesis, our attention is devoted to the susceptible-infected-susceptible model (SIS). In this model, an individual can be reinfected multiple times as in diseases such as flu [24] and malaria [25]. The SIS model presents a transition from a disease-free (absorbing state) to an endemic phase and the infection rate for which it occurs is called epidemic threshold. The model was simulated in this work using the optimized Gillespie scheme described in Refs. [26,27]. To circumvent the difficulties of dealing with the absorbing state in finite-size networks, we implement the quasistationary simulations (QS) [28], in which the dynamics returns to a previously visited active configuration whenever the absorbing states is visited.

Theoretical descriptions of the SIS transition are performed considering mean-field theories [14]. In homogeneous mean-field the dynamical and topological features are replaced by their average, respectively. However, heterogeneities on complex networks can change drastically the behavior of the epidemic spreading. In particular, for networks presenting a power-law degree distribution, $P(k) \sim k^{-\gamma}$, and absence of degree correlation, localization takes place in the SIS transition and different set of nodes triggers the activation of the epidemic in the network for $\gamma < 2.5$ and $\gamma > 2.5$ [18,29,30], and the epidemic threshold is null regardless of the value of γ [31,32]. Heterogeneities can be included in mean-field approximations in different forms [14]. Two widely used approximations are the heterogeneous mean-field (HMF) [33–35] and the quenched mean-field (QMF) [36,37] theories. The former consists of a cross-graining where only the degree of the nodes and the statistical degree correlations are included in the dynamical equations for the probability that a node is infected [33,38] and neglects the dynamical correlations. The latter includes the full connectivity structure of the networks but still neglects dynamical correlations assuming that states of nearest-neighbors are independent. A further quantitative improvement of the QMF theory has been achieved in Ref. [39] by means of the explicit inclusion of pairwise dynamical correlations [40–42], hereafter called of pair QMF (PQMF) theory. These theories can

be handled using linear stability analysis [14,43] where the behavior of the transition near to the threshold can be obtained from the spectral properties of matrices related to the network structure [44,45]. However, the mechanism of long-term epidemic activity in subgraphs of the network and a long-range interaction among them provides a rigorous mathematical description for the SIS dynamics on uncorrelated networks with power-law degree distributions [31,32]. A review of the SIS model presenting the main results, theoretical descriptions, and simulation methods is done in chapter 4.

Since absence of correlations usually simplifies theoretical approaches, they are typical benchmarks for the investigation of dynamical processes on networks and have been considered in many studies [1,14,15]. However, the ubiquitousness of correlations in real networks naturally calls for the investigation of the effect of correlated interaction patterns. While the effects of degree correlations have been considered for several dynamical processes [38,46], a full understanding of their effects on the performance of theoretical approaches is still missing. In chapter 5, we present a comparison between stochastic simulations and mean-field theories for the epidemic threshold of the SIS model on correlated networks (both assortative and disassortative) with a power-law degree distribution $P(k) \sim k^{-\gamma}$. The main results of this chapter were published in Ref. [47]. We confirm the vanishing of the threshold regardless of the correlation pattern and the degree exponent γ. Thresholds determined in stochastic simulations are compared with QMF and PQMF theories. Degree correlations do not change the overall picture: The QMF and PQMF theories provide estimates that are asymptotically correct for $\gamma < 5/2$, while they only capture the vanishing of the threshold for $\gamma > 5/2$, failing to reproduce quantitatively how this occurs. For a given size, PQMF theory is more accurate. We relate the variations in the accuracies of QMF and PQMF predictions with the spectral properties (spectral gap and localization) of standard and modified adjacency matrices, which rule the epidemic prevalence near the transition point, depending on the theoretical framework.

We extended our analysis to the prevalence (order parameter) as a function of the infection rate (control parameter) in chapter 6, whose main results were published in Ref. [48]. We perform a nonperturbative numerical analysis of the QMF and

PQMF theories for the SIS model on synthetic and real networks. For $\gamma < 5/2$, we show that while QMF theory provides an estimate of the epidemic threshold which is asymptotically exact, it fails to reproduce the prevalence asymptotically close to the transition. Moreover, we show that the PQMF theory considerably outperforms the standard QMF on synthetic networks of distinct levels of heterogeneity and degree correlations, providing extremely accurate predictions when the system is not too close to the epidemic threshold, while QMF theories deviates substantially from simulations for networks with a degree exponent $\gamma > 5/2$. The scenario for real networks is more complicated, still with PQMF significantly outperforming the QMF theory. However, despite a high accuracy for most investigated networks, in a few cases PQMF deviations from simulations are not negligible. We found correlations between accuracy and average shortest path while other basic networks metrics seem to be uncorrelated with the theory accuracy. Our results show the viability of the PQMF theory to investigate the high prevalence regimes of recurrent-state epidemic processes on networks, a regime of high applicability.

Localization phenomena permeates many branches of physics playing a fundamental role on dynamical processes evolving on heterogeneous networks. Localization is frequently grounded on eigenvectors of specific matrices, such as the adjacency [44] or non-backtracking matrices [49,50] defined in Sec. 2.1 and chapter 8 respectively, which emerge in theories of dynamic processes near to an active to inactive transition. In chapter 7, we advance in this problem gauging nodal activity to quantify the localization in dynamical processes whether they are near to a transition or not. This work was published in Ref. [103]. We investigated activity vectors of spreading processes on a wide spectrum of networks, from annealed to scale-free, both analytically and numerically showing that nodal activity patterns depend on the network structure. Using annealed networks, in which connections are rewired at an infinite rate, we show that a localized state at the transition and an endemic phase just above it are not incompatible features of an epidemic process. We also report that epidemic prevalence near to the transition is determined by the delocalized component of the network even when the inverse of the partition rate (IPR) analysis indicates a localized

activity. Also, dynamical processes with distinct critical exponents can be described by a same localization pattern. Turning to quenched networks, a more complex picture, depending on the type of activation and on the range of degree exponent are observed and discussed.

Another relevant aspect of real epidemic spreading is the mobility of an individual through the network [23]. The addition of diffusion in epidemic spreading models is usually performed considering a metapopulation model, in which a node is occupied by a population (bosonic model) and individuals migrate through the nodes [51]. We investigate a fermionic (each node corresponds to an individual) SIS model with mobility of infected individuals on uncorrelated networks with power-law degree distributions $P(k) \sim k^{-\gamma}$ of exponents $2 < \gamma < 3$ [52]. Considering the standard diffusion, in which an infected agent swaps its position with one of its nearest-neighbors with equal chance, a non-monotonic dependence of the epidemic threshold on the diffusion rate D is observed and an optimum diffusion rate D^*, for which the epidemic spreading is more efficient, is found. The effects of diffusion can be quite different when it is biased as we observed if the destiny of the moving infected is chosen with probability proportional to the neighbor's degree, resulting in a monotonic decay of the epidemic threshold as a function of the diffusion rate. Extending our analysis for $\gamma > 3$, stronger effects of localization were observed in biased diffusion, and we expect to apply the methodology based on nodal activity to characterize it. Even though a big part of this work was performed and the corresponding publication written and published [52] during the PhD, it was not included as result of this thesis to avoid overlap with my former master dissertation [53].

Finally, in chapter 8, the main results of this thesis are summarized and some perspectives are discussed.

Chapter 2

Complex Networks: An Introduction

In this chapter, a brief review of the basic concepts of complex networks is presented. The main statistical properties such as degree distribution, average shortest path length and average nearest neighbors degree of a node are defined following Refs. [1,11,54]. In the end of this chapter we present how to determine the relevance of a network node considering some types of centrality [55]. These concepts are explored to determine, for example, who are responsible to generate outbreaks in epidemic processes [17]. Latter in this thesis, we investigate the relation between topological properties and performance of theoretical description of epidemic spreading.

2.1 Basic concepts and statistical characterization

Complex network is a powerful tool to represent and study a diversity of biological, infra-structural, technological, and social systems. The protein-protein interaction in yeast [56], the western states' power grid of the United States [16] and Internet [12] are a few examples which can be cited. Basic concepts of networks can be presented considering the friendship between 34 members of a karate club in the 1970s [57]. In this social network, each individual corresponds to a node. A friendship between two members of this karate club is represented by an edge that can be

weighted if each edge has different levels of importance, directed such as in social media, in which a user i can follow another user j without the latter follows back the former, representing a directed edge pointing from i to j. In particular, the karate club network is unweighted and undirected, in which each edge has the same level of importance and the direction is not a relevant aspect of the connection linking two nodes. In Figure[1] 2.1, a representation of the karate club network is shown illustrating some interesting aspects of complex networks. At a first glance, the degree of a node, defined as the number of nearest neighbors of a node, can be different. For example, node 8 has degree 4 while node 21 has degree 3. Their neighborhoods are highlighted in Figure 2.1 in orange and yellow, respectively.

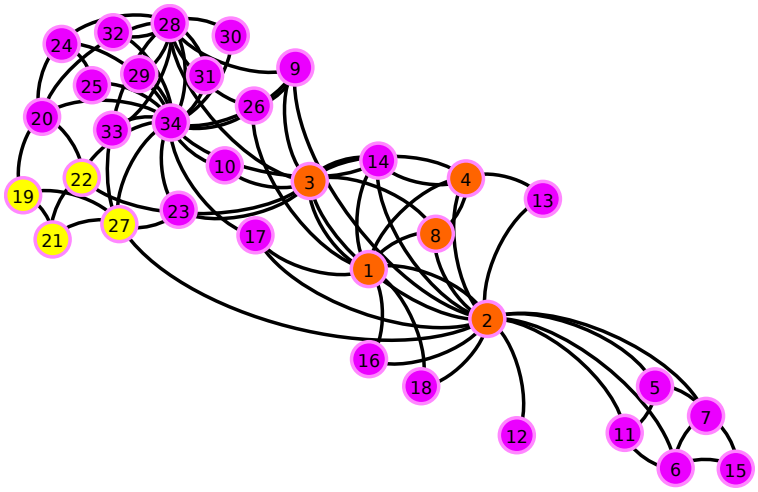


Figure 2.1: Network representing friendship between 34 members of a karate club at a US university in the 1970s [57]. Orange and yellow nodes represent the neighborhood of nodes 8 and 21, respectively, including themselves.

Matrices allow mapping connections among nodes into a useful mathematical structure. In this context, the adjacency matrix is defined as

$$A_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

In particular, for undirected unweighted networks the adjacency matrix is symmetric, $A_{ij} = A_{ji}$. The degree of a node i, k_i, can be obtained from this definition as

$$k_i = \sum_{j=1}^{N} A_{ij}. \tag{2.2}$$

[1] All figures in this thesis were originally made by us. The representation of the Karate club network was obtained using the software Gephi [58].

For large networks, we need to use their statistical properties. A very useful one is the degree distribution defined as the probability that a randomly chosen node has degree k and it is given by

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

in which N_k is the number of nodes of degree k and N is the network size. The degree distribution for the karate club network is illustrated in Figure 2.2. Other examples in which this quantity is computed are presented in Figure 2.3. In Figure 2.3(a), the pet social media dogster.com is considered. Each node is a pet profile and an edge linking them represents a friendship between two profiles [59,60]. In the Youtube network, Figure 2.3(b), each node is a user and the process of sharing a video establishes an edge [61,62]. In actor collaboration network, Figure 2.3(c), nodes are actors and they are connected if they have acted in a same movie [20]. In these examples, one can observe a power-law tail in the degree distribution. This occurs for many other networks representing real systems such as the Internet [12] and flights between international airports [1]. This feature is associated with the robustness against random failures observed in the Internet [11], in which a node and its connections are removed at random and a fragmentation in isolated subgraphs usually occurs when some finite fraction of nodes is removed [1,11]. However, when the network presents a power-law degree distribution, $P(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$, it happens only when almost all nodes are removed [11]. Thus, in general, the elimination of nodes at random does not break the Internet apart.

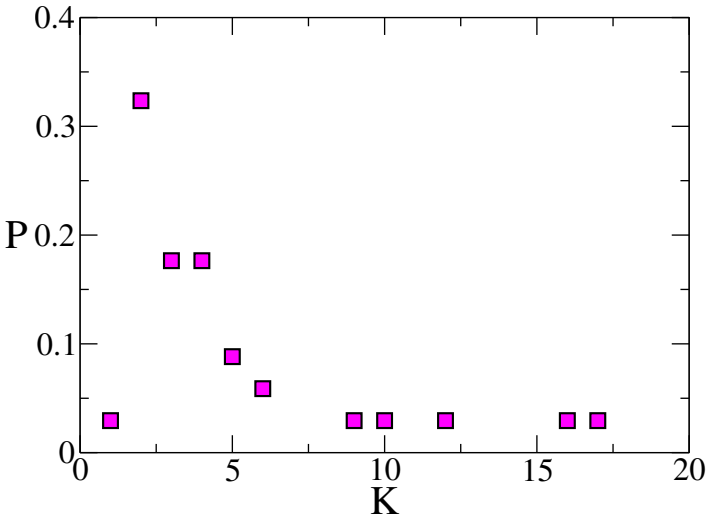


Figure 2.2: Degree distribution ot the karate club network [57].

Some features of the networks and quantities related to dynamical processes

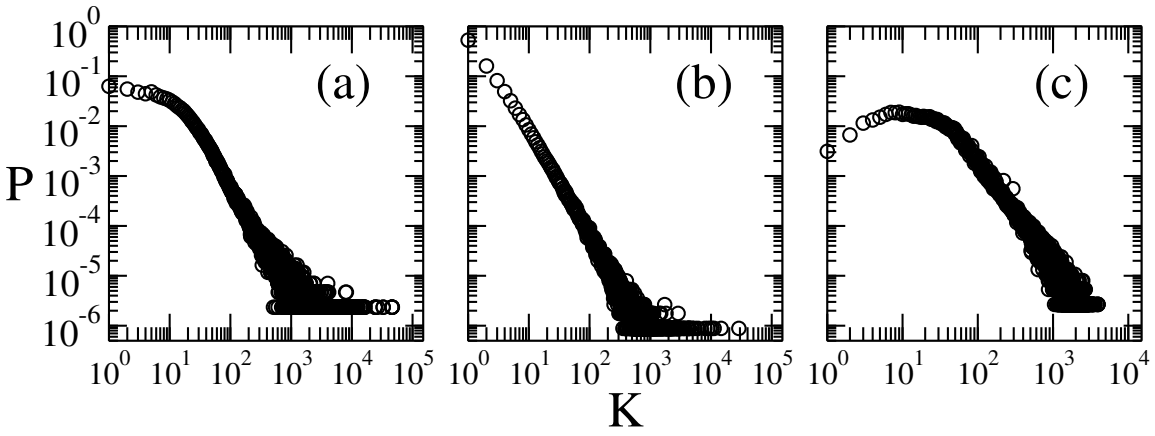


Figure 2.3: Degree distribution of networks representing the (a) social media dogster.com [59,60], (b) sharing of videos among Youtube users [61,62] and (c) actors which have acted in the same movie [20].

occurring on their top have a straightforward relation with the moments of the degree distribution. The nth moment is given by

$$\langle k^n \rangle = \sum_k k^n P(k). \tag{2.4}$$

The first and second moment will be especially very helpful hereafter. For example, they can be used to quantify how heterogeneous the degree distribution is, using the coefficient of heterogeneity defined as [1]

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{2.5}$$

When $\kappa \sim \langle k \rangle$, the network is considered homogeneous while a scale-free network presents $\kappa \to \infty$ in the limit of infinite network size (thermodynamic limit). In general, a finite scale-free network is characterized by $\kappa >> \langle k \rangle$ [1]. It happens, for example,

Network	$\langle k \rangle$	$\langle k^2 \rangle$	κ
karate club	4.6	35.6	7.7
dogster.com	40.1	82361.5	2053.9
Youtube users	5.3	2603.7	491.3
actors collaboration	80.2	33542.2	418.2

Table 2.1: First and second moment and the coefficient of heterogeneity associated with the degree distribution of real networks of Figures 2.1 and 2.3.

in networks presenting power-law degree distribution, $P(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$ [11].

In table 2.1, we show the first and second moments and the coefficient of heterogeneity associated with this degree distribution of real networks presented in Figures 2.2 and 2.3. In particular, the real networks in Figure 2.3, in which power-law tails in the degree distribution are observed, have a high heterogeneity coefficients.

Back to the karate club network, we will consider the following situation: The individual corresponding to node 15 has forgotten his black belt at the mat and the individual corresponding to node 20 has found out it. How could node 20 hand in the black belt to node 15 respecting the network structure? One solution would be the individual 20 to pass it to one of his colleagues, which would pass it to another and so on until the black belt arrives to individual 15. Thus, a path connecting nodes 20 and 15 is built following the connections of the network. This process can be performed for any pair of nodes i and j of the network, and the length of the path linking these nodes, denoted by l_{ij}, is equal to the number of edges crossed from the nodes i to j. Note that a path connecting the nodes i and j may not exist if the network is fragmented. If there is at least one path connecting any pair of nodes, the network is connected. Set of nodes presenting at least a path connecting them form a component. The giant component is a component whose size scales linearly with the network size. The existence of a giant component indicates that an extensive fraction of the network is connected. In particular, for connected networks the giant component is the network itself. Some pairs of nodes can present many paths connecting them and the shortest one is usually the most important. Figure 2.4 shows the shortest

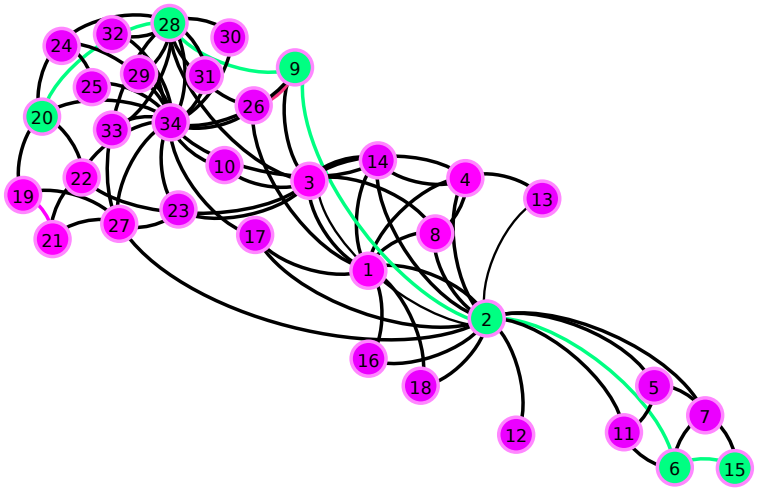


Figure 2.4: The cyan nodes and edges represent the shortest path linking nodes 15 and 20 in karate club network [57].

path from the individuals 20 and 15 of the karate club network of length $l_{15,20} = 5$. Computing the shortest path length for all pairs of nodes, its average value is given by

$$\langle l \rangle = \frac{1}{N(N-1)} \sum_{ij} l_{ij}. \tag{2.6}$$

For the karate club network the average shortest path length is $\langle l \rangle = 2.4$. As a function of the network size N, we have $\langle l \rangle \sim N^{1/D}$ in lattices, in which D is the Euclidian spatial dimension. This scaling usually assumes the form $\langle l \rangle \sim \ln N$ for random complex networks. Therefore, the increasing of distance with size in these networks is much slower than in lattices. This property is called "small-world" [16]. Small-world networks have infinity dimension due to the logarithmic dependence with size.

Networks can present loops in which a path starting from a node i passes through distinct edges and returns to i. In special, triangles are formed when three nodes i, j and l are connected to each other. The existence of triangles on a network is related to the clustering C_i of a node i defined as the ratio of the number of edges e_i connecting two neighbors of i and the maximum number of these edges that could occur [16]. Mathematically, it is given by

$$C_i = \frac{2e_i}{k_i(k_i - 1)}, \tag{2.7}$$

in which k_i is the degree of the node i. In addition, the average clustering coefficient is given by,

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i. \tag{2.8}$$

For the karate club network, we observe a high clustering coefficient given by $\langle C \rangle = 0.571$.

Correlations can appear in the process of connection between two nodes. Considering the node degree, there are three basic patterns of correlations. If nodes of similar degrees tend to be preferentially connected among them, the degree correlation is assortative. It means that large-degree nodes tend to connect with each other. Conversely, if the nodes are more likely to attach to nodes of very different degree, the

correlation is disassortative. In this pattern of degree correlation, large-degree nodes tend to connect to small-degree ones. Finally, when the nodes are connected at random, the network is uncorrelated or neutral. To quantify the degree of correlation, Newman [13] has proposed to use the Pearson coefficient,

$$P = \frac{\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{N\langle k\rangle}\right) k_i k_j}{\sum_{ij}\left(k_i \delta_{ij} - \frac{k_i k_j}{N\langle k\rangle}\right) k_i k_j}, \tag{2.9}$$

in which δ_{ij} is the Kronecker delta. It is constrained to the interval $-1 < P < 1$, being negative for disassortative, null for uncorrelated and positive for assortative degree correlations. For karate club network $P = -0.476$, which indicates a disassortative degree correlation. The Pearson coefficient can provide misleading conclusions when the correlation behavior is very complicated, calling for the introduction of another way of determining the degree correlation. A commonly used quantity is the average degree of the nearest-neighbors of the node i given by [12]

$$k_{\mathrm{nn,i}} = \frac{1}{k_i} \sum_{j \epsilon \nu(i)} k_j, \tag{2.10}$$

in which the sum runs over the nearest neighbors of i, $\nu(i)$. Considering all nodes of degree k, one obtains the average degree of the nearest neighbors for nodes of degree k which is given by

$$k_{\mathrm{nn}}(k) = \frac{1}{N_k} \sum_{i|k_i=k} k_{\mathrm{nn,i}}, \tag{2.11}$$

in which N_k is the number of nodes of degree k. It is associated with the conditional probability that a node of degree k is connected to a node of degree k', $P(k'|k)$, through

$$k_{\mathrm{nn}}(k) = \sum_{k'} k' P(k'|k). \tag{2.12}$$

If $k_{\mathrm{nn}}(k)$ is a decreasing function of k, the correlation pattern is dissortative as in the case of the karate club network shown in Figure 2.5(a). If $k_{\mathrm{nn}}(k)$ is an increasing function, the correlation pattern is assortative. The Cond-mat network, in which each

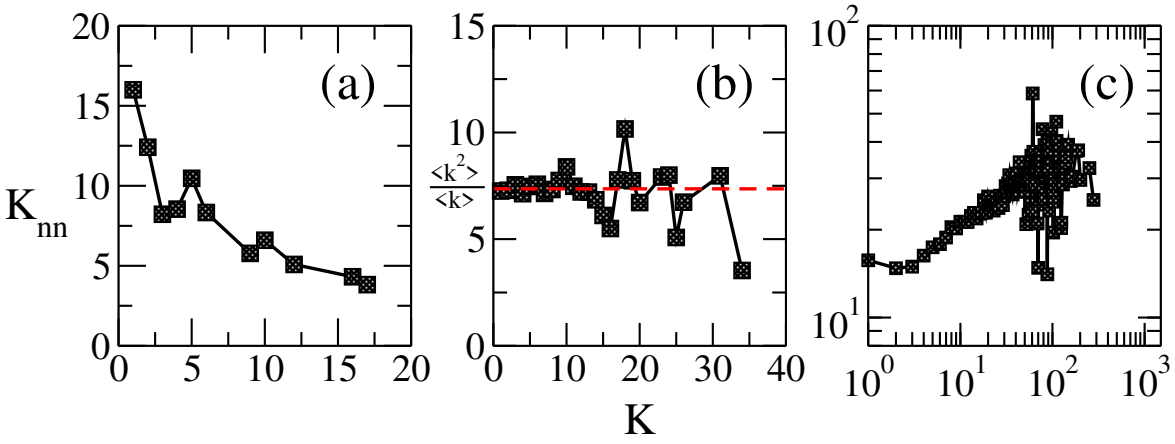


Figure 2.5: Average degree of the nearest neighbor as a function of degree for (a) karate club [63], (b) air Traffic [63] and (c) Cond-Mat 1993-2003 [63] networks. In (b), the dashed line corresponds to $k_{\rm nn}(k) = \frac{\langle k^2\rangle}{\langle k\rangle}$ for the air traffic network.

edge represents a connection between two co-authors in an arXiv paper submission in the condensed matter category, presents this pattern as shown in Figure 2.5(c). The absence of correlation is marked by a constant value of $k_{\rm nn}$. This constant can be determined taking into account that $P(k'|k)$ is only a function of k'. Thus,

$$P(k'|k) = \frac{k' N_{k'}}{\sum_k k N_k}, \tag{2.13}$$

in which the numerator is the number of edges emanating from nodes of degree k' and the denominator is the total number of these edges. Applying Eq. (2.4), we found

$$P(k'|k) = \frac{k' P(k')}{\langle k\rangle}. \tag{2.14}$$

Finally, replacing Eq. (2.14) in Eq. (2.12), the result is

$$k_{\rm nn}(k) = \frac{\langle k^2\rangle}{\langle k\rangle}. \tag{2.15}$$

In the air traffic network [64], where the nodes represent airports or service centers and links are created from routes recommended by the National Flight Data Center (NFDC), USA's Federal Aviation Administration (FAA), presents absence of correlations. Furthermore, the constant value is very close to the prediction of Eq. (2.15); see

Figure 2.5(b).

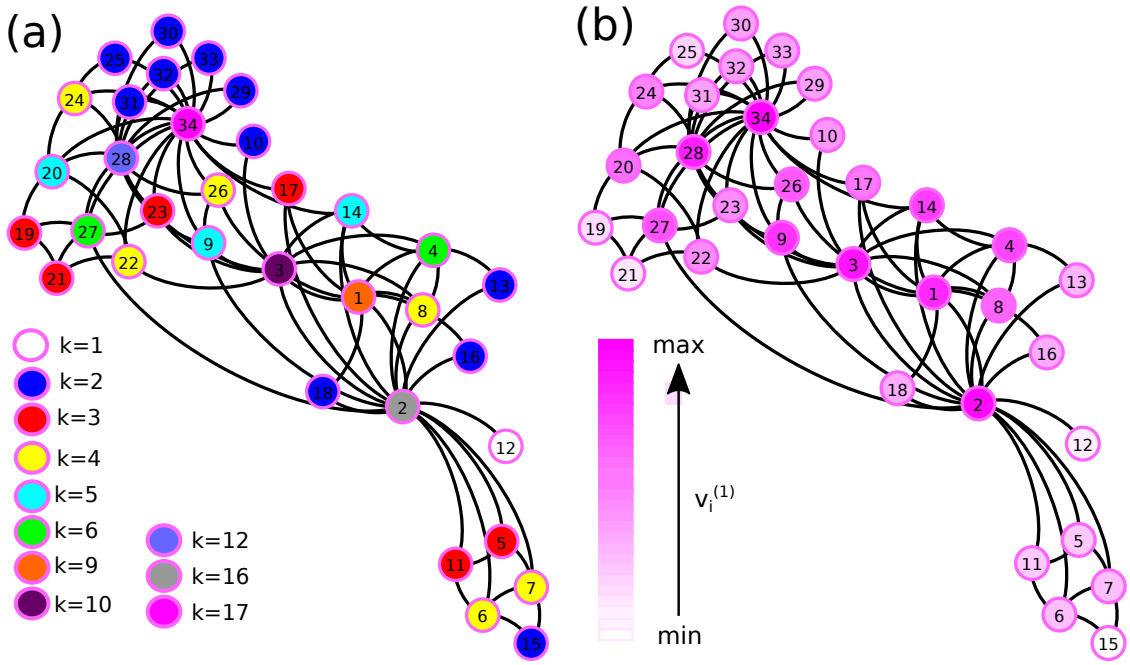


Figure 2.6: Representation of the (a) and degree centrality, (b) eigenvector centrality in the karate club network. In the former, each color represents a set of nodes of same degree centrality such as nodes 5, 11, 17, 19, 21 and 23. In the latter, a gradient of colors from a lighter (smallest $v_i^{(1)}$) to deeper magenta (highest $v_i^{(1)}$) indicates the eigenvector centrality of a node. The eigenvector centrality to the nodes 5, 11, 17, 19, 21 and 23 is distinct.

Each node can present a certain level of relevance in the network with respect to a specific property and this concept is called centrality. Here, we will present centralities which are important to the dynamical processes studied in this thesis. The degree centrality [54] is based on the node degree. According to this centrality, the highest ranked node in the karate club network is the 34, while the nodes 5, 11, 17, 19, 21 and 23 have the same relevance, as shown in Figure 2.6(a). In a social context, the degree centrality would correspond to who knows more people is more relevant. However, it may not be a good indicator of the relevance of an individual since, for example, some people are relevant by who they know. To incorporate this feature, we can defined a centrality based on the neighborhood of a node, mathematically given by [65]

$$v_i^{(1)} = \frac{1}{\Lambda^{(1)}} \sum_j A_{ij} v_j^{(1)}, \tag{2.16}$$

in which $\Lambda^{(1)}$ is the largest eigenvalue of the adjacency matrix and $v_i^{(1)}$ is the component i of its associated eigenvector. These components are called eigenvector centrality. The choice of the largest eigenvalue of the adjacency matrix and its associated eigenvector to define the eigenvector centrality explores the fact that both are unique and positive. It is ensured by the Perron-Frobenius theorem, which states that for non-negative and irreducible matrices, such as the adjacency matrix, the largest eigenvalue is unique and positive and there is only one eigenvector associated with this eigenvalue. In addition, all components of this eigenvector are non-negative and it is the only eigenvector with all non-negative elements [54].

To determine the eigenvector centrality from an arbitrary vector $\vec{x}^{(0)}$ with positive entries the power method can be applied [55]. In this method, we have

$$\vec{x}^{(k)} = \mathbf{A}^k \vec{x}^{(0)}. \tag{2.17}$$

The initial condition as function of the eigenvectors of the adjacency matrix, $v^{(j)}$, is given by

$$\vec{x}^{(0)} = \sum_j c_j \vec{v}^{(j)} \tag{2.18}$$

for an appropriated choice of the constants c_j. Replacing Eq. (2.18) in Eq. (2.17), we obtain

$$\vec{x}^{(k)} = \Lambda^{(1)} \sum_j \left(\frac{\Lambda^{(j)}}{\Lambda^{(1)}} \right)^k \mathbf{A}^k c_j \vec{v}^{(j)}, \tag{2.19}$$

in which $\Lambda^{(j)}$ is the eigenvalue associated with $v^{(j)}$. In particular, $\Lambda^{(1)}$ is the largest and non-degenerate eigenvalue (LEV) associated with $v^{(1)}$ which is the principal eigenvector (PEV). Thus, in the limit of $k \to \infty$, we have that $\vec{x}^{(k)} \sim \vec{v}^{(1)}$. In Figure 2.6(b), one can see that the eigenvector centrality of the karate club network differs from its degree centrality. In the latter, the nodes 5, 11, 17, 19, 21 and 23 have the same centrality while in the former, node 23 is more central than 19 and 21, for example.

The redistribution of the importance presents in Eq. (2.16) does not avoid concentration of this centrality in a few nodes as a instance, when the weight of the eigenvector components is almost concentrated in hubs [55]. Hubs are nodes present-

ing high degree with values much larger than the average degree [11]. In the case of scale-free networks, in which $\kappa >> \langle k \rangle$, hubs are always present in meaningful amounts. In especial, the degree of the largest hub on random scale-free networks is limited by a natural cutoff which presents a dependence with the network size, $k_c \sim N^{1/(\gamma-1)}$ [66]. A more detailed discussion about the natural cutoff will be done in the next chapter.

The concentration of the components of the eigenvector could occur in other structures as the maximum k-core; densely connected component identified by a k-core decomposition [67]. This process of decomposition starts from nodes with the minimum degree in the network, $k_{\min}$. These nodes and the edges connected to them are eliminated and their neighbors have their degree reduced by one. After this reduction, two cases are possible. In the first one, no node in this new network has degree equal to or smaller than the $k_{\min}$ and the set of removed nodes forms the $k_{\min}$-shell and the remaining nodes represent the $(k_{\min}+1)$-core. In the second case, some nodes still present $k \leq k_{\min}$ and the procedure above is repeated until that there is no node of degree $k \leq k_{\min}$ and $k_{\min}$-shell is identified. The identification of shells are performed until all nodes of the network have been included into one of them. This process is repeated and the maximum k-core (of index k_s) is composed by the nodes such that one more iteration of the procedure removes all of them. In Figure 2.7, the k-core decomposition is shown for the karate club network. The maximum k-core, with index $k_s = 4$, is composed by the blue nodes.

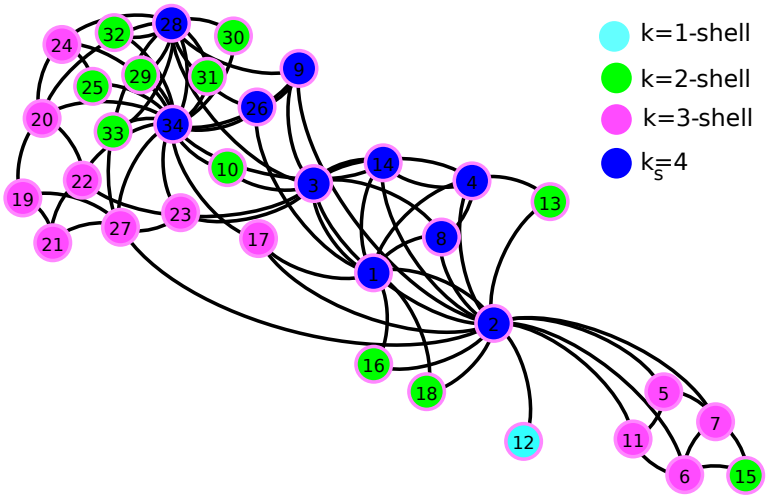


Figure 2.7: K-core decomposition in the karate club network. Cyan node belongs to the 1-shell, green nodes to the 2-shell and magenta nodes to the 3-shell. The maximum k-core, with $k_s = 4$, is composed by the blue nodes.

The eigenvector centrality highlights the leading role of the eigenvalues and their respective eigenvectors associated with the adjacency matrix in the network. During the development of theoretical descriptions to predict the behavior of dynamical processes on the top of networks based on mean-field approaches, more attention will be devoted to this issue, introducing tools to identify and compute localization and other properties such as the spectral gap, which is the difference between the largest and the second largest eigenvalue. A broad investigation of the accuracy of mean-field theories and spectral properties, localization and spectral gap, will be performed in this thesis.

Chapter 3

Modeling complex networks

In this chapter, a brief review of some models capable of generating networks with specified topological features observed on real networks is presented. These models allow to control properties such as the degree correlation and degree distribution, for example. A description of the main algorithms to implement them and some results obtained from their application are reproduced. A especial attention is devoted to the uncorrelated configuration model (UCM) [68] and the Weber-Porto model [69] (WPCM) to generate the synthetic networks used hereafter.

3.1 An overview

The characterization of real networks exploring the concepts presented in chapter 2 reveals that very complex patterns can arise such as power-law degree distribution, small-word, and degree correlations, for example. These complex patterns are consequence of the self-organized dynamical evolution of networks [1]. Synthetic networks constitute an important tool to understand the origin of these patterns and their roles on dynamics processes on the top of real networks. Models have been proposed [16,20,70,71] to generate these synthetic networks with specific topological properties.

The first model we consider to generate synthetic networks is the Erdős-Rényi model [70]. Erdős-Rényi-like networks are generated connecting each pair of nodes with

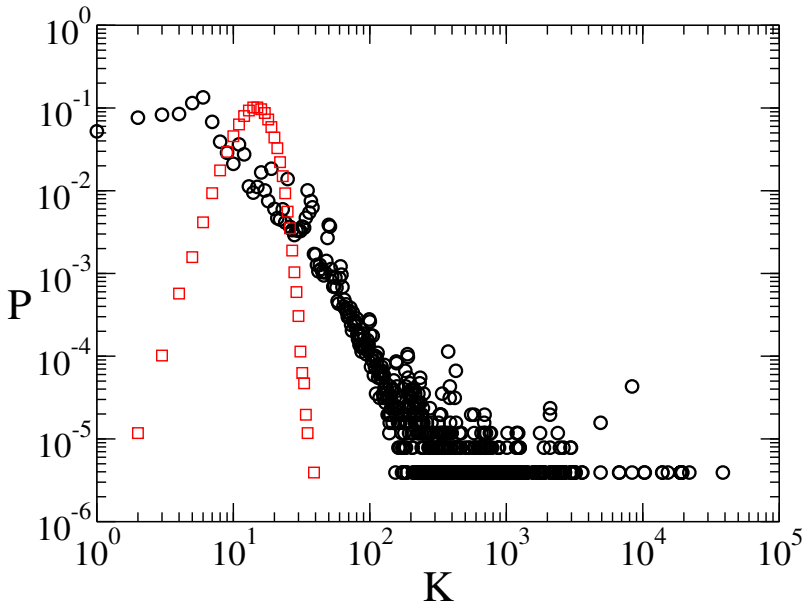


Figure 3.1: Comparison between the degree distributions of the Web Stanford network (black circles) and a Erdős-Rényi-like network (red squares) presenting the same average degree and network size.

a probability p [72]. These networks allow to compute the set of structural properties presented in chapter 2, making possible to identify more complex patterns caused by the presence of some bias in the edge connections. Following the above rules, an Erdős-Rényi-like network presents a binomial degree distribution which becomes a Poissonian when $\langle k \rangle << N$ [11], being, thereby completely described by $\langle k \rangle$ [11]. A comparison between a degree distribution of an Erdős-Rényi-like network with the same $\langle k \rangle$ and N as the Web Stanford network, where nodes are pages from Stanford University and edges represent connections among them, with its actual degree distribution is presented in Figure 3.1. It highlights the existence of nodes having a degree much larger than the average degree, indicating existence of hubs on this real network absent in its randomized version.

We introduce the next model starting with the lattice shown in Figure 3.2(a). The edges emanating from a node are rewired clockwisely with a probability p. In this rewiring, the new end-point is a node randomly chosen avoiding self- and multiple connections. With probability $1-p$, the present connection is preserved. When $p = 0$, the lattice is not modified exhibiting both large clustering coefficient and average shortest

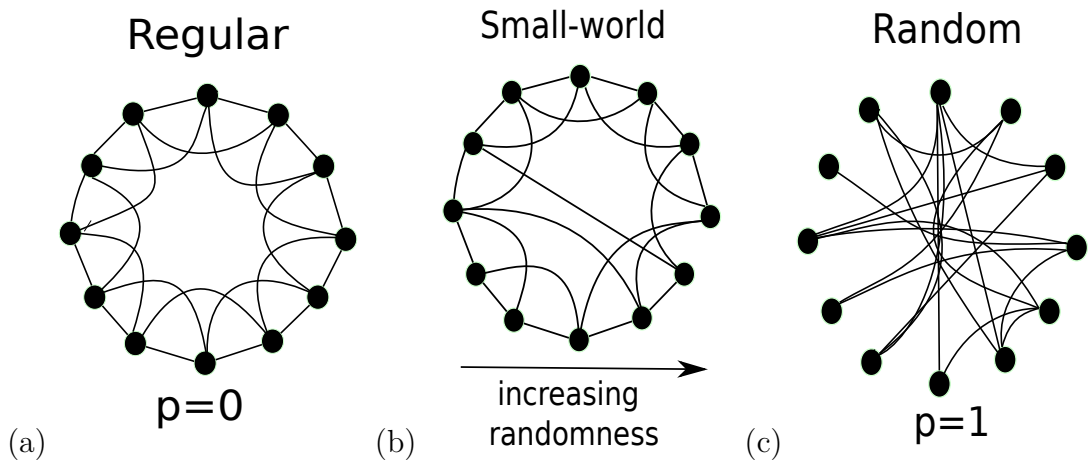


Figure 3.2: Evolution of the Watts-Strogatz model considering the values of the probability of rewiring. (a) Ring lattice in which each node shows the same degree for $p = 0$. (b) Increasing p the small-world is observed. (c) A random network occurs for $p = 1$.

path length. For $p = 1$, all edges are rewired generating an almost random network as shown in Figure 3.2(c), in which very small clustering coefficient and average shortest path length are observed. Thus, the parameter p allows to control the clustering coefficient and the shortest path length, interpolating between lattices and random graphs as in Figure 3.2(b). This model was proposed by Watts and Strogatz [16] inspired by social networks, in which the clustering coefficient is high and, concomitantly, the average shortest path length is small [1]. This behavior observed for intermediate values of p are detected, for example, in neural network of the worm *Caenorhabditis elegans* and for the western states power grid of United States network [16].

A class of relevant models assuming the evolutionary mechanism and the preferential attachment condition [20] is capable of mimicking the power-law degree distribution of several networks observed in real world, such as those in Figure 2.4. The model rules are as follows. A group of m_0 nodes connected to each other represents the initial condition. A new node attaches to $k_0 < m_0$ nodes previously added at random avoiding multiple connections. The attachment to a node occurs with a probability proportional to its degree k. As consequence of this choice we have $P(k) \sim k^{-3}$ [20]. Although a power-law degree distribution has been obtained, this choice of attachment probability does not allow to control the corresponding exponent. It can be done, for example, choosing the node to attach with probability proportional to $k + a$, in which

a is a constant [71]. This modification provides a scaling-law $P(k) \sim k^{-\gamma}$ in which $\gamma = 3 - a/k_0$ and the average degree of nearest neighbors takes the form [71]

$$k_{\mathrm{nn}}(k) \sim \begin{cases} k^{-3+\gamma} \text{ if } \gamma < 3 \\ \ln(k) \text{ if } \gamma > 3. \end{cases} \tag{3.1}$$

The disassortative pattern observed for $\gamma < 3$ is expected since the preferential attachment mechanism promotes links between nodes of minimum degree k_0 and nodes of largest degree of the network. The degree distribution and the average degree of the nearest neighbors as a function of degree, for a network generated applying this model adopting $a = 2.4$, $k_0 = 3$, are shown in Figure 3.3. As predicted by Barrat et al. [71], they scale as $P(k) \sim k^{-2.2}$ and $k_{\mathrm{nn}} \sim k^{-0.8}$.

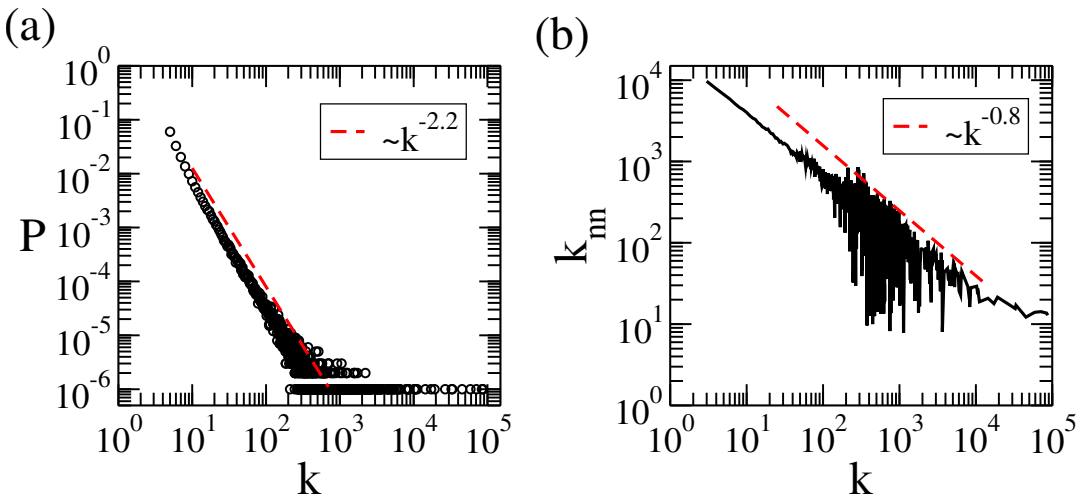


Figure 3.3: (a) Degree distribution and (b) average degree of the nearest neighbors as a function of degree for a network generated with the preferential attachment model. The model parameters were $a = 2.4$ and $k_0 = 3$. The network size is $N = 10^6$.

3.1.1 Uncorrelated configuration model (UCM)

In the preferential attachment model, the power-law degree distribution and the degree correlation pattern are not independent. A choice of the exponent of the degree distribution has a defined degree correlation pattern associated with it. In the rest of this chapter, our attention will be devoted to introducing a model capable of generating networks where both the exponent of the power-law degree distribution and

the correlation pattern can be turned independently.

The configuration model (CM) [73] is an alternative to build a network controlling the power-law degree distribution without presence of degree correlation. In this model, each node i receives a degree k_i according to the degree distribution. In an auxiliary list, each index i is repeated k_i times. Two elements of this list are selected at random and an edge between the corresponding nodes is established and the elements are removed from the list. Since, any restrictions are imposed, edges between the same pair of nodes (multiple edges) and connecting a node to itself (self-connections) are allowed in this model. The connections are performed until the list is empty. A variation of this model is the uncorrelated configuration model (UCM), in which the multiple and self-connections are forbidden. These constraints impose some restrictions to the size of the largest hub of the network. To understand this, we consider that the degree cutoff k_c [66] is the average value of the actual largest degree k_{max} of the network in a sample [74], it means $k_c = \langle k_{max} \rangle$. Therefore, fluctuations in the degree cutoff are expected. Even if no restriction is imposed in the largest degree or during the formation of connections allowing self- and multiple-connections, a named natural cutoff will appear according to the degree distribution in finite networks [68]. This natural cutoff is reckoned considering that above this degree k_c one expects observed at most one node [66], that can be put in mathematical terms as

$$N \int_{k_c}^{\infty} P(k) dk \sim 1. \tag{3.2}$$

For the case $P(k) \sim k^{-\gamma}$, the cutoff that emerges for a power-law distribution is

$$k_c \sim N^{1/(\gamma-1)}. \tag{3.3}$$

It is impossible to connect a large network using the natural cutoff for $\gamma < 3$ without intrinsic correlations when multiple and self-connections are not allowed. Thus, a cutoff [68]

$$k_c \simeq (\langle k \rangle N)^{1/2}, \tag{3.4}$$

named structural cutoff, has to be proposed. In the UCM, the largest degree of the network have an upper bound given by Eq.(3.4). When $\gamma > 3$, the natural cutoff is smaller than the structural one and no restriction is required to build a uncorrelated network with $P(k) \sim k^{-\gamma}$ without any self- and multiple-connections. This choice for k_c implies in $k_{max} \approx k_c$ for $\gamma < 3$ (called hard cutoff [74]), while for $\gamma > 3$, large fluctuations in the value of k_{max} are expected and $k_c = \langle k_{max} \rangle \sim N^{1/(\gamma-1)}$ [74].

The UCM allows to control the exponent of the power-law degree distribution, but there is no degree correlation. This model can be modified to control the degree correlations of the network by introducing a rule for the formation of an edge which depends on the degree. This is the basic mechanism behind the Weber and Porto model [69] which allows to generate networks presenting both a power-law degree distribution and degree correlation in a controlled way. In this model, the connection probability relies on the correlation function

$$f(k,k') = \frac{P(k,k)}{P(k,k')_{unc}} = \frac{P(k,k')}{P_e(k)P_e(k')}, \tag{3.5}$$

in which $P(k,k')$ is the joint probability that nodes with degrees k and k' are connected, $P(k,k')_{unc}$ corresponds to the case in which there is no correlation and $P_e(k)$ is the probability that an edge end in a node of degree k given by

$$P_e(k) = \frac{kP(k)}{\langle k \rangle}. \tag{3.6}$$

Let $\langle A(k) \rangle_e$ represents the average of $A(k)$ computed with respect to $P_e(k)$.

We have that

$$P(k,k') = \frac{E_{kk'}}{\langle k \rangle N}, \tag{3.7}$$

in which $E_{kk'}$ is the number of edges connecting nodes of degrees k and k' and $\langle k \rangle N$ is twice the total number of edges on a network of size N. Note that the probability $P_e(k)$ can be written as

$$P_e(k) = \sum_{k'} P(k,k'). \tag{3.8}$$

Combining Eqs. (3.5) and (3.8) we find

$$P_{\mathrm{e}}(k) = \sum_{k'} P_{\mathrm{e}}(k) P_{\mathrm{e}}(k') f(k,k') = P_{\mathrm{e}}(k) \langle f(k,k') \rangle_{\mathrm{e}}^{k'} \tag{3.9}$$

what implies in

$$\langle f(k,k') \rangle_{\mathrm{e}}^{k'} = 1. \tag{3.10}$$

Since we are considering undirected networks, we have $E_{kk'} = E_{k'k}$. Thus, from Eqs. (3.7) and (3.5), the function $f(k,k')$ is symmetric with respect to k and k'. A simple form of the function $f(k,k')$ which satisfies Eq. (3.10) and the symmetry is

$$f(k,k') = 1 + h(k)h(k'), \tag{3.11}$$

in which h is a function of k which fulfills

$$\langle h(k) \rangle_{\mathrm{e}} = \sum_{k} h(k) P_{\mathrm{e}}(k) = 0. \tag{3.12}$$

The conditional probability of a node of degree k be connected to another of degree k',

$$P(k'|k) = \frac{E_{kk'}}{kN_k} = \frac{P(k,k')}{P_{\mathrm{e}}(k)}, \tag{3.13}$$

is written taking into account the correlation function as

$$P(k'|k) = P_{\mathrm{e}}(k') f(k,k'), \tag{3.14}$$

which results in

$$k_{\mathrm{nn}}(k) = \sum_{k'} k' P(k'|k) = \langle k \rangle_{\mathrm{e}} + h(k) \langle k' h(k') \rangle_{\mathrm{e}} \tag{3.15}$$

and

$$h(k) = \frac{k_{\mathrm{nn}}(k) - \langle k \rangle_{\mathrm{e}}}{\langle k' h(k') \rangle_{\mathrm{e}}}, \tag{3.16}$$

in which the form of $f(k,k')$ in Eq. (3.11) was considered. The denominator can

be computed multiplying both sides of Eq. (3.15) by $kP_e(k)$ and summing over all k resulting in

$$\langle k'h(k')\rangle_e = \sqrt{\langle kk_{nn}(k)\rangle_e - \langle k\rangle_e^2}. \tag{3.17}$$

Combining Eqs. (3.17), (3.16) and (3.11), we obtain

$$f(k,k') = 1 + \frac{(k_{nn}(k) - \langle k\rangle_e)(k_{nn}(k') - \langle k\rangle_e)}{\langle kk_{nn}(k)\rangle_e - \langle k\rangle_e^2}. \tag{3.18}$$

Finally, the connection probability is defined as

$$P_{link} = \frac{f(k,k')}{f_{max}}, \tag{3.19}$$

in which f_{max} is the maximum value of the correlation function $f(k,k')$.

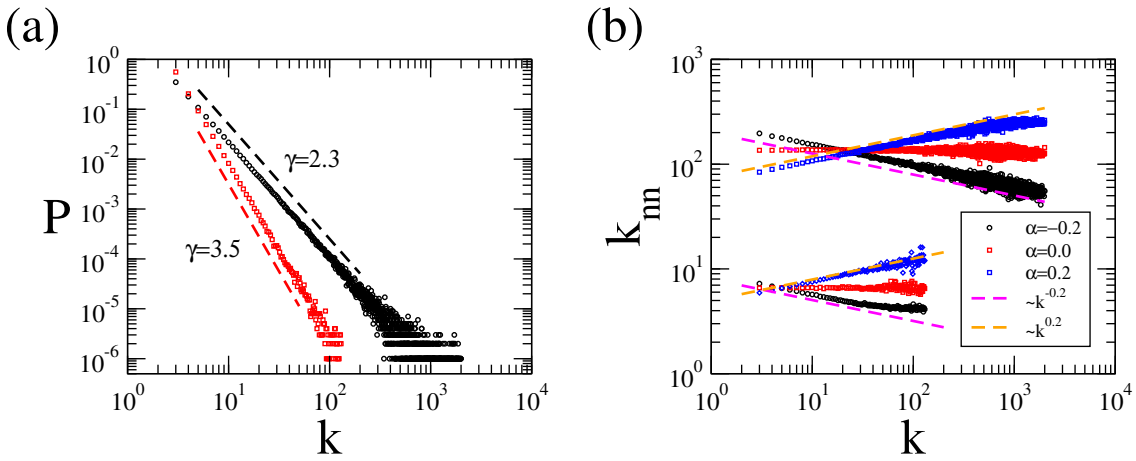


Figure 3.4: (a)Degree distribution in the form $P \sim k^{-\gamma}$ with $\gamma = 2.3$ (black circle) and $\gamma = 3.5$ (red square). (b) Average degree of the nearest neighbors as a function of degree presenting the form $k_{nn}(k) = Ak^{\alpha}$, with $\alpha = -0.2$ (black circles), $\alpha = 0.0$ (red squares) and $\alpha = 0.2$ (blue diamonds). They correspond to disassortative, uncorrelated and assortative patterns, respectively. The top (bottom) patterns correspond to $\gamma = 2.3$ ($\gamma = 3.5$). The network size is $N = 10^6$, the lower degree is $k_{min} = 3$ and the cutoffs adopted are $k_c = 2\sqrt{N}$ for $\gamma = 2.3$ and $NP(k_c) = 1$ for $\gamma = 3.50$.

In the rest of this thesis, the synthetic networks will be generated by the Weber and Porto model, hereafter called by Weber and Porto configuration model (WPCM), assuming

$$k_{nn}(k) = \frac{\langle k^2\rangle}{\langle k\rangle}k^{\alpha}, \tag{3.20}$$

where $\alpha = -0.2$ and $\alpha = 0.2$ were adopted to disassortative and assortative networks, respectively. The other parameters are a minimum degree $k_{\min} = 3$ and a structural cutoff $k_c = 2\sqrt{N}$ for $\gamma < 3$ and $NP(k_c) = 1$ when $\gamma > 3$, which accelerates the convergence to the thermodynamic limit and avoids large fluctuations on the size of the largest hub [74], respectively. In Figure 3.4, the degree distribution with $\gamma = 2.3$, 3.5 and $k_{nn}(k)$ obtained from networks with size $N = 10^6$ generated by the WPCM model considering this set of parameters are shown. The small deviations observed in k_{nn} for large values of k in the presence of correlation are due to the network finite size that prevents k_{nn} from decaying or increasing indefinitely with k.

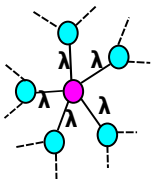
Chapter 4

Modeling epidemic spreading

This chapter is devoted to studying epidemic spreading and, in particular, the SIS model. The main features of this model are described with focus on the transition from an inactive phase, in which the epidemics is eradicated, to an active one, with an endemic state in which a finite fraction of the population is infected. The theoretical frameworks based on mean-field approaches applied to determine the epidemic threshold and the behavior of the density of infected individuals are reviewed. The dynamics of hubs and the interaction among them are related to the physical mechanism responsible to generate outbreaks in this model. Finally, a description of computational methods to simulate the SIS dynamics are described, presenting tools to circumvent the difficulties to deal with the absorbing state in finite-size networks and optimized algorithms to implement the stochastic simulations.

4.1 SIS dynamics

An individual can be infected multiple times in the SIS dynamics. This feature occurs in diseases as the seasonal influenza [24] and gonorrhea [25]. In the SIS model, an infected individual transmits the disease for each of his nearest neighbors with a rate λ. This infected individual becomes susceptible spontaneously with a rate μ. We assume $\mu = 1$ without loss of generality since this choice only fixes the time unit. Representations of both processes are shown in Figure 4.1(a). Increasing the infection


λ
λ
λ
λ
λ

takes the form

$$\frac{d\rho}{dt} = -\rho + \lambda\langle k\rangle(1-\rho)\rho. \tag{4.1}$$

The first term on the right-hand side is related to the spontaneous healing. The second one corresponds to the infection events, in which the term, $(1-\rho)$ is the fraction of susceptible individuals, ρ is the fraction of infected individuals, λ is the infection rate, $\langle k\rangle$ is the average number of contacts of an individual in the population. In the steady-state, $\frac{d\rho}{dt} = 0$, Eq.(4.1) has two solutions: The trivial one $\rho = 0$ and

$$\rho = \frac{\lambda\langle k\rangle - 1}{\lambda\langle k\rangle}. \tag{4.2}$$

Using Eq.(4.1), a linear stability analysis around the trivial solution provides

$$\frac{d\rho}{dt} = -\rho + \lambda\langle k\rangle\rho \tag{4.3}$$

implying in

$$\rho(t) \sim e^{(-1+\langle k\rangle\lambda)t}. \tag{4.4}$$

For $-1+\lambda\langle k\rangle < 0$, it converges to the trivial solution, $\rho = 0$. For $-1+\lambda\langle k\rangle > 0$, it moves away from the trivial solution and converges to the infected fraction of individuals given by Eq.(4.2), implying in an epidemic threshold $\lambda_c = 1/\langle k\rangle$.

Near to the epidemic threshold, Eq. (4.2) can be written as

$$\rho = \frac{\lambda - \lambda_c}{\lambda}, \tag{4.5}$$

which provides a $\rho \sim (\lambda - \lambda_c)^\beta$ with an exponent $\beta = 1$.

4.2.1 Heterogeneous mean-field theory (HMF)

It is intuitive to guess that nodes having larger degrees would be better spreaders. To include this feature, the heterogeneous mean-field (HMF) assumes that all nodes presenting the same degree k are statistically equivalent [33,34]. In this approach, the dynamical correlations are still absent. The dynamical equation to the

density of infected nodes of degree k, ρ_k, is

$$\frac{d\rho_k}{dt} = -\rho_k + \lambda k(1-\rho_k)\sum_{k'} P(k'|k)\rho_{k'}. \tag{4.6}$$

The first term in the right-hand side is the spontaneous healing of nodes in the compartment k. The second one corresponds to the infection term, which is proportional to the infection rate λ, the density of healthy sites of degree k, $(1-\rho_k)$, the degree k and the probability that a link emanating from a node of degree k points to an infected node $\theta = \sum_{k'} P(k'|k)\rho_{k'}$. Remember that $P(k'|k)$ is the conditional probability of a node of degree k to be connected to one of degree k' (see Sec. 2.1).

A trivial solution of Eq.(4.6) is $\rho_k = 0$. Considering a linear stability analysis around this solution we obtain

$$\frac{d\rho_k}{dt} = \sum_{k'} L_{kk'}\rho_{k'}, \tag{4.7}$$

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

$$L_{kk'} = -\delta_{kk'} + \lambda k \sum_{k'} P(k'|k) \tag{4.8}$$

and $\delta_{kk'}$ is the Kronecker delta symbol. The solution $\rho_k = 0$ will be unstable for $-1 + \lambda\Lambda_m > 0$, where Λ_m is the largest eigenvalue of the connectivity matrix $C_{kk'} = kP(k'|k)$. As consequence, the epidemic threshold is [38]

$$\lambda_c = \frac{1}{\Lambda_m}. \tag{4.9}$$

Equation (2.14) holds for uncorrelated networks leading to the connectivity matrix

$$C_{kk'} = \frac{kk'P(k')}{\langle k\rangle}. \tag{4.10}$$

Note that $v_k = k$ is an eigenvector of the matrix $C_{kk'}$ in Eq.(4.10) and its associated eigenvalue is $\Lambda_m = \langle k^2\rangle/\langle k\rangle$. Since this matrix is positive and irreducible, the Perron-

Frobenius theorem ensures its unicity. Therefore, the threshold is given by

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{4.11}$$

For networks exhibiting a power-law degree distribution $P(k) \sim k^{-\gamma}$, the threshold goes to zero in the limit $N \to \infty$ for $\gamma < 3$, since $\langle k^2 \rangle \to \infty$ within this range of γ. One can show that the vanishing threshold as the network size goes to infinity occurs regardless of the degree correlation pattern of the network for $\gamma < 3$ [38]. However, the threshold converges to a finite value for $\gamma > 3$ as the network size goes to infinity. For annealed networks [74] defined as networks in which $\tau_{\text{Net}} << \tau_{\text{dyn}}$, where τ_{Net} and τ_{dyn} are the characteristic evolution time of the network and the dynamics running on its top respectively, the HMF predictions for the threshold are exact in the thermodynamic limit [74].

An analysis of the prevalence,

$$\rho = \sum_{k'} P(k')\rho_{k'}, \tag{4.12}$$

on uncorrelated network close to the critical point can be performed considering

$$\theta^{\text{nc}} = \sum_{k'} k' \frac{P(k')}{\langle k \rangle} \rho_{k'}. \tag{4.13}$$

In this limit, $\rho \to 0$ and $\theta^{\text{nc}} \to 0$. Considering the steady-state and Eq.(4.13), Eq.(4.6) is reduced to

$$\rho_k = \frac{\lambda k \theta^{\text{nc}}}{1 + \lambda k \theta^{\text{nc}}}, \tag{4.14}$$

and

$$\rho \approx \lambda \langle k \rangle \theta^{\text{nc}}. \tag{4.15}$$

Replacing Eq.(4.14) in Eq.(4.13), we found

$$\theta^{\text{nc}} = \sum_{k'} \frac{k' P(k')}{\langle k \rangle} \left[\frac{\lambda k' \theta^{\text{nc}}}{1 + \lambda k' \theta^{\text{nc}}} \right]. \tag{4.16}$$

Considering a power-law degree distribution and a continuous approximation for k, in which a sum over k is replaced by an integral, we have

$$\theta^{\mathrm{nc}} = \frac{(\gamma-2)}{k_0^{-\gamma+2}}\lambda\theta^{\mathrm{nc}}\int_{k_0}^{\infty}\frac{k^{2-\gamma}}{1+\lambda k\theta^{\mathrm{nc}}}dk, \tag{4.17}$$

resulting in,

$$\theta^{\mathrm{nc}} = F\left(1,\gamma-2,\gamma-1;\frac{-1}{(3-\gamma)\lambda\theta^{\mathrm{nc}}}\right), \tag{4.18}$$

in which $F(a,b,c;x)$ is the Gauss hypergeometric function [75]. Taking the asymptotic limit of $F(a,b,c;x)$, different solutions are observed depending on the value of γ.

For $2<\gamma<3$, we have

$$\theta^{\mathrm{nc}} \approx \Gamma(3-\gamma)\Gamma(\gamma-1)\left[\frac{1}{(3-\gamma)\lambda\theta^{\mathrm{nc}}}\right]^{\gamma-2}, \tag{4.19}$$

in which $\Gamma(x)$ is the gamma function. Plugging Eq.(4.19) into Eq.(4.15) results in

$$\rho \approx \lambda\langle k\rangle\theta^{\mathrm{nc}} \sim \lambda^{\frac{1}{3-\gamma}}. \tag{4.20}$$

Since λ_{c} goes to zero in this range of γ, it provides an exponent $\beta = 1/(3-\gamma)$ [34].

For $3<\gamma<4$, the result considering the lower order term becomes

$$\theta^{\mathrm{nc}} = \frac{\Gamma(\gamma-3)\Gamma(\gamma-1)}{\Gamma(\gamma-2)^2}(3-\gamma)\lambda\theta^{\mathrm{nc}} + \Gamma(\gamma-3)\Gamma(\gamma-1)[(3-\gamma)\lambda\theta^{nc}]^{\gamma-2}. \tag{4.21}$$

After some math and considering Eq.(4.11), we obtain

$$\rho \sim (\lambda-\lambda_{\mathrm{c}})^{\frac{1}{\gamma-3}}, \tag{4.22}$$

and $\beta = 1/(\gamma-3)$ [34].

Finally, for $\gamma>4$, the result is

$$\theta^{\mathrm{nc}} = \frac{\Gamma(\gamma-3)\Gamma(\gamma-1)}{\Gamma(\gamma-2)^2}(\gamma-3)\lambda\theta^{\mathrm{nc}}\left[\frac{(\gamma-3)^2}{\gamma-4}\lambda\theta^{\mathrm{nc}}-1\right], \tag{4.23}$$

implying

$$\rho \sim (\lambda - \lambda_c). \tag{4.24}$$

Therefore, $\beta = 1$ as in the homogeneous mean-field theory [34].

4.2.2 Quenched mean-field theory (QMF)

The full structure of the network is enclosed in the adjacency matrix. The introduction of this quantity in the theoretical framework is expected to produce a more accurate approach. So, the probability that the node i is infected, ρ_i, evolves as

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda(1 - \rho_i) \sum_j A_{ij}\rho_j. \tag{4.25}$$

Here, the first term in the right-hand side corresponds to the spontaneous recovering of an infected node, i. The second term computes the infection role of a susceptible node i by each infected neighbor j.

In an analogous analysis performed on HMF theory, a linear stability analysis around the trivial solution $\rho_i = 0$ leads to a Jacobian matrix,

$$L_{ij} = -\delta_{ij} + \lambda A_{ij}, \tag{4.26}$$

implying in a threshold,

$$\lambda_c = \frac{1}{\Lambda^{(1)}}, \tag{4.27}$$

in which $\Lambda^{(1)}$ is the LEV of the adjacency matrix. In especial, for random uncorrelated networks presenting power-law degree distribution, $P(k) \sim k^{-\gamma}$, the LEV is given by [76]

$$\Lambda^{(1)} \approx \begin{cases} \frac{\langle k^2 \rangle}{\langle k \rangle}, & \text{if } \gamma < 5/2 \\ \sqrt{k_{\max}}, & \text{if } \gamma > 5/2 \end{cases}, \tag{4.28}$$

in which $k_{\max}$ is the largest degree of the network. Combining Eqs.(4.28) and (4.27), the QMF theory predicts a vanishing epidemic threshold in the limit of network size $N \to \infty$ regardless of the value of γ, since $\langle k^2 \rangle$ and $k_{\max}$ diverge for $\gamma < 5/2$ and

$\gamma > 5/2$, respectively. Both HMF and QMF theories predict a vanishing threshold for $\gamma < 3$. However, if $\gamma > 3$, the HMF theory predicts a finite threshold in contrast with the QMF prediction. An attempt to generalize the form of the LEV for networks presenting power-law degree distribution suggests [77]

$$\Lambda^{(1)} \approx \max(\sqrt{k_{\max}}, \langle k \rangle_{\mathrm{KM}}), \tag{4.29}$$

in which $\langle k \rangle_{\mathrm{KM}}$ is the average degree of the nodes which belong to the maximum k-core (see Sec.2.1). In special, in the absence of degree correlation, we have that $\langle k \rangle_{\mathrm{KM}} = \langle k^2 \rangle / \langle k \rangle$ [67], recovering Eq.(4.28) for $\gamma < 5/2$. The form of the LEV in Eq.(4.29) suggests that the LEV is determined by a competition between two subgraphs: the star subgraph centered on the largest hub and the maximum k-core [78]. However, the application of Eq.(4.29) is not possible when the principal eigenvector (PEV) associated to the LEV is localized on any of these subgraphs and when the PEV is localized in a set of nodes without overlap with neither the star centered in the largest hubs or the maximum k-core, which happens for some real networks [78].

An important relation between the PEV of A_{ij} and the prevalence defined as

$$\rho = \frac{1}{N} \sum_{i=1}^{N} \rho_i, \tag{4.30}$$

can be established. The eigenvectors $\vec{v}^{(j)}$ of A_{ij}, associated to the eigenvalues $\Lambda^{(j)}$, generate an orthonormal basis for this matrix since A_{ij} is symmetric and real and, thus, Hermitian. Then, we can express ρ_i in this basis as

$$\rho_i = \sum_{j=1}^{N} c_j v_i^{(j)}, \tag{4.31}$$

in which the coefficient c_j is the projection of the vector $\vec{\rho} = (\rho_1, \ldots, \rho_N)$ on $\vec{v}^{(j)}$ and $v_i^{(j)}$ is the ith component of the eigenvector associated to the eigenvalue $\Lambda^{(j)}$.

From Eq.(4.25), the steady state provides

$$\rho_i = \frac{\lambda \sum_j A_{ij}\rho_j}{1+\lambda \sum_j A_{ij}\rho_j}. \tag{4.32}$$

Replacing Eq.(4.31) in Eq.(4.32), we find

$$\sum_k c_k v_i^{(k)} = \frac{\lambda \sum_j A_{ij} \sum_k c_k v_j^{(k)}}{1+\lambda \sum_j A_{ij} \sum_k c_k v_j^{(k)}}. \tag{4.33}$$

The sums over j and k are independent and can be permuted. Consequently, terms of the form $\sum_j A_{ij} v_j^{(k)}$ appear, which can be replaced by $\sum_j A_{ij} v_j^{(k)} = \Lambda^{(k)} v_i^{(k)}$. Multiplying both sides by $v_i^{(l)}$ and using the orthonormality, we obtain

$$c_k = \lambda \sum_k \Lambda^{(k)} c_k \sum_i \frac{v_i^{(k)} v_i^{(l)}}{1+\sum_m \Lambda^{(m)} c_m v_i^{(m)}}. \tag{4.34}$$

Near to the epidemic threshold, it is expected that the PEV domains the expansion in Eq.(4.31). Consequently, Eq.(4.34) becomes

$$c_1 = \lambda \Lambda^{(1)} c_1 \sum_i \frac{(v_i^{(1)})^2}{1+\lambda \Lambda^{(1)} c_1 v_i^{(1)}}. \tag{4.35}$$

Since the analysis is around the critical point, $\lambda\Lambda^{(1)} \approx 1$ and $\rho_i \approx 0$, the denominator can be expanded leading to

$$c_1 = \frac{1}{\sum_i (v_i^{(1)})^3} (\lambda \Lambda^{(1)} - 1). \tag{4.36}$$

Finally,

$$\rho \approx \frac{1}{N} \sum_i c_1 v_i^{(1)} = a \left(\lambda \Lambda^{(1)} - 1\right) \tag{4.37}$$

with

$$a = \frac{\sum_i v_i^{(1)}}{N \sum_i (v_i^{(1)})^3}. \tag{4.38}$$

This result was obtained considering that the prevalence is proportional to the PEV. This hypothesis is better for larger spectral gap $(\Lambda_1 - \Lambda_2)$, in which Λ_2 is the second

LEV [44,45]. In addition, one can see from Eq.(4.37) that $\beta = 1$ regardless of the degree distribution, degree correlation or any other topological feature of the network.

As previously mentioned, the PEV can be localized on a set of nodes. To determine this localization, the inverse participation ratio (IPR), defined as [44]

$$Y_4 = \sum_{i=1}^{N} (v_i^{(1)})^4, \tag{4.39}$$

can be studied as a function of the system size. Assuming a scaling law relation of the form $Y_4 \sim N^{-\nu}$, two situations are possible: an exponent $\nu = 1$ and $\nu < 1$ [79]. In the former situation, the PEV is delocalized and its components are given by $v_i^{(1)} \sim 1/\sqrt{N}$. In the latter, some localization is present. In particular, when the PEV components are finite in a finite set of nodes, the PEV is extremely localized and $\nu = 0$ in the limit of $N \to \infty$ [79]. A special case is the uncorrelated annealed network presenting power-law degree distribution. In these networks the PEV components are proportional to the degree of the nodes, $v_i^{(1)} \sim k_i$, as consequence $Y_4 \sim N^{(3-\gamma)/2}$ for $2 < \gamma < 3$ and a structural cutoff [79]. According to Ref. [79], the PEV components for quenched networks presenting uncorrelated power-law degree distribution with $\gamma < 5/2$ are similar to the annealed one and there is a localization in a set of nodes corresponding to the maximum k-core. For $\gamma > 5/2$, the PEV localization is around the largest hub. In particular for $5/2 < \gamma < 3$, the PEV will be ruled by the largest hub for N large enough [79]. A relation between the spectral gap and the PEV localization suggests that there is no spectral gap when the PEV presents an extreme localization [80]. The previous discussion establishes a relation between the prevalence and the spectral properties (PEV localization and spectral gap) near to the epidemic threshold of the QMF theory [44]. In chapter 5, these results will be studied in the presence of degree correlation and an attempt to establish a relation between the accuracy of the QMF theory in predicting the epidemic threshold and the spectral properties of the Jacobian matrices will be performed. In chapter 7 the discussion about localization will be analyzed considering a new methodology ruled by the node activity associated with a dynamical process and some puzzles will be elucidated.

4.2.3 Pair-quenched mean-field theory (PQMF)

The introduction of dynamical correlations in the previously presented mean-field theories can be done in a pairwise approximation [81–83]. It can be performed in different ways. For example, it can take into account the cluster structure of the network [83]. Here, it is performed following Ref. [81]. Firstly, we introduce the general notation in which, $[A_i]$ is the probability that node i is in the state A; $[A_i, B_j]$ is the probability that nodes i and j are in the states A and B, respectively. $[A_i, B_j, C_k]$ is the extension to three nodes. In particular for the SIS dynamics, in which the states susceptible and infected are represented by 0 and 1, respectively, the notation is alternatively written as $[1_i] = \rho_i$, $[0_i] = 1 - \rho_i$, $[0_i, 1_j] = \varphi_{ij}$, $[0_i, 0_j] = \omega_{ij}$, $[1_i, 1_j] = \psi_{ij}$ and $[1_i, 0_j] = \bar{\varphi}_{ij}$. The pairs ω_{ij} and ψ_{ij} are symmetric with respect to i and j. In addition, the relations

$$\begin{aligned} \psi_{ij} + \varphi_{ij} = \rho_j, \quad \psi_{ij} + \bar{\varphi}_{ij} &= \rho_i, \\ \omega_{ij} + \varphi_{ij} = 1 - \rho_i, \quad \omega_{ij} + \bar{\varphi}_{ij} &= 1 - \rho_j. \end{aligned} \tag{4.40}$$

hold for any pair of nodes.

In this approach, the SIS dynamics is mathematically described by the set of equations [81]

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda \sum_j \varphi_{ij} A_{ij}. \tag{4.41}$$

and

$$\frac{d\varphi_{ij}}{dt} = -\varphi_{ij} - \lambda\varphi_{ij} + \psi_{ij} + \lambda \sum_{\substack{l \epsilon \nu(j) \\ l \neq i}} [0_i, 0_j, 1_l] - \lambda \sum_{\substack{l \epsilon \nu(i) \\ l \neq j}} [1_l, 0_i, 1_j]. \tag{4.42}$$

The first equation describes the temporal evolution of the probability that a node i is infected, ρ_i. Considering one-node approximation, the pair φ_{ij} is factorized as $\varphi_{ij} \approx \rho_i(1 - \rho_j)$ and Eq.(4.41) assumes the form of Eq.(4.25) corresponding to the QMF theory. In Eq.(4.42), the first three terms represent infection and healing of φ_{ij} taking into account nodes i and j. The remaining terms represent changes caused

by the interaction with neighbors of i and j. The sums in these terms run over the neighborhood of j and i, $\nu(j)$ and $\nu(i)$ respectively, excluding the connection between nodes i and j.

To solve this system of equations, the triplets are approximated by [84,85]

$$[A_i, B_j, C_k] = \frac{[A_i, B_j][B_j, C_k]}{[B_j]}. \tag{4.43}$$

In the triplet, the nodes i and k are neighbors of j but i and k are not neighbors. It means, the cluster structure of the network is neglected in this approach. Terms of higher orders such as quadruplets and so forth are not considered. Using the approximation of Eq.(4.43) and the relations of Eq.(4.40), Eq.(4.42) leads to

$$\begin{aligned} \frac{d\varphi_{ij}}{dt} &= -(2+\lambda)\varphi_{ij} + \rho_j + \lambda \sum_l \frac{(1-\varphi_{ij}-\rho_i)\varphi_{jl}}{1-\rho_j}(A_{jl}-\delta_{il}) \\ &- \lambda \sum_l \frac{\varphi_{ij}\varphi_{il}}{1-\rho_i}(A_{il}-\delta_{lj}). \end{aligned} \tag{4.44}$$

A trivial solution for the system of equations formed by Eqs.(4.41) and (4.44) is given by $\rho_i = \varphi_{ij} = 0$. A linear stability analysis around the trivial solution simplifies Eq.(4.44) to

$$\frac{d\varphi_{ij}}{dt} = -(2+\lambda)\varphi_{ij} + \rho_j + \lambda \sum_l \varphi_{jl}(A_{jl}-\delta_{il}). \tag{4.45}$$

Using a quasi-static approximation in Eq.(4.41), in which $d\rho_i/dt \approx 0$, we obtain

$$\rho_i \approx \lambda \sum_j \varphi_{ij} A_{ij}. \tag{4.46}$$

Plugging Eqs.(4.40) and (4.46) in Eq.(4.45) and considering $d\varphi_{ij}/dt \approx 0$, we find

$$\varphi_{ij} \approx \frac{(2+\lambda)\rho_j - \lambda\rho_i}{2+2\lambda}. \tag{4.47}$$

Replacing Eq.(4.47) in Eq.(4.41) and performing some algebraic manipulation, the

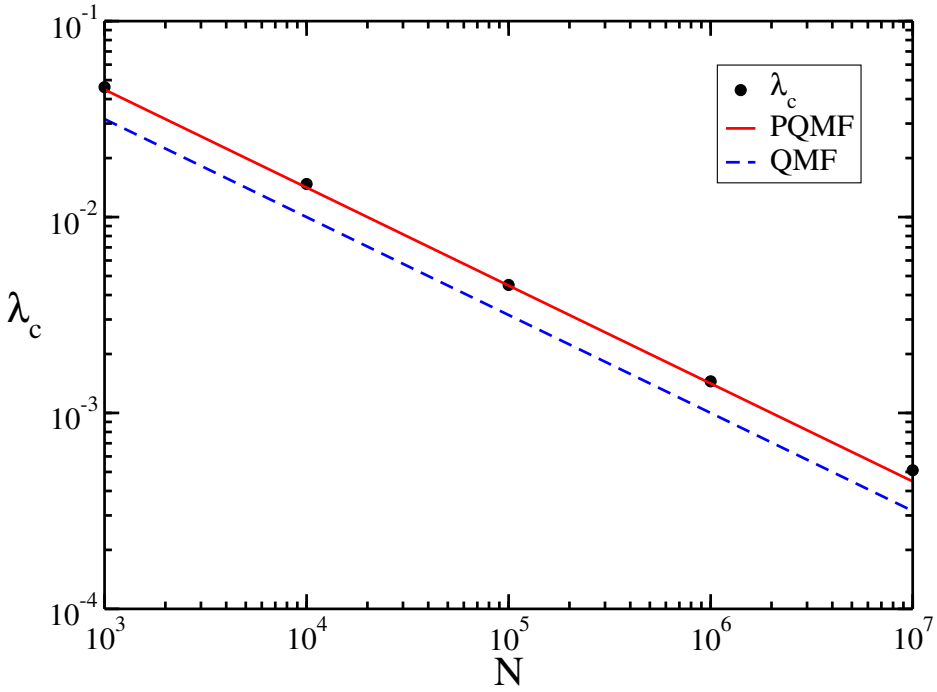


Figure 4.2: SIS epidemic threshold as a function of the star graph size. Here, black dots represent simulations, red line PQMF theory and dashed blue line QMF theory.

following Jacobian matrix is derived

$$L_{ij} = -\left(1 + \frac{\lambda k_i^2}{(2+2\lambda)}\right)\delta_{ij} + \frac{\lambda(2+\lambda)}{(2+2\lambda)}A_{ij}. \tag{4.48}$$

The threshold is given by the value of λ in which the largest eigenvalue of this Jacobian matrix is null. A special case is the star graph in which k nodes (leaves) are connected to a central node (center). The threshold as a function of the star graph size becomes

$$\lambda_c^{\mathrm{PQMF}} \simeq \sqrt{\frac{2}{N}}, \tag{4.49}$$

for $N \gg 1$. It differs from the QMF theory prediction, $\lambda_c^{\mathrm{QMF}} = 1/\sqrt{N}$ [26], by a multiplicative factor. In Figure 4.2, a comparison among results obtained by simulations (See Sec. 4.4), QMF, and PQMF theories for the threshold shows that this factor represents a meaningful improvement in predicting the epidemic threshold. It calls attention for the relevant role of the dynamical correlation in the SIS dynamics. In the next chapter, the PQMF spectral properties will be investigated, drawing a parallel

with the QMF spectral properties presented in Sec. 4.2.2.

4.3 The activation mechanisms of the SIS dynamics

The physics of SIS dynamics is based on the communication among hubs of a network. The following discussion is based on Ref. [29] which puts some mathematical results [31,32] into physical grounds. In this development, two stages are considered. The first one estimates the lifespan of the SIS dynamics in a hub, using a discrete-time approximation for SIS on a star graph as follows. The time unit corresponds to the inverse of the cure rate μ. A cycle is considered as shown in Figure 4.3. At $t = 0$, only the center is infected. After one time step, some leaves are infected and the center is healed and finally, in the last step, only the center can be again reinfected and all are healed.

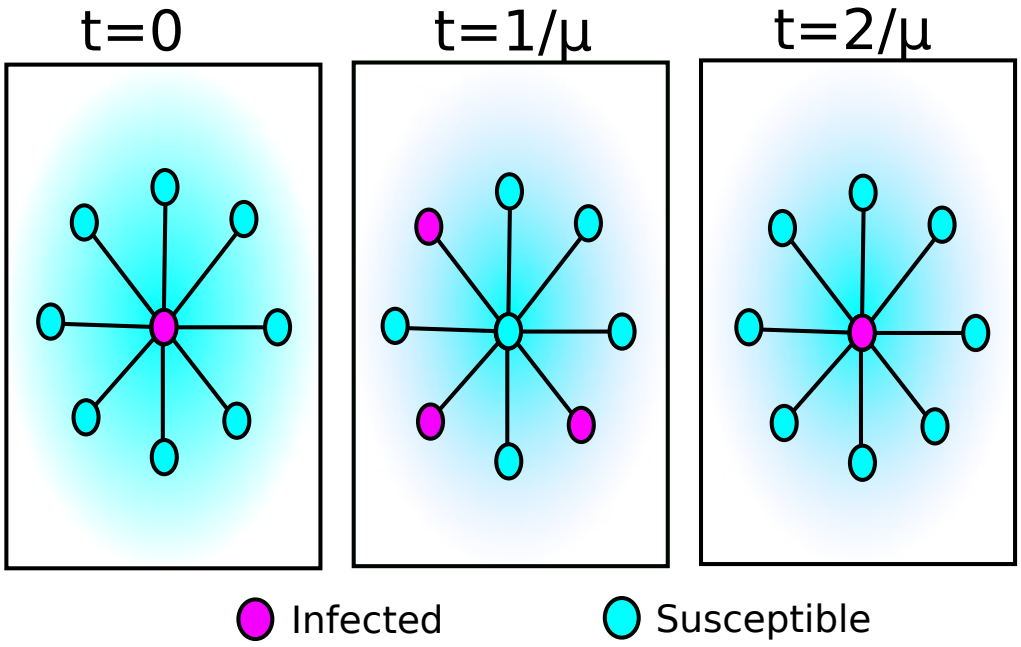


Figure 4.3: Cycle of infection in a hub considering an approximated discrete-time SIS dynamics.

The dynamics above reveals an exponential increase of the epidemic lifespan with the size k of the star given by [30]

$$\tau_k = \frac{2}{\mu} \exp\left(\frac{\lambda^2}{\mu^2} k\right). \tag{4.50}$$

The time required by a hub of degree k to infect a hub of degree k' can also be estimated. We consider that the hub of degree k is always infected (source) what is justified by Eq.(4.50), where $\tau_k \gg 1$ for a hub large enough. A shortest path of size d links the source to another hub of degree k' (target). Considering only this path leads to an upper bound for the time of propagation of the infection between these hubs, since other paths are neglected. A representation of the system is shown in Figure 4.4.

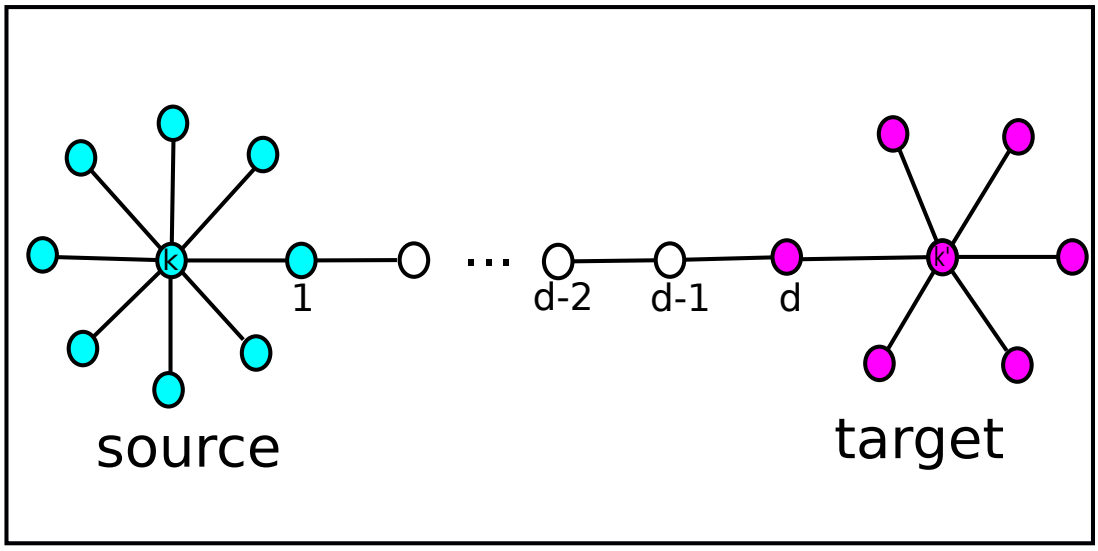


Figure 4.4: Communication between hubs of degree k and k' separate by a shortest path of size d.

For random networks presenting power-law degree distribution, the average distance separating two nodes of degree k and k' is given by [86]

$$d_{kk'} = 1 + \frac{\ln(N\langle k\rangle/kk')}{\ln(\kappa - 1)}, \tag{4.51}$$

in which κ is the coefficient of heterogeneity in Eq. (2.5). Thus, the time of infection between nodes having degree k and k' can be determined as [30]

$$\tau_{kk'} = \frac{1}{\lambda}\left(\frac{N\langle k\rangle}{kk'}\right)^{b(\lambda)} \tag{4.52}$$

in which $b(\lambda) = \ln(1 + \mu/\lambda)/\ln(\kappa - 1)$. Comparing Eqs.(4.50) and (4.52), one can conclude that hubs are capable of infecting each other, sustaining an endemic phase, since the epidemic lifespan increases exponentially while the mutual infection time does

algebraically implying that $\tau_k >> \tau_{kk'}$.

Rigorous mathematical results show that the epidemic threshold is null for random networks regardless the value of γ [31,32]. Also, it was shown that the prevalence near the critical point takes the form [32]

$$\rho = \begin{cases} \lambda^{\frac{1}{3-\gamma}}, \text{ if } 2 < \gamma < 2.5 \\ \frac{\lambda^{2\gamma-3}}{\log^{\gamma-2}\left(\frac{1}{\lambda}\right)}, \text{ if } 2.5 < \gamma < 3 \\ \frac{\lambda^{2\gamma-3}}{\log^{2\gamma-4}\left(\frac{1}{\lambda}\right)}, \text{ if } \gamma > 3 \end{cases} . \tag{4.53}$$

This activation mechanism has influenced other theoretical approaches. Equation (4.50) suggests an effective healing rate depending on the degree of a node, which could be incorporated in a HMF theory of the SIS model [87]. Moreover, for models such as the susceptible-infected (SI), susceptible-infected-recovered (SIR), susceptible-exposed-infected-recovered (SEIR), in which an individual cannot be reinfected, the message-passing approach [88] and its variations [49,89] have been successfully used to describe the epidemic threshold and other properties of this class of models. This theory relies on a hypothesis that a node i infected by a node j cannot reinfect j, in odds with the essential mechanism of the SIS dynamics. Therefore, the message-passing approach and its variations are not suitable for describing the SIS dynamics accurately [90].

Another accurate mathematical framework for SIS describes the communication between hubs by the cumulative massive percolation (CMP) [91]. The presence of a giant component obtained by this percolation corresponds to the existence of an endemic stationary state allowing to determine the epidemic threshold [92]. Computing the scaling of the cumulative giant component size, one is capable of establishing regimes of applicability of different theories [92]. For example, when only nearest neighbors are capable of forming clusters, nodes are added to the cumulative percolation giant component in an order established by the degree of the nodes. This degree ordered percolation giant component (DOPGC) is associated with independent activation of network hubs, providing a finite epidemic threshold for SIS when $\gamma > 3$ [93].

In the other hand, nodes forming clusters by a long-range interaction are added to the cumulative percolation giant component reproducing the null epidemic threshold and the scaling predicted to the prevalence in Ref. [32].

The next section will be devoted to simulation for the SIS and other epidemic dynamics in the forthcoming chapters.

4.4 How to simulate epidemic spreading

4.4.1 The quasistationary method (QS)

As previously discussed, the absence of infected individuals would be associated with an inactive phase, in which the epidemic does not propagate. It corresponds to an absorbing state. For finite sizes and long enough times, the system always reaches the absorbing state, which is the only real stationary-state. Since simulations are performed in finite systems, we need tools able to overcome this drawback. Here, we present the standard quasistationary method following the discussion of Ref. [28].

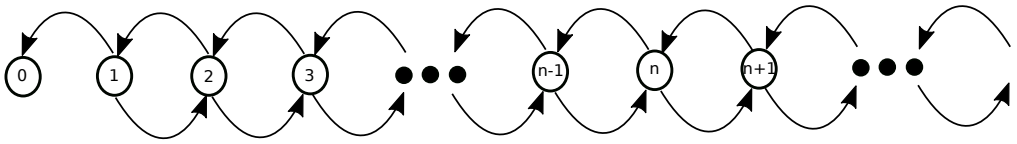


Figure 4.5: Representation of a random walk having an absorbing state. The walker is trapped when it visits the position 0.

Taking into account a continuous-time random walk in one-dimension, in which the walker performs discrete steps and is permanently trapped when it visits the position 0. It means, a walker can hop from 1 to 0, but the converse is forbidden, as illustrated in Figure 4.5. Assuming the walker is not at position 0 at the time $t = 0$, the probability that the walker is found at position n at time t, $P_n(t)$, is described by the master equation

$$\frac{dP_n(t)}{dt} = \sum_m w_{n,m} P_m(t) - \sum_m w_{m,n} P_n(t), \tag{4.54}$$

in which $w_{m,n}$ represents the transition rate from n to m.

The survival probability that the random walks are not trapped at $n = 0$ is given by

$$P_{\rm s}(t) = \sum_{n\geq 1} P_n(t) = 1 - P_0(t), \tag{4.55}$$

which represents the probability that the dynamics does not fall into an absorbing state. One can define

$$\bar{P}_n = \lim_{t\to\infty} \frac{P_n(t)}{P_{\rm s}(t)} \quad (n \geq 1), \tag{4.56}$$

being $\bar{P}_n$ the quasistationary distribution. Furthermore, $\bar{P}_0 = 0$ and $\sum_{n=1}^{N} \bar{P}_n = 1$ by construction.

To develop a simulation scheme to obtain the quasistationary distribution, we consider a modified dynamics where every time the walker visits the position $n = 0$, it is redirected to some place that it has already visited in the past, chosen with a weight proportional to the probability that the walker has visited that place. The probability of finding the walker in a position $n > 0$ is described by

$$\frac{dQ_n(t)}{dt} = \sum_{m\neq n} w_{n,m} Q_m(t) - \sum_{m\neq n} w_{m,n} Q_n(t) + q_0 Q_n(t), \tag{4.57}$$

in which $q_0 = \sum_m w_{0,m} Q_m$ is the flux of probability into the state $n = 0$. Equation (4.57) differs from Eq.(4.54) by the third term on the right-hand side. It reflects the redirection process from $n = 0$ to any other $n = m$. Furthermore, it is a term associated with memory. Thus, Eq.(4.57) is not a genuine master equation.

To implement the simulation of the quasistationary state, a list with the past configurations of the dynamical process is saved. Because of the computational limitations, a finite number of configurations, $N_{\rm conf}$, is stored. In each time step, one of them is replaced by the current state with a probability $P_{\rm rep}$ per unit of time. In the simulations presented in this thesis were adopted $N_{\rm conf} = 50$ and $P_{\rm rep} = 0.01$. If the system is in the absorbing state, one configuration of the list is selected at random to restart the simulation and the dynamics follows the usual rules of the system out of the absorbing state.

The time interval between two successive visits to the absorbing configuration

in a quasistationary regime is named lifespan. It can be determined considering that the probability that the system falls into the absorbing state is given by

$$\frac{dP_0}{dt} = \sum_m w_{0,m} P_m(t), \tag{4.58}$$

and the differentiation in relation to time of Eq.(4.55) results in

$$\frac{dP_s}{dt} = -\frac{dP_0}{dt}. \tag{4.59}$$

Combining Eqs.(4.58) and (4.59), we have

$$\frac{dP_s(t)}{dt} = -P_s(t) \sum_m w_{0,m} \bar{P}_m, \tag{4.60}$$

which has an exponential solution with a characteristic time which corresponds the lifespan. It is given by

$$\tau = \frac{1}{\sum_m w_{0,m} \bar{P}_m}. \tag{4.61}$$

For the random walk described above, $\tau = 1/\bar{P}_1$.

There are other ways to perform the quasistationary simulations [94] as for example the weak external field method. A new node is infected (or activated in a more general context) in any position of the system at random. The list of visited configurations is not necessary in this case. However, the choice of the creation rate has to be small enough to produce a negligible activity when compared with the quasistationary prevalence. The creation rate will be specified when the method be applied.

The SIS dynamics can be implemented following the optimized Gillespie algorithm [27]. The dynamics starts from a fully infected network. Other configurations could be adopted providing the same results [39]. At each time step, the number of infected nodes, N_{inf}, and the total number of edges emanating from them, N_{SI}, are computed. With probability,

$$Q = \frac{\mu N_{\text{inf}}}{\mu N_{\text{inf}} + \lambda N_{\text{SI}}}, \tag{4.62}$$

an infected node is chosen and healed. With complementary probability

$$1 - Q = \frac{\lambda N_{\mathrm{SI}}}{\mu N_{\mathrm{inf}} + \lambda N_{\mathrm{SI}}}, \tag{4.63}$$

an infected node i is chosen with probability proportional to its degree. Among the nearest-neighbors of i, a node j is chosen with equal probability. If j is susceptible, it becomes infected. However, nothing happens if j is infected. Finally, the time is incremented by

$$\delta t = \frac{-\ln u}{\mu N_{\mathrm{inf}} + \lambda N_{\mathrm{SI}}}, \tag{4.64}$$

in which u is a pseudo random number uniformly distributed in the interval $(0, 1)$.

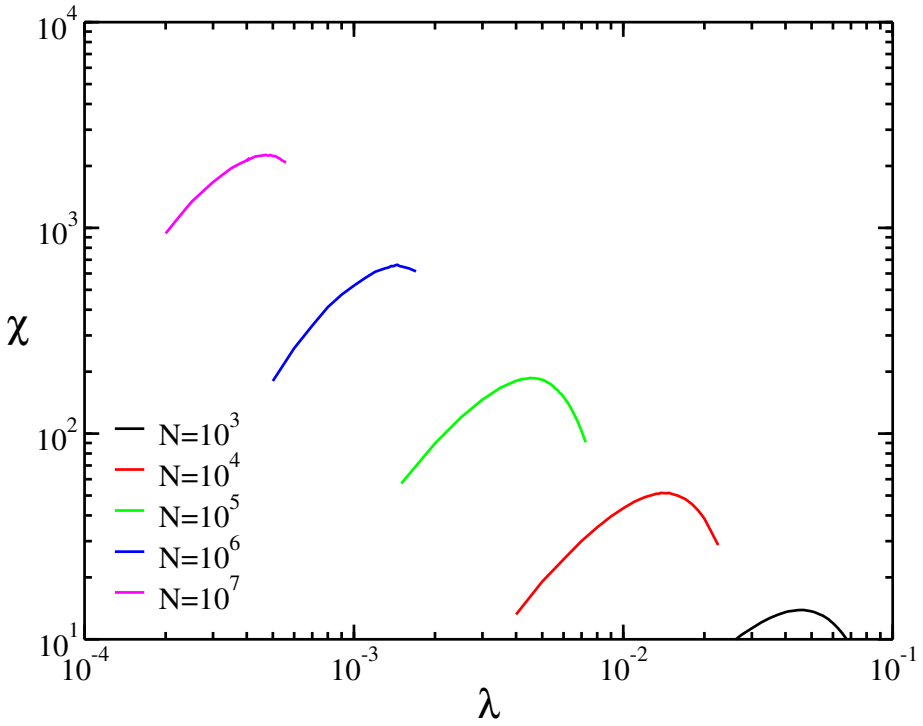


Figure 4.6: Susceptibility as a function of the infection rate for SIS dynamics on star graphs of different sizes.

The threshold can be determined using the dynamical susceptibility [26]

$$\chi = \frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle} \tag{4.65}$$

as a function of the infection rate. From the QS method, $\langle \rho^2 \rangle$ and $\langle \rho \rangle$ are the second and

the first moment of the quasistationary distribution computed applying the definition

$$\langle \rho^{\alpha} \rangle = \sum_{n \geq 1} n^{\alpha} \bar{P}_n. \tag{4.66}$$

Finally, the value of the infection rate λ at which χ is maximum is an estimate of the epidemic threshold λ_c [26]. In particular, for power-law degree distribution with $\gamma > 3$ multiple transitions can appear if no restriction to $k_{\max}$ is applied [39]. To avoid multiple transitions when $\gamma > 3$, we can adopt an upper cutoff $k_c \sim N^{-1/\gamma}$ [39,90]. In Figure 4.6, the susceptibility as a function of the infection rate for star graphs of different sizes is shown. The epidemic threshold estimated by the maximum of the dynamical susceptibility as a function of the network size is presented in Figure 4.2.

Chapter 5

Relating mean-field approaches accuracy and their spectral properties

In this chapter, a detailed study of the SIS threshold on complex networks is performed. Synthetic networks presenting power-law degree distributions, $P(k) \sim k^{-\gamma}$, with distinct values of γ and a variety of degree correlation patterns are considered. It will allow to investigate if some results reported for the uncorrelated case holds or does not in the presence of degree correlations. Moreover, a general picture can be established determining the role of degree correlation on the spreading process. The further development of the PQMF theory will allow to explore the spectral properties of the corresponding weighted adjacency matrix. A comparison between stochastic simulations and mean-field approaches (HMF, QMF and PQMF) is done to evaluate their performance on predicting the epidemic threshold. In particular, this performance will be associated with the spectral properties of standard and weighted adjacency matrices related to QMF and PQMF theories, respectively. Finally, these analyses are performed in real networks. The results in this chapter were published in Ref. [47].

5.1 Effects of degree correlation on the SIS threshold

The degree correlated power-law networks were generated applying the Weber-Porto configuration model (WPCM) described in Sec. 3.1.1. In this model, the average degree of the nearest neighbors as a function of degree assumes a form, $k_{\rm nn}(k) \sim k^{\alpha}$ and the degree distribution is given by $P(k) \sim k^{-\gamma}$. Here, we investigate $\alpha = -0.20$ and $\alpha = 0.20$ which correspond to disassortative and assortative patterns, respectively. When $\alpha = 0.00$, this function assumes the constant value $k_{\rm nn}(k) = \langle k^2 \rangle / \langle k \rangle$, and the network is uncorrelated. The SIS dynamics was simulated following the procedures described in Sec. 4.4. The epidemic threshold as a function of the network size for different values of γ and α are shown in Figure 5.1. At a first glance, the attention will be devoted to the cases $\gamma < 5/2$ and $\gamma > 3$, for which the physical mechanisms underlying the epidemic transition are clear [18]. Later we will discuss the case $5/2 < \gamma < 3$, whose interpretation is hampered by extremely long crossover phenomena in the spectral properties [77,79].

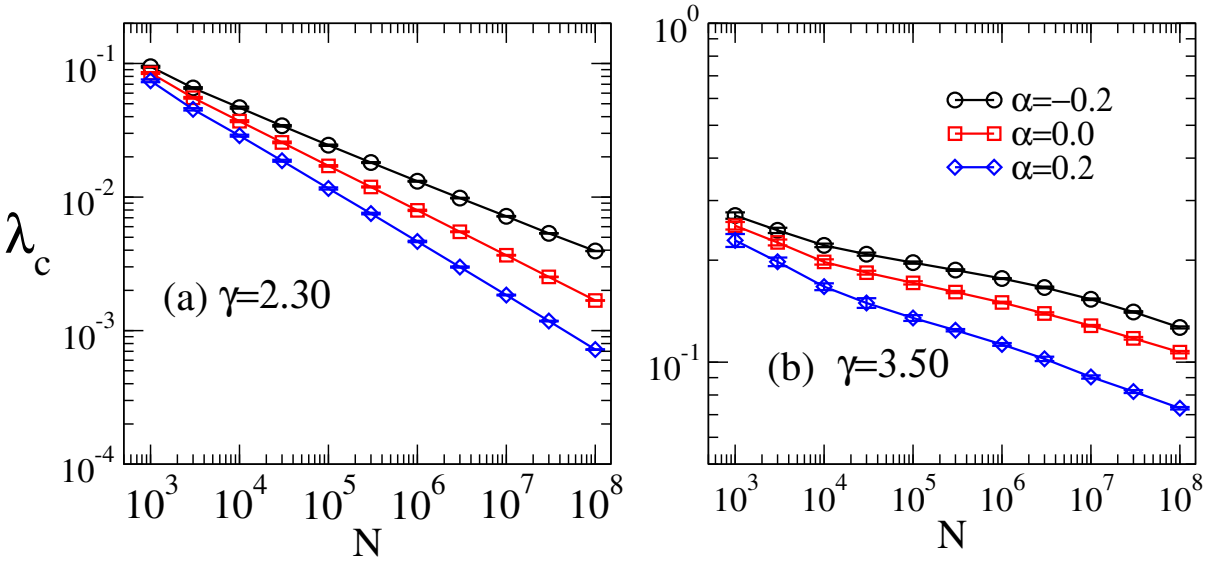


Figure 5.1: Epidemic threshold as a function of the network size for different values of α and (a) $\gamma = 2.3$ and (b) $\gamma = 3.5$. The lower cutoff is $k_{\rm min} = 3$ for all curves while the upper cutoff is $k_{\rm c} = 2\sqrt{N}$ ($\sim N^{1/\gamma}$) for $\gamma < 3$ ($\gamma > 3$). Curves are averages over 10 networks; error bars are smaller than symbols.

As we can see in the Figure 5.1, all thresholds vanish as N diverges, regardless of the correlation level α and heterogeneity γ. Compared to the uncorrelated case, assortative networks ($\alpha > 0$) have a smaller threshold, while the threshold is larger for $\alpha < 0$, i.e., disassortative mixing, in agreement with the behavior of the LEV of the adjacency matrix [44,95]. In the case $\gamma > 3$, this phenomenology can be qualitatively explained by considering the mechanism of long-range mutual reinfection of hubs (see Sec. 4.3), which triggers the epidemic transition. According to this mechanism, the subgraph consisting of the hub plus its nearest neighbors can sustain in isolation an active state for times long enough to permit the activation of other hubs, even if they are not directly connected. This mechanism is at work independently of degree correlations, as long as distances among hubs increase slowly enough with network size. In assortative networks, communication among hubs is enhanced since they have higher probability to be closer; for disassortative topology the converse is true and larger values of λ are needed to trigger the transition.

5.2 Revisiting the PQMF theory

In the QMF theory, a direct relation between spectral properties of the adjacency matrix (A_{ij}) and the prevalence, is established (see Sec. 4.2.2). An analogous development can be done for the PQMF theory. As previously seem in Sec. 4.2.3,

$$\frac{d\rho_i}{dt} = -\rho_i + \lambda \sum_j \varphi_{ij} A_{ij} \tag{5.1}$$

and

$$\frac{d\varphi_{ij}}{dt} = -(2+\lambda)\varphi_{ij} + \rho_j + \lambda \sum_j \frac{\omega_{ij}\varphi_{jl}}{(1-\rho_j)}(A_{jl} - \delta_{il}) - \lambda \sum_l \frac{\varphi_{ij}\varphi_{il}}{1-\rho_i}(A_{il} - \delta_{lj}), \tag{5.2}$$

describe the SIS dynamics in this theory. Considering the steady state of Eqs. (5.1) and (5.2) and keeping only leading terms up to second order in ρ_i in the latter, we

obtain

$$\varphi_{ij} \approx \frac{(2+\lambda)\rho_j - \lambda\rho_i}{2+2\lambda} - \rho_i\rho_j + O(\rho^3, \lambda\rho^2), \tag{5.3}$$

in which only leading order terms in $\lambda \approx \lambda_c << 1$ [39] for quadratic terms in ρ_i were kept. Finally, replacing the result above in the steady state of Eq. (5.1),

$$\rho_i = \frac{\lambda \sum_j B_{ij}(\lambda)\rho_j}{1 + \lambda \sum_j B_{ij}(\lambda)\rho_j}. \tag{5.4}$$

in which

$$B_{ij} = \frac{2+\lambda}{2+2\lambda} \frac{A_{ij}}{1 + \frac{\lambda^2 k_i}{2+2\lambda}} \simeq \frac{A_{ij}}{1 + \frac{\lambda^2 k_i}{2}} \tag{5.5}$$

is an effective weight adjacency matrix. The last passage was performed considering $\lambda << 1$. Equation (5.4) has exactly the same form of Eq. (4.32) obtained for QMF theory, with B_{ij} taking place of A_{ij}. Therefore, all the description and tools related to the spectral analysis developed for the QMF theory (see Sec. 4.2.2) can be extended to the PQMF case replacing the spectral properties of A_{ij} by those of B_{ij}. The threshold is given by

$$\lambda_c^{\mathrm{PMQF}} \Omega^{(1)}(\lambda_c^{\mathrm{PMQF}}) = 1, \tag{5.6}$$

where $\Omega^{(1)}$ is the LEV of B_{ij}. This result is exactly the same obtained from the Jacobian matrix [39]. The prevalence can be expanded on the eigenvectors of B_{ij}, $\boldsymbol{w}^{(n)}$. For $\lambda \gtrsim \lambda_c^{\mathrm{PMQF}}$, $\rho_i \sim w_i^{(1)}$, in which $w_i^{(1)}$ is the component i of the PEV of B_{ij}. As consequence,

$$\rho \simeq b_1(\lambda \Omega^{(1)}(\lambda_c^{\mathrm{PMQF}}) - 1), \tag{5.7}$$

in which,

$$b_1 = \frac{\sum_i w_i^{(1)}}{N \sum_i (w_i^{(1)})^3}. \tag{5.8}$$

It has exactly the same form of the factor reported in Eq. (4.37). Finally, the IPR for the PEV of B_{ij},

$$Y_4(\boldsymbol{w}^{(1)}) = \sum_{i=1}^{N} (w_i^{(1)})^4, \tag{5.9}$$

allows to quantify the localization in the PQMF theory.

5.3 Accuracy of theoretical estimates for epidemic threshold

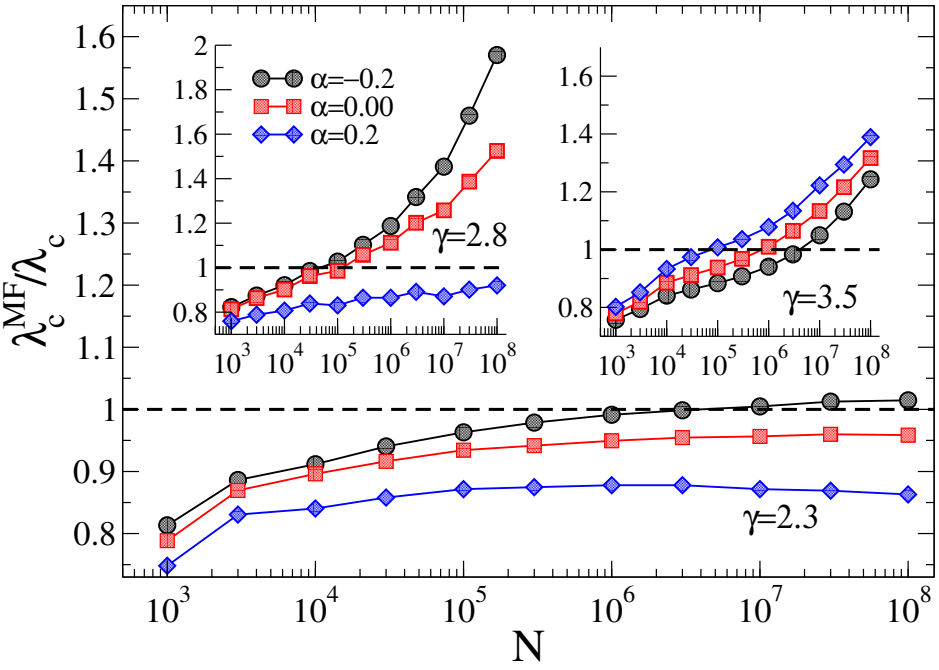


Figure 5.2: Ratio between thresholds of HMF (λ_c^{MF}) and simulations (λ_c) as a function of the network size for different values of γ and α. Main panel, right and left insets correspond to $\gamma = 2.3, 2.8$, and 3.5, respectively. An upper cutoff $k_c = 2\sqrt{N}$ ($k_c \sim N^{1/\gamma}$) is considered for $\gamma < 3$ ($\gamma > 3$). Averages correspond to 10 network realizations.

The accuracy of HMF theory is tested with respect to simulations in Figure 5.2. The existence of correlation reveals a non-negligible asymptotic discrepancy between them when $\gamma = 2.3$. The theory overestimates (underestimates) the threshold for disassortative (assortative) networks. For larger values of γ, this discrepancy is remarkable, and the epidemic threshold is significantly overestimated, as can be seen in the insets of Figure 5.2.

When the comparison is performed considering QMF and PQMF, both theories appear to converge asymptotically to the epidemic threshold observed in simulations when $\gamma = 2.3$, as shown in Figure 5.3(a). Moreover, for this value of γ, the PQMF theory displays a faster convergence than the QMF theory, mainly for smaller α. The qualitative success of PQMF and QMF theories in predicting a null threshold in the

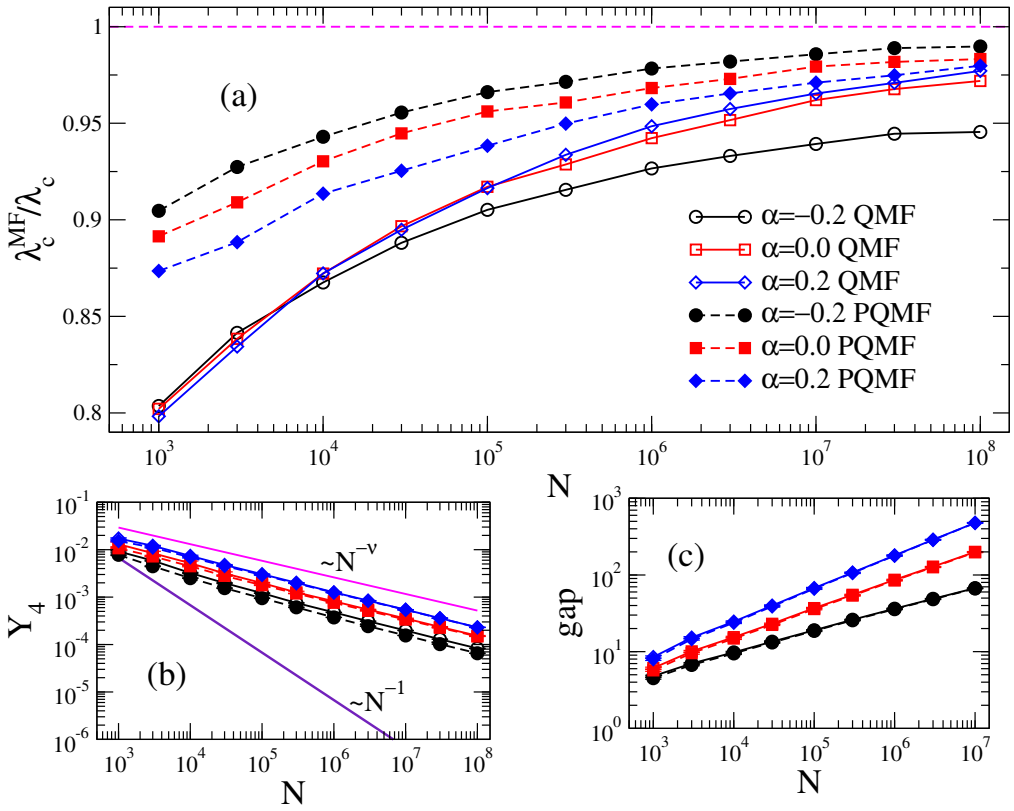


Figure 5.3: (a) Comparison of the QMF and PQMF theories, (b) IPR and (c) spectral gap of A_{ij} and B_{ij} against size for $\gamma = 2.3$ and different values of α. Averages correspond to 10 networks realizations. In (b), solid lines are power-law decays $Y_4 \sim N^{-\nu}$, with $\nu = (3-\gamma)/2$ and $Y_4 \sim N^{-1}$ corresponding to subextensive localization in the maximum K-core and delocalization, respectively. Solid lines and open symbols correspond to the QMF theory and A_{ij}, while dashed lines and closed symbols correspond to the PQMF theory and B_{ij}.

thermodynamic limit regardless of the correlation pattern is not repeated quantitatively for $\gamma = 3.5$ as shown in Figure 5.4(a). Theoretical estimates of the threshold deviate from simulation results leading to decreasing ratios $\lambda_{\mathrm{MF}}/\lambda_{\mathrm{c}}$ in the large network limit. The rigorous results [96] suggest this ratio would decrease asymptotically as $1/\ln(k_{\max})$ [92]. Finally, PQMF theory presents a better quantitative performance and the relative improvement over QMF theory growths with α.

5.3.1 Relation to spectral properties

What is the origin of the discrepancies between theoretical predictions and numerical results observed in Sec. 5.3 ? In this section we investigate which spectral feature is correlated with the performance of the theoretical approaches. We consider

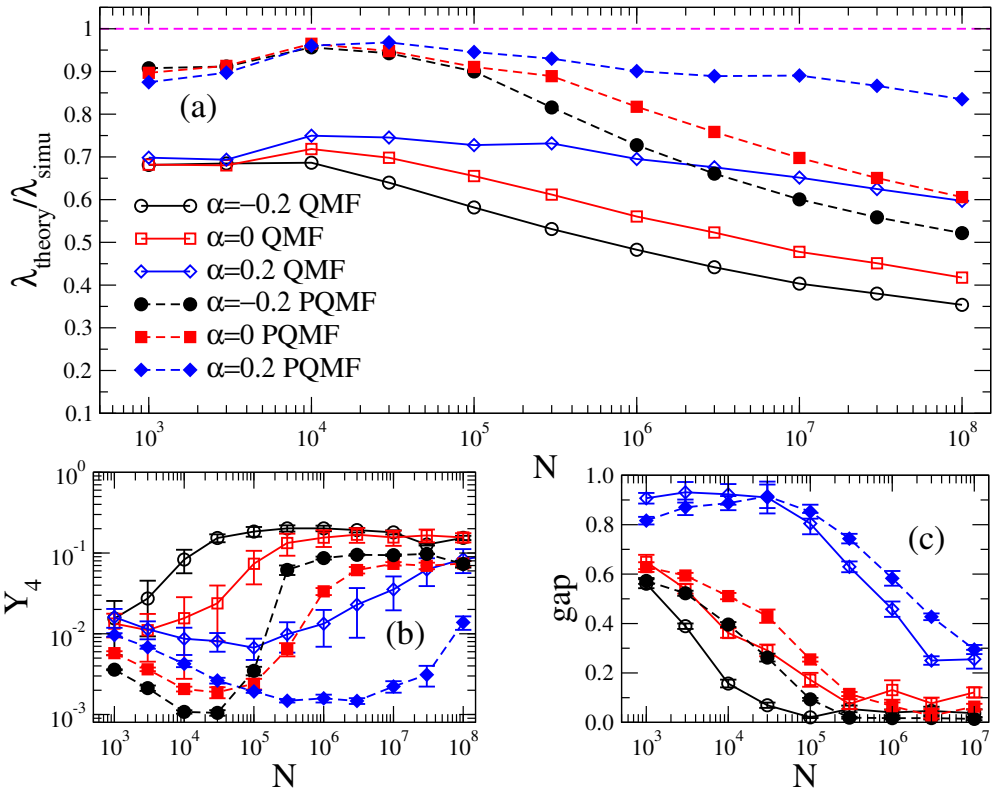


Figure 5.4: (a) Comparison of the QMF and PQMF mean-field theories, (b) IPR and (c) spectral gap of A_{ij} and B_{ij} against size for $\gamma = 3.5$ and different values of α, using an upper cutoff $k_c \sim N^{1/\gamma}$. Averages correspond to 10 networks realizations. Solid lines and open symbols correspond to the QMF theory and A_{ij}, while dashed lines and closed symbols correspond to the PQMF theory and B_{ij}.

both QMF and PQMF theories, testing their accuracy against the spectral properties of adjacency matrices A_{ij} and B_{ij}, respectively.

Let us consider first the case $\gamma = 3.5$. The real threshold is not the QMF one because the PEV is localized. As pointed out in Ref. [44], this in principle implies that the actual threshold coincides with the inverse of the largest eigenvalue corresponding to a delocalized PEV, coinciding with the HMF threshold $\lambda_c^{\mathrm{HMF}} = \langle k \rangle / \langle k^2 \rangle$. Actually, however, the QMF neglects dynamical correlations which have the effect of allowing mutual reinfection events among different hubs in the network. In this way an endemic global state can be established due to the long-range interactions among localized states [92] setting the actual threshold to an intermediate value: $\lambda_c^{\mathrm{QMF}} < \lambda_c < \lambda_c^{\mathrm{HMF}}$. With this picture in mind, we can predict that, if the localization is stronger (higher values of the IPR Y_4), the actual threshold will be farther from λ_c^{QMF} and thus the

performance (accuracy) of the QMF approach will be reduced.

We plot the dependence of Y_4 on the system size N for $\gamma = 3.5$ in Figure 5.4: The IPR of A_{ij} converges to a finite value in the thermodynamic limit, irrespective of the correlation degree, representing a PEV localized on a finite set of nodes [44,79]. The saturation with size occurs earlier for disassortative correlations and later for assortative, compared to the uncorrelated case. In general, for a given size N, Y_4 is larger for smaller α. As conjectured, a better QMF performance occurs for smaller Y_4.

The IPR analysis for the PQMF theory, involving B_{ij}, has a qualitatively similar behavior to QMF theory, but presents lower values for the IPR. Hence, the PQMF steady-state solution is less localized than that of the QMF theory. Correspondingly, the PQMF performance is better than the QMF performance. We also calculate, in Figure 5.4(c), the dependence of the spectral gap on the system size, for both the adjacency matrix A_{ij}, involved in QMF theory, and B_{ij}, in PQMF theory. The spectral gap is defined as the difference $\Lambda^{(1)} - \Lambda^{(2)}$ between the largest and second largest eigenvalues of the adjacency matrices. The gap of the adjacency matrix A_{ij} is small and it decreases as N grows, as predicted by Ref. [76]. The gap is smaller for smaller α. The dependence of the spectral gap of B_{ij} on size is qualitatively similar to the gap of A_{ij}.

Let us consider the case $\gamma = 2.3$. In this case, the physical mechanism underlying the epidemic transition is different, as it does not involve the interaction between distant hubs but rather the extension of activity from the maximum k-core to the rest of the network. The connection between QMF theory performance and localization is not easily predictable. As shown in Figure 5.3(b), the IPR for $\gamma = 2.3$ follows a power-law $Y_4 \sim N^{-\nu}$, with $\nu \approx (3-\gamma)/2$, which is associated to the IPR localized in the maximum k-core of the network [79]. Correlations leave the scaling exponent unchanged, altering only the prefactor: The smaller the α, the smaller the IPR. This means that the PEV is still localized on a subextensive fraction of nodes. However, since Y_4 increases with α, the PEV is more localized for positive α than for negative. The same is true for the matrix B_{ij} of the PQMF theory. Interestingly, the effect on the performance of the theoretical approaches is opposite. The QMF theory works better for larger Y_4 and the

PQMF theory works better for smaller Y_4. We have no simple interpretation for this result.

Figure 5.3(c) shows the spectral gap for the WPCM networks. In this case, the gap increases with the network size and it is smaller for smaller α. This is true also for the gap spectral of the PQMF theory. Finally, let us observe that there is almost no difference between the spectral properties of A_{ij} and B_{ij} for $\gamma = 2.3$. This is indeed not surprising for $\alpha = 0$, since the term $\lambda^2 k_i$ in the denominator of Eq. (5.5) is asymptotically negligible, because $\lambda_c k_{\max}^2 \sim k_{\max}^{5-2\gamma} \to 0$ as $N \to \infty$.

5.3.2 Intermediate case $5/2 < \gamma < 3$

As for the other values of γ, in this range the vanishing of the epidemic threshold with N is observed regardless of the correlation pattern. The localization phenomenon of the PEV in the case $5/2 < \gamma < 3$ is asymptotically equivalent to the case $\gamma > 3$. However very strong crossover effects are observed in this case, because of the presence of a localization process on the maximum k-core ($\gamma < 5/2$) competing with the localization around the hub [79]. As a consequence, already in the uncorrelated case, the PEV gets strongly localized around the largest hub only for very large values of N. Correlations further complicate the picture: Figure 5.5(b) shows that disassortative correlations accelerate the convergence to the final localized state. For $\alpha > 0$ instead, Y_4 is a decreasing function of N. The upward bend of the curve hints at an incipient crossover, but one cannot exclude that the asymptotic behavior is different for $\alpha > 0$. A similar pattern is observed for the case of the spectral gap as shown in Figure 5.5(c).

With regard to the performance of the theoretical approaches, for negative or zero correlations the scenario perfectly matches what happens for $\gamma > 3$, in which all theories somehow fail in capturing the way the threshold vanishes, with the PQMF theory being more accurate than the others. In the case $\alpha = 0.2$, numerical results seem to suggest that both theories describe quite well how the threshold changes with the system size. However, the large crossover effects mentioned above do not allow one to draw any firm conclusion.

We can summarize our findings by stating that the performance in predicting the

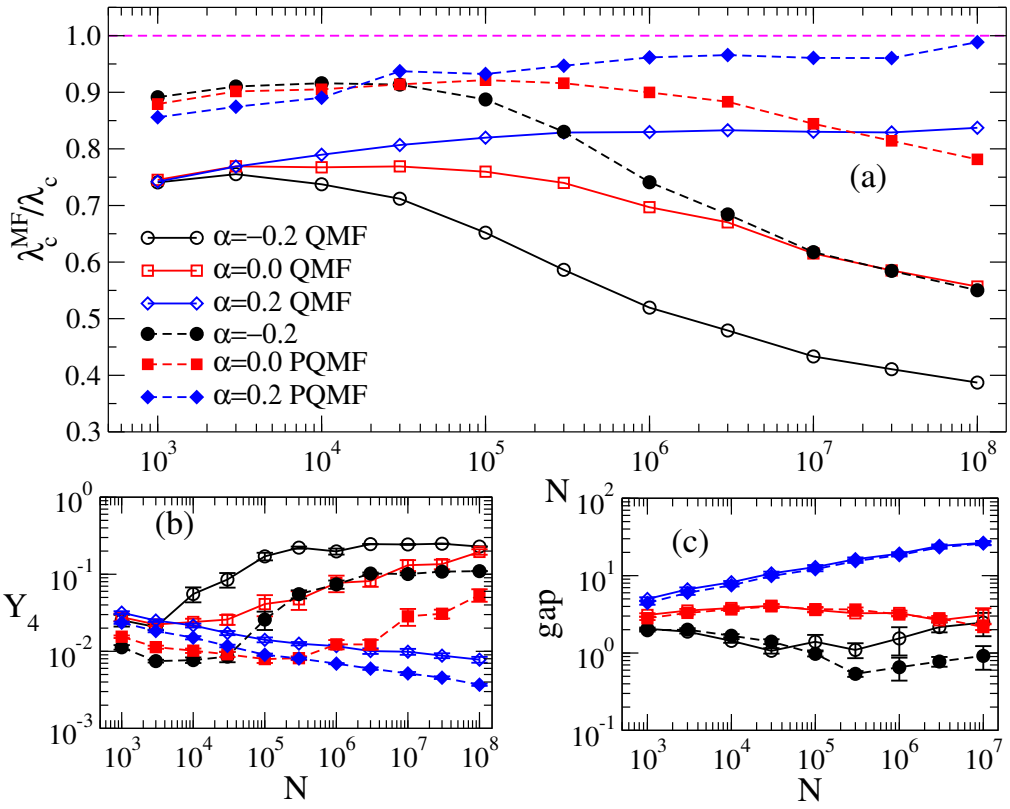


Figure 5.5: (a) Comparison of the QMF and PQMF mean-field theories, (b) IPR and (c) spectral gap of A_{ij} and B_{ij} against size for $\gamma = 2.8$ and different values of α, using an upper cutoff $k_c = 2\sqrt{N}$. Averages correspond to 10 networks realizations. Solid lines and open symbols correspond to the QMF theory and A_{ij}, while dashed lines and closed symbols correspond to the PQMF theory and B_{ij}.

behavior of epidemic threshold of the QMF and PQMF theories on WPCM networks is correlated with the size of the spectral gap and the IPR of the PEV of the respective A_{ij} and B_{ij} matrices that rule the prevalence near the transition point. A large spectral gap or a low IPR leads to a good performance of the mean-field theories, while the converse, small gap or large IPR, leads to deviations from the theoretical predictions. The QMF theory seems to be more correlated with the spectral gap, while the PQMF theory is with the IPR, at least in the regime where the gap is significant and the theories are accurate.

5.3.3 Analysis on Real networks

We extend our analysis to a set of 99 networks composed by social, technological, biological, infrastructure networks having different sizes and topological features was considered. The list and some properties of these networks are presented in appendix A). The spectral gap and the IPR of matrices A_{ij} and B_{ij} are compared in the scatter plots shown in Figures 5.6(a) and (b). We see that the spectral gap is almost the same for both matrices, while the IPR extracted from B_{ij} is smaller than the one extracted from A_{ij}, in particular in the range of large IPR values. This shows that the PQMF matrices B_{ij} are less localized than the matrix A_{ij} which emerges in the QMF theory. The relative errors between the QMF or PQMF theory and simulations, defined as

$$\varepsilon = \frac{\lambda_c - \lambda_c^{MF}}{\lambda_c}, \tag{5.10}$$

are compared in the scatter plot shown in Figure 5.6(c). As in the case of random synthetic networks, the PQMF theory outperforms the QMF theory for all networks investigated.

On this set of networks we test the relation observed for synthetic correlated networks, connecting qualitatively the accuracy of QMF and PQMF threshold predictions with the properties of the adjacency matrices (spectral gap and IPR), respectively, and with the level of correlation determined by the Pearson coefficient P defined in Eq. (2.9) that lies in the interval $-1 < P < 1$, being negative for disassortative, null for uncorrelated, and positive for assortative networks. The analyses for QMF theory are shown in the scatter plots of the relative error ε against the corresponding topological properties in Figures 5.6(d) and (e). Qualitatively similar patterns are obtained for the PQMF theory. We can see that in real networks, the correlation between the performance of the theoretical prediction and the spectral gap is on average the same as that observed for the WPCM: A larger spectral gap is associated with a higher accuracy. The inverse correlation with the IPR is again preserved: A smaller Y_4 corresponds to a more accurate prediction. We do not find a significant correlation with the Pearson coefficient. Statistical analyses were performed using the correlation

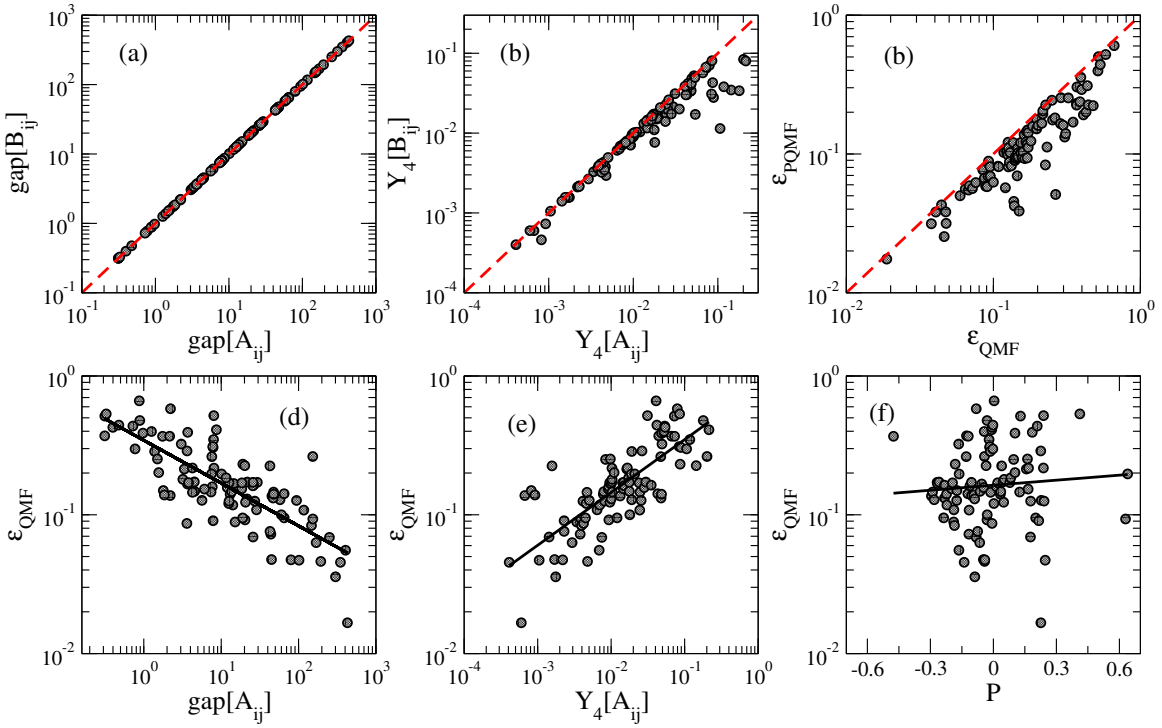


Figure 5.6: Scatter plots for a set of 99 real networks (see appendix A). Each point corresponds to a single network. (a) Spectral gap and (b) IPR of the matrix B_{ij} plotted versus the corresponding values for the matrix A_{ij} . (c) Relative errors of QMF and PQMF theoretical predictions with respect to the simulation, defined by Eq. (5.10). Dashed red lines denote the diagonal. Relative errors of the QMF theory are plotted vs (d) the spectral gap, (e) the IPR, and (f) the Pearson coefficient.

coefficients obtained from either power-law, in the case of the spectral gap and IPR, or from exponential, in the case of the Pearson coefficient, regressions of the scatter plots. We obtain strong statistical correlations with $|r| \gtrsim 0.70$ ($p - \text{value} < 10^{-5}$) for both QMF and PQMF theories using either the IPR and the spectral gap of the corresponding matrices. Values $|r| \lesssim 0.20$ ($p - \text{value} > 0.05$) for correlation with the Pearson coefficient of the network confirms no significant statistical correlations.

5.4 Summarizing this chapter

The determination of the epidemic threshold in models of disease spreading in complex topologies is a nontrivial problem in network science. Several theoretical approaches have been proposed, applying approximations with different levels of stringency, that provide contrasting predictions on the epidemic threshold. Among

the main theoretical approaches at the mean-field level we can consider, in decreasing order of approximation, the heterogeneous mean-field theory, neglecting dynamical correlations and the actual pattern of connections in the network (preserving only its statistical properties); the quenched mean-field theory, also neglecting dynamical correlations but keeping the network structure; and the pair-quenched mean-field theory, which incorporates dynamical correlations between pairs of connected nodes. In this chapter we have presented a comparison of the predictions of these three approximated theories for the case of the SIS epidemic model, focusing on the case of networks with a power-law degree distribution and degree correlations, representative of many real network systems.

Comparing the predictions with actual stochastic simulations of the SIS process, we observed that, independently of the degree of correlations, the predictions of PQMF theory are more accurate than those of QMF theory, while both outperform HMF theory, which fails to predict the vanishing threshold observed for a degree exponent $\gamma > 3$. While in general PQMF theory is more accurate than QMF theory, the two approximations show different levels of accuracy when compared in networks with different levels of correlations. Thus, for the case of synthetic networks generated with the Weber-Porto algorithm [69] (see Sec. 3.1.1), we observed that, for fixed network size and degree heterogeneity, QMF predictions are more accurate in assortative networks than in disassortative networks. On the other hand, PQMF theory is increasingly accurate in the presence of disassortative correlations for small degree exponent, while it is more accurate when correlations are assortative if the degree exponent is large.

We proposed a criterion for the accuracy of the QMF and PQMF approaches based on the spectral properties of the networks. We observed that the accuracy is positively correlated with the amplitude of the spectral gap of the adjacency matrix and is inversely related to the degree of localization of the principal eigenvector, as measured by the inverse participation ratio. This general observation was corroborated by the analysis of a large set of real correlated networks, covering a wide range of sizes and topological features.

Chapter 6

Non-perturbative analysis of the high prevalence regime for the SIS model

In this chapter, the analysis of mean-field approaches' accuracy will be extended to the prevalence. We perform a nonperturbative analyses of both QMF and PQMF theories for SIS on networks using numerical integration of the corresponding dynamical equations. Firstly, we investigate the regime of low prevalence, in which a critical behavior of the form $\rho \sim (\lambda - \lambda_c)^\beta$ is expected. We study a special case in which the QMF theory presents high accuracy in predicting the epidemic threshold. Finally, we analyze the regime of high prevalence, aiming at filling this gap of the literature, using synthetic and real networks, presenting degree-correlation and power-law degree distribution. The relation between some topological metrics, spectral properties and the accuracy of QMF and PQMF theories in predicting the prevalence is investigated. This chapter presents the results published in Ref. [48].

6.1 Prevalence of SIS model in mean-field theories

In chapter 5, we verified that despite the detailed microscopic description of the QMF theory, neglecting dynamical correlations in the SIS model can lead to modest

accuracy with significant deviations from simulations if the epidemic involves, for example, activation localized in the hubs that spreads to the rest of network [18,31]. The inclusion of dynamical correlations reckoned by individual pairwise interactions greatly improve the predictions of the epidemic thresholds of the QMF theory in the PQMF theory.

Since PQMF theory has been mainly analyzed perturbatively in the limit of very low prevalence to investigate the position of the epidemic thresholds, Matamalas et al. [82] claimed that it has limitations to compute high-epidemic-incidence regimes and proposed that a microscopic Markov chain approach (MMCA) [97], which is a discrete-time version of the QMF theory, could be used instead. However, a nonperturbative approach is possible through numerical integration of both QMF and PQMF dynamical equations. Since large discrepancies between discrete- and continuous-time approaches can be present in the SIS dynamics [98], a nonperturbative analysis of QMF and PQMF theories is necessary.

We perform a nonperturbative analyses of both QMF and PQMF theories for SIS on networks using numerical integration of the corresponding dynamical equations. We consider both large synthetic networks generated with the Weber-Porto model (see Sec. 3.1.1) and real networks with different levels of heterogeneity, sizes and degree correlations. We numerically integrated QMF and PQMF equations using a fourth-order Runge-Kutta method with time step $\delta t = 10^{-2}$ to 10^{-1}. Initial conditions consistent with the exact closure equations relating pairwise and single node probabilities such as $\phi_{ij} + \psi_{ij} = \rho_j$ must be chosen and the steady state is insensitive to a particular choice. An example of the robustness of the steady state in relation to an initial condition choice is shown in Figure 6.1, where we adopted $\rho_i = \bar{\rho}$ and $\phi_{ij} = \rho_i(1 - \rho_j)$ with different values of $\bar{\rho}$ for a fixed infection rate. We performed stochastic simulations of the SIS dynamics on networks using the optimized Gillespie algorithm. The absorbing states, which are rigorously the unique real stationary state in finite-size networks, were circumvented using quasistationary simulations. A description of how to simulate SIS dynamics is found in Sec. 4.4.

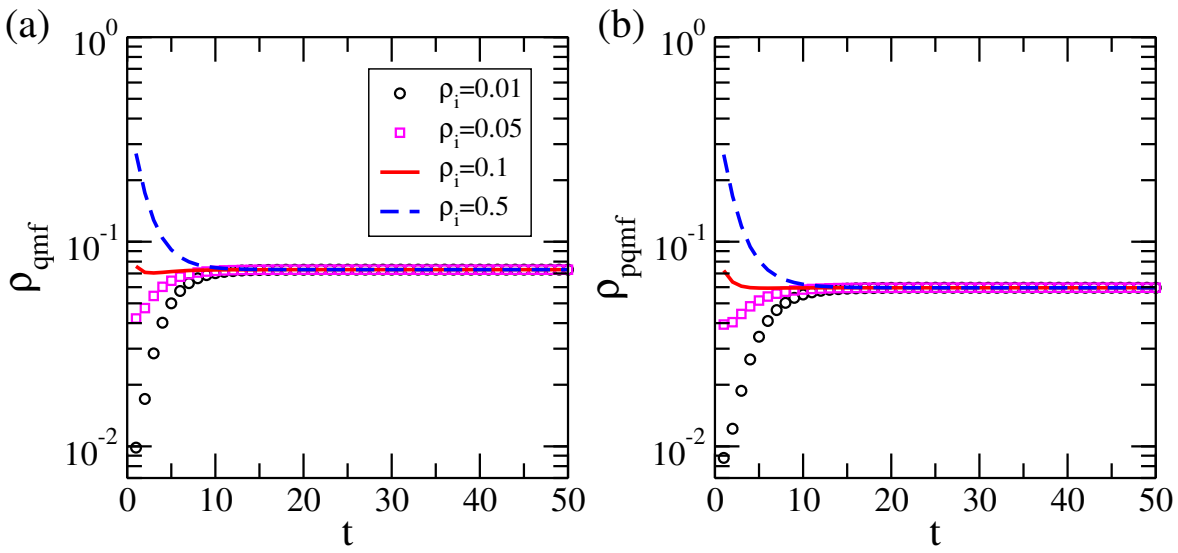


Figure 6.1: Robustness of the steady state determined by integration of (a) QMF and (b) PQMF theories considering different uniform initial conditions. In the initial conditions considered, the prevalence in each node is $\rho_i = \bar{\rho}$ and the pairs are given by $\phi_{ij} = (1-\rho_i)\rho_j$. The distinct values of $\bar{\rho}$ are indicated in the legend. The integrations were performed on a UCM network of size $N = 10^6$ and $\gamma = 2.8$ considering the healing and infection rates $\mu = 1$ and $\lambda = 0.1$, respectively.

6.1.1 Epidemic prevalence near to the epidemic threshold

Figure 5.3 shows that the QMF prediction for the epidemic threshold tends to the same limit of numerical simulations for both uncorrelated and correlated networks for $\gamma = 2.3$. This observation naturally leads one to wonder what is the extent of accuracy of the QMF theory for SIS dynamics on random networks with $\gamma < 5/2$. In order to answer this question we test the other prediction of the QMF theory, Eq. (4.37), stating that the fraction of infected individuals decays to zero linearly as the threshold is approached from above. Numerical results, for the case of uncorrelated networks $\alpha = 0$ are shown in Figure 6.2, in which the density and the infection rates are rescaled to conform to Eq. (4.37). We can clearly see the existence of two scaling regimes. For $\lambda\Lambda^{(1)} - 1 \ll 1$ the density scales with an exponent larger than the prediction $\beta^{\mathrm{QMF}} = 1$ (see Sec. 4.2.2). The observed exponent is consistent with the exact result of Ref. [32] (see Sec.4.3), $\beta = 1/(3-\gamma)$, which is also the value predicted by the HMF theory [34,35] (see Sec. 4.2.1). This exponent is observed in a regime very close to the transition, where the system is kept asymptotically active only by virtue of the QS

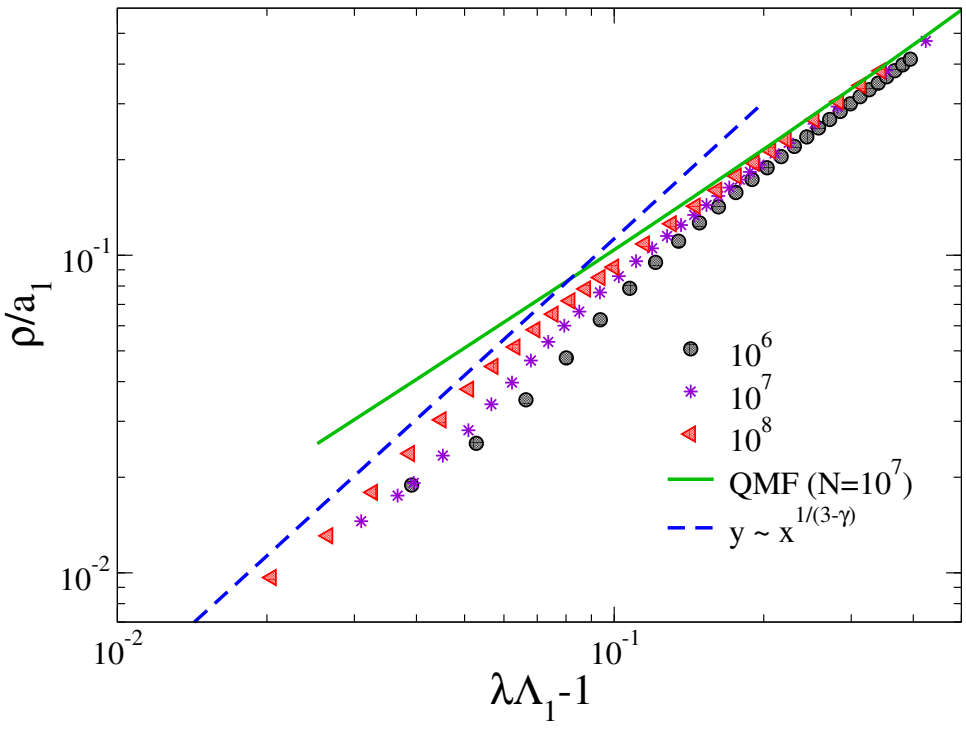


Figure 6.2: Rescaled average density as a function of the distance from the epidemic threshold. Quasistationary simulations for different sizes are indicated in the legend. The solid line is a numerical integration of the QMF theory Eq. (4.25) for $N = 10^7$, while the dashed line is a power law with exponent predicted analytically in Ref. [32]. The results were obtained on uncorrelated networks ($\alpha = 0$) and $k_c = 2\sqrt{N}$.

method. A comparison between simulation results and a nonperturbative analysis by integrating the QMF equation, Eq. (4.37), for $\lambda > \frac{1}{\Lambda_{(1)}}$ and $N = 10^7$ confirms that the QMF theory correctly predicts the linear behavior of the prevalence ρ around the epidemic transition, but only sufficient far from it. In the immediate neighborhood of the threshold the decay is more rapid.

On a linear scale[1], the region that departs from linearity is squeezed around $\lambda = \lambda_c \to 0$ as the network size increases, as indicated by the horizontal arrows in Figure 6.3, in which simulations and QMF theory are compared. The slope of the linear region decreases with size since a_{QMF} defined by Eq. (4.38), also does: We found $a_{\mathrm{QMF}} = 0.00382$ for $N = 10^7$ and $a_{\mathrm{QMF}} = 0.00130$ for $N = 10^8$. Finally, we see that QMF is not able to capture quantitatively the amplitude of the linear region observed in simulations reinforcing the need of nonperturbative analyses of the PQMF theory.

[1]Obviously, this region will not shrink in a logarithm scale and the asymptotic scaling is the theoretical one given by Eq. (4.37).

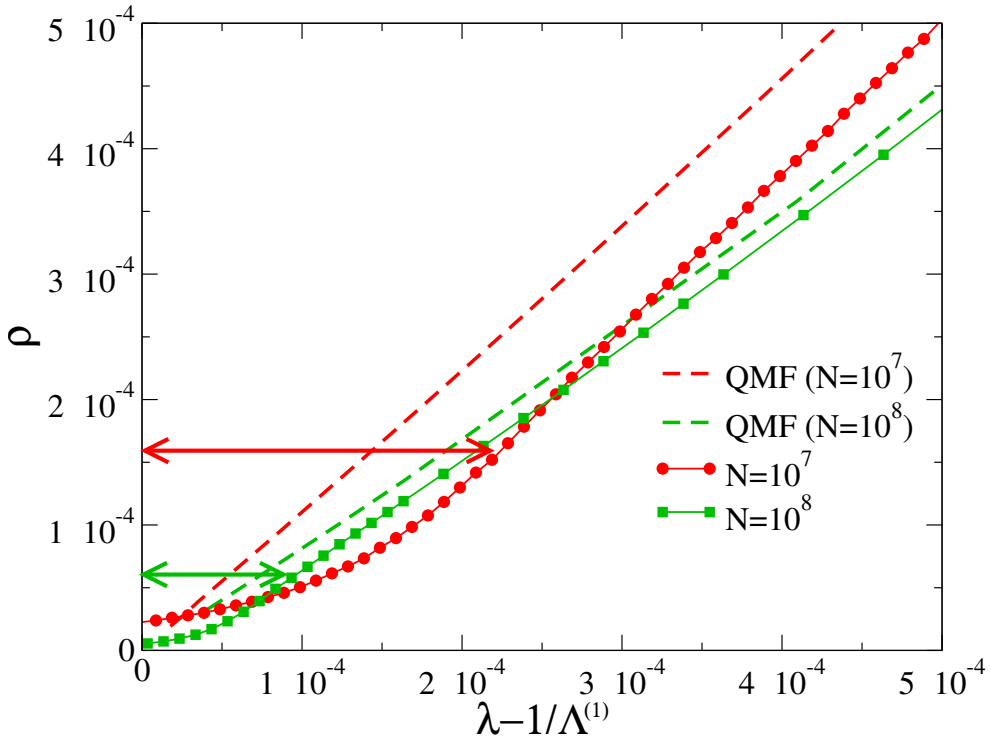


Figure 6.3: Epidemic prevalence around the threshold $\lambda_c = \frac{1}{\Lambda_1}$ for UCM networks with $\gamma = 2.3$, $k_{\min} = 3$ and $k_c = 2\sqrt{N}$. Simulations are represented by solid lines with symbols and numerical integration of the QMF Eq. (4.25) is given by the dashed lines. Horizontal arrows indicate the interval where curves depart from linearity.

6.1.2 Synthetic networks

Epidemic prevalences obtained from theories and stochastic simulations are compared in Figures 6.4, 6.5 and 6.6. PQMF outperforms QMF theory in all investigated cases, as intuitively expected since the former includes higher-order correlations. While QMF theory deviates from simulation for regimes of high densities of infected nodes, PQMF cannot be distinguished from simulations in the presented scales. Under low-density regimes, QMF and PQMF agree very well and are indistinguishable from simulations for $\gamma = 2.3$ as shown in Figure 6.4, while larger deviations of QMF can be seen for larger values of γ, as shown in Figures 6.5 and 6.6.

The accuracy of the theories at low prevalence is better for assortative and worse for disassortative networks compared with the neutral case. Another interesting dependence on the assortativity can be observed in these curves. For low prevalences, assortative and disassortative networks possess, respectively, higher and lower den-

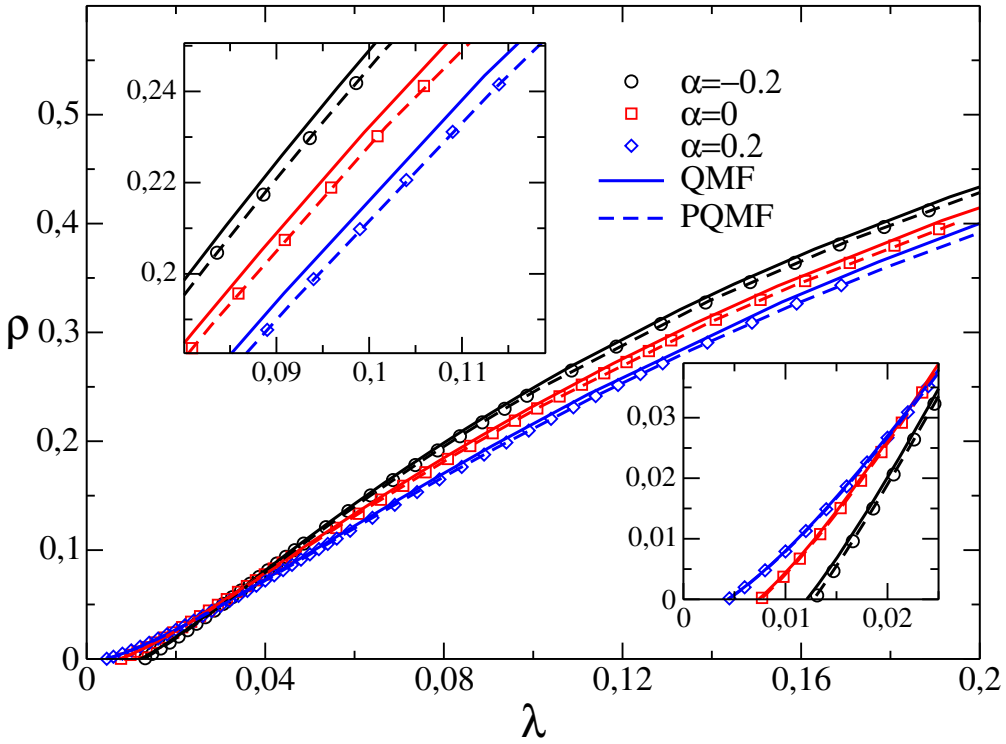


Figure 6.4: Epidemic prevalence as a function of the infection rate for WPCM networks with $N = 10^6$, $\gamma = 2.3$ and different levels of degree correlations. Symbols represent stochastic simulation while solid and dashed lines the numerical integration of the QMF and PQMF theories, respectively. Bottom and top insets show zoom of low- and high-prevalence, respectively.

sities compared with the uncorrelated networks. At high prevalences, the converse is observed, in which disassortative networks present higher densities than the assortative and neutral networks. The same behavior is observed for all values of γ, indicating that it is related to the degree correlations. The behavior at low densities can be explained in terms of the reduced capacity to transmit infection when hubs are surrounded by low-degree nodes in the disassortative case rather than being directly connected with higher probability in the assortative case. We cannot provide simple arguments for the inverted dependence on the assortativity degree at high densities, but it is very precisely reproduced by the PQMF theory.

We also investigate the role of the heterogeneity comparing the PQMF theory with a pair homogeneous mean-field (PHMF) theory [99]. A closed solution for Eqs. (4.41) and (4.44) can be derived for the particular case of homogeneous networks where $P(k) = \delta_{k,m}$ for which $\rho_i = \rho$ and $\phi_{ij} = \phi$. The expression for stationary

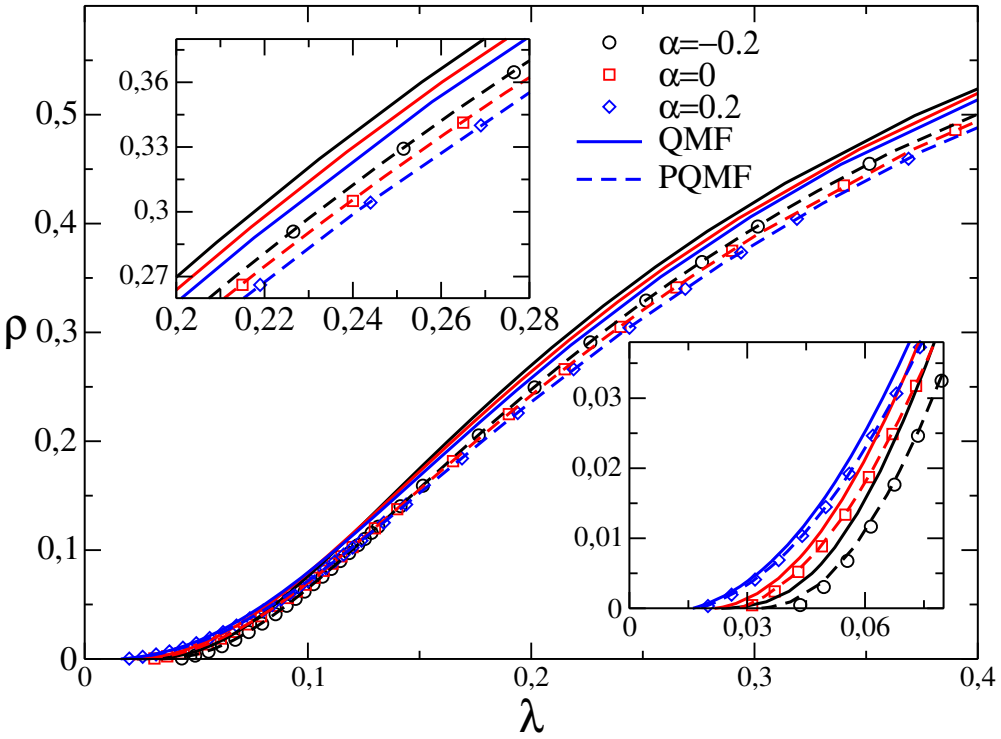


Figure 6.5: Epidemic prevalence as a function of the infection rate for WPCM networks with $N = 10^6$, $\gamma = 2.8$ and different levels of degree correlations. Symbols represent stochastic simulation while solid and dashed lines the numerical integration of the QMF and PQMF theories, respectively. Bottom and top insets show zoom of low- and high-prevalence, respectively.

epidemic prevalence is

$$\rho = \frac{\lambda - \lambda_c}{\langle k \rangle^{-1} + \lambda - \lambda_c}, \tag{6.1}$$

where

$$\lambda_c = \frac{1}{\langle k \rangle - 1} \tag{6.2}$$

is the epidemic threshold predicted by the PHMF theory. This solution is the same one derived for the contact process (CP) in, e.g., Ref. [99], replacing the infection rate $\lambda_{CP} = m\lambda_{SIS}$, where λ_{CP} is the infection rate for the contact process and λ_{SIS} is the infection rate for SIS.

To perform the comparison between the PQMF an PHMF theories, we replaced m in Eq. (6.1) by the average degree $\langle k \rangle$ of the network [100,101]. The densities of infected nodes obtained in both pairwise approaches are shown in Figure 6.7. Beyond the expected discrepancy for describing the low-prevalence regimes, since one theory

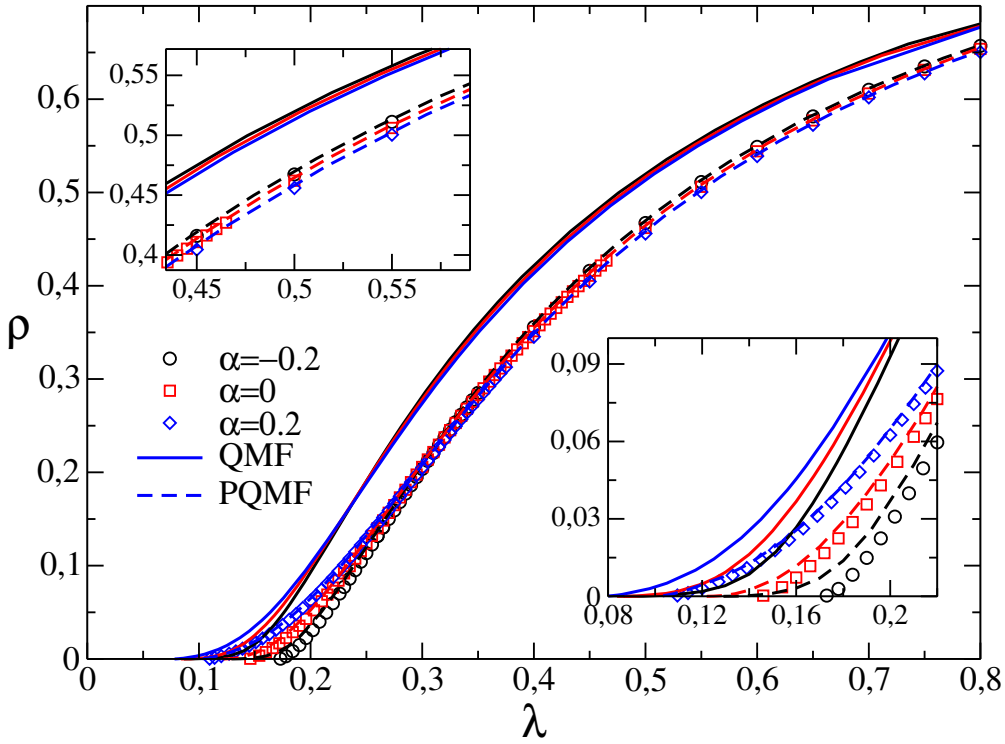


Figure 6.6: Epidemic prevalence as a function of the infection rate for WPCM networks with $N = 10^6$, $\gamma = 3.5$ and different levels of degree correlations. Symbols represent stochastic simulation while solid and dashed lines the numerical integration of the QMF and PQMF theories, respectively. Bottom and top insets show zoom of low- and high-prevalence, respectively.

predicts a finite and the other a vanishing threshold, the regime of high epidemic prevalence is affected by the inclusion of heterogeneity. As one could expect, the more heterogeneous networks present the larger discrepancies between homogeneous and heterogeneous theories.

6.1.3 Real networks

Real networks usually present some degree of correlation and, in many cases, the patterns can be quite complex exhibiting both assortative and disassortative correlations for distinct ranges of degree [1,11]. Therefore the comparison between mean-field theories and simulations are necessary in order to determine to which extent the accuracy observed in synthetic networks holds in the real-world counterparts. We selected some networks with different levels of heterogeneity, sizes, and correlations used in

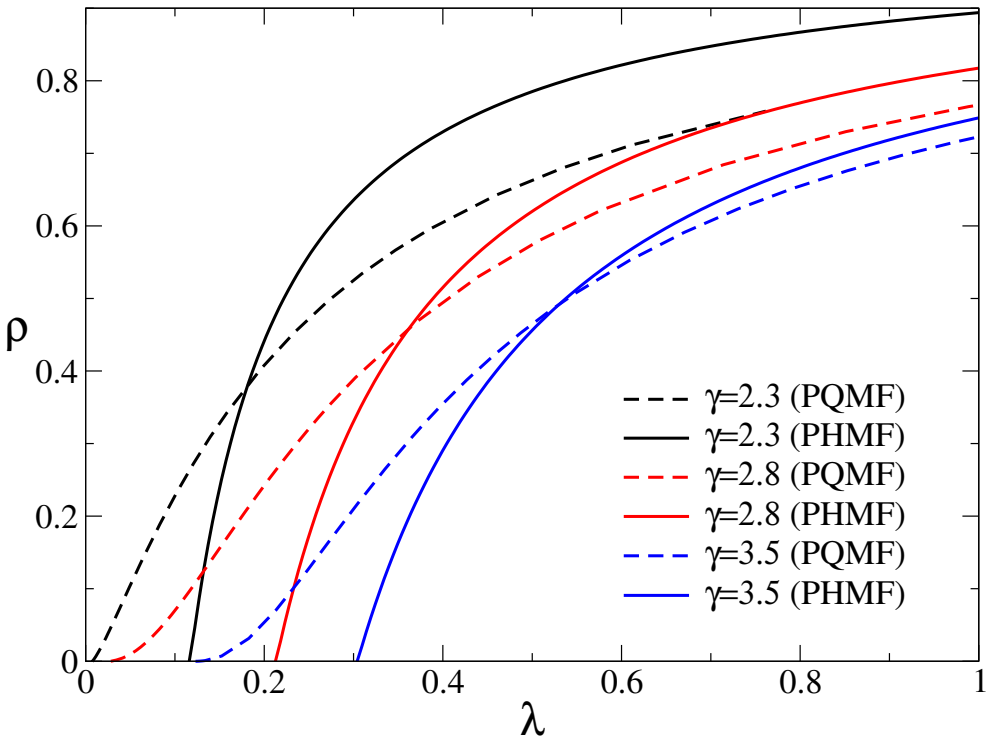


Figure 6.7: Comparison of pair-quenched (dashed lines) and pair-homogeneous (solid lines) mean-field theories for uncorrelated networks with different levels of heterogeneity. The network size is $N = 10^6$.

chapter 5. Only the largest connected components are used if there is more than one.

Figure 6.8 presents the prevalence as a function of the infection rate for 12 real networks. We remark that data asymptotically close to the epidemic threshold mismatch simulations, as shown in chapter 5, and are beyond the scope of this chapter. In some cases, QMF and PQMF are indistinguishable from each other and agree almost perfectly with simulations in the scale presented in these figures. In other cases, QMF theory deviates considerably from simulations while PQMF theory remains accurate. In order to quantify the differences we define the relative deviation of densities obtained in simulations ($\rho_{\rm sim}$) with the QMF theory ($\rho_{\rm QMF}$) as

$$\eta_{\rm QMF} = \frac{\int_{\lambda_1}^{\lambda_2} [\rho_{\rm QMF}(\lambda) - \rho_{\rm sim}(\lambda)] d\lambda}{\int_{\lambda_1}^{\lambda_2} \rho_{\rm sim}(\lambda) d\lambda}, \tag{6.3}$$

where λ_1 and λ_2 are the initial and final infection rates in the simulations presented in Figure 6.8, and an equivalent definition of $\eta_{\rm PQMF}$ for the PQMF theory. The intervals

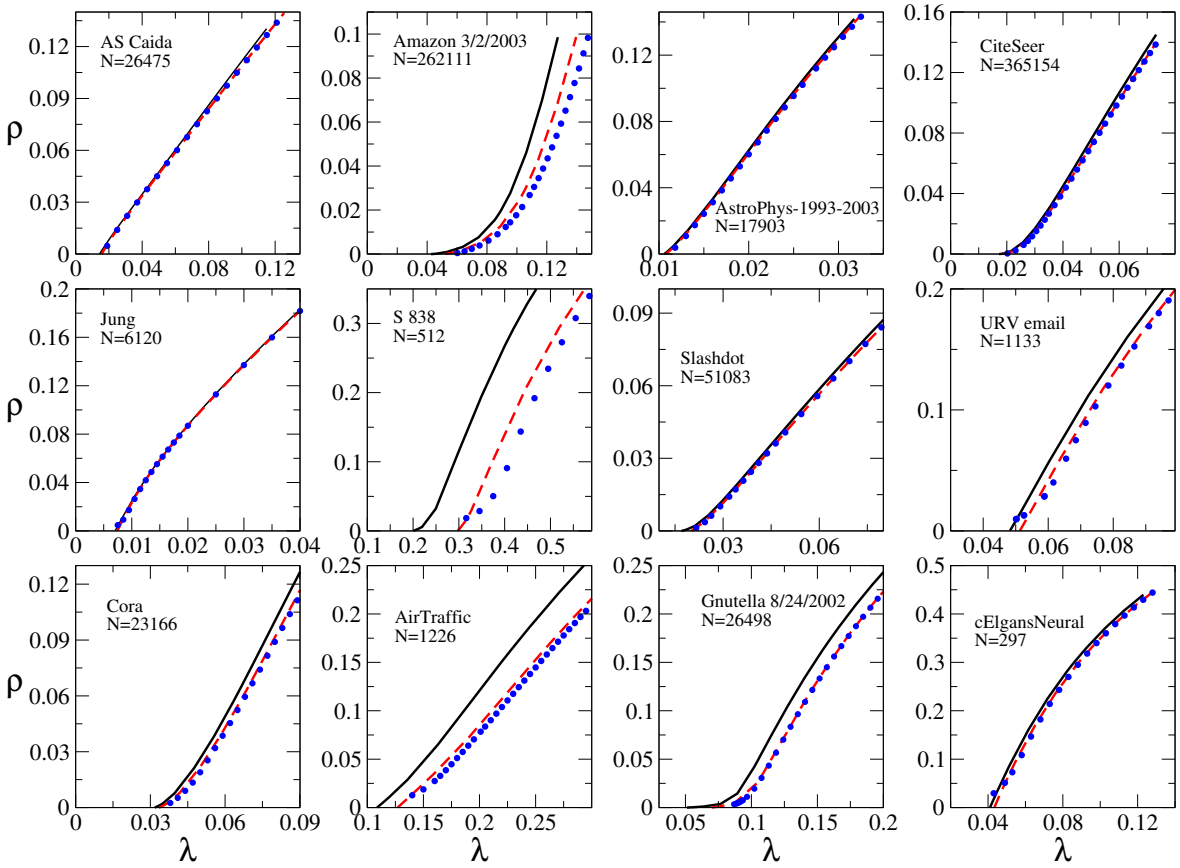


Figure 6.8: Epidemic prevalence on real networks. Symbols represent stochastic simulations while solid and dashed lines represent the numerical integration of the QMF and PQMF equations, respectively. In each panel, the usual name and size of networks are given. See appendix A for other properties of these networks.

$[\lambda_1, \lambda_2]$ were chosen to include sufficiently high densities to guarantee that the effects of the absorbing states are small and prevalences are not close to 1. The relative deviations are listed in Table 6.1.

The positivity of η shows that the mean-field theories overestimate the density obtained in simulations as expected since dynamical correlations, pruned in mean-field theories, reduce the spreading capacity of the epidemic process. As can be seen, we have $\eta_{\text{PQMF}} \ll \eta_{\text{QMF}}$ such that PQMF is always more precise than QMF. Three networks present significant deviations of the PQMF theory from the simulations, namely, for Amazon customers, electronic circuit S838, and Air Traffic networks, with 24%, 7%, and 6.6% of deviation, respectively.

6.1.4 Accuracy versus structural properties

The gain of PQMF theory with respect to QMF in real networks is expressive but it still deviates from simulations in some cases, as shown in Table 6.1. One central question is to determine when either QMF or PQMF performance is satisfactory. Near the transition point, when the prevalence is very small, a relation between the accuracy and the localization of the PEV of the weighted adjacency matrices obtained in the linearization was proposed in chapter 5. This is justified by the fact that a leading contribution to the probability that a node i is infected in the mean-field theories is proportional to the corresponding PEV of A_{ij} or $B_{ij}(\lambda_c^{\mathrm{PQMF}})$ for QMF and PQMF theories, respectively. For sake of completeness of chapter 5, in which the accuracy at the epidemic threshold was discussed thoroughly, Figure 6.9 shows the steady-state density calculated slightly above the epidemic threshold of the PQMF theory versus the network size. The PQMF theory is much more accurate than the QMF but also starts to deviate from simulations as the network size increases. In both cases the accuracy is greater for networks with a less localized PEV as quantified by the IPR mathematically given in Eq. (4.39). The larger the IPR, the more localized the PEV. The insets in Figure 6.9 show the IPR for both QMF and PQMF theories where we

Network	η_{QMF}	η_{PQMF}	Y_4^{QMF}	Y_4^{PQMF}	κ	$\langle C\rangle$	$\langle l\rangle$	Q
AS Caida	0.0320	0.0022	0.0240	0.0139	280.2	0.333	3.88	0.670
Amazon 3/2/03	0.7500	0.2400	0.1060	0.0114	11.1	0.430	8.85	0.902
Astrophysics 93/03	0.0350	0.0130	0.0045	0.0043	65.7	0.669	4.19	0.622
CiteSeer	0.0670	0.0100	0.0177	0.0109	48.5	0.212	6.46	0.797
Jung	0.0079	0.0021	0.0478	0.0335	991.8	0.680	2.11	0.480
S838	0.2800	0.0700	0.1790	0.0340	4.02	0.058	6.86	0.791
Slashdot	0.0320	0.0011	0.1440	0.0347	81.6	0.047	4.52	0.480
URVmail	0.0220	0.0028	0.0096	0.0087	18.7	0.254	3.61	0.568
Cora	0.152	0.0370	0.0100	0.0090	23.7	0.307	5.85	0.790
Air Traffic	0.3800	0.0660	0.0191	0.0154	7.36	0.073	5.93	0.695
Gnutella 8/24/02	0.1700	0.0066	0.2140	0.0800	2.39	0.015	4.64	0.463
cElegans Neural	0.0570	0.0078	0.0189	0.0175	26.0	0.308	2.46	0.394

Table 6.1: Relative deviations and inverse participation ratios for QMF and PQMF theories applied to real networks and basic metrics, namely, heterogeneity coefficient κ, average clustering coefficient $\langle C\rangle$, average shortest distance $\langle l\rangle$ and modularity coefficient Q.

see that the latter is much less localized than the former but still increases towards a finite value as the network size increases, indicating localization asymptotically.

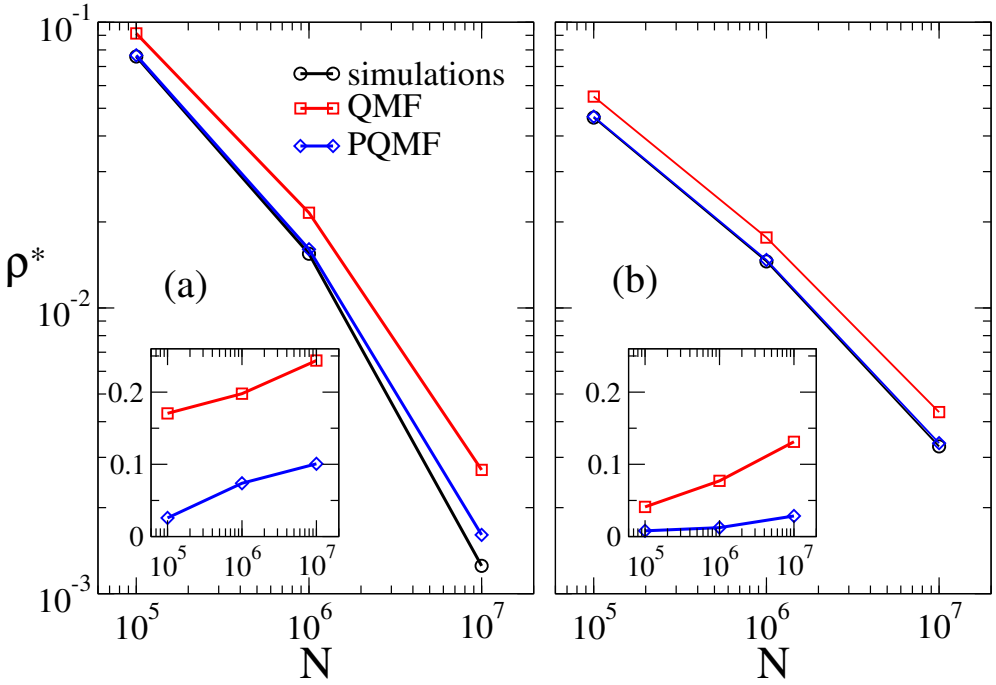


Figure 6.9: Finite size scaling of the steady-state density evaluated at $\lambda = 2\lambda_c^{\mathrm{PQMF}}$. WPCM networks with degree exponent $\gamma = 2.8$, presenting disassortative ($\alpha = -0.2$) and uncorrelated ($\alpha = 0.0$) degree correlations are considered. Insets show the corresponding IPR calculated for PEV of the corresponding mean-field theory.

However, the nonperturbative theory accounts for the contributions of the complete basis of eigenvectors, whether it is of A_{ij} or $B_{ij}(\lambda_c^{\mathrm{PQMF}})$, and this comparison is no longer justifiable. Indeed, as shown in Table 6.1, the accuracy of the QMF theory can be high even when the PEV is localized, as, for example, in the case of the Slashdot network. In other cases, such as the Air Traffic and Cora networks, the PEV localization corresponding to the QMF and PQMF theories are similar but the performance of the latter is much better. We performed a statistical analysis of the correlations between $\ln \eta$ and $\ln Y^{(1)}$ and found nonstatistically significant p values of $p_{\mathrm{QMF}} = 0.29$ and $p_{\mathrm{PQMF}} = 0.46$. It is noteworthy that the statistical analyses considering the linear data present even lower statistical significance.

We also checked (logarithm) statistical correlations of η with other basic network metrics, namely, the heterogeneity coefficient κ [Eq. (2.5)] , the modularity coefficient

Q [54], the average clustering coefficient $\langle C \rangle$ [Eq. (2.8)], and the average shortest distances $\langle l \rangle$ [Eq. (2.6)]. The modularity coefficient was determined using the Louvain method [102] available in network software such as Gephi [58]. This coefficient measures the level of homophily in communities, in which there are a higher density of connections within a community than among nodes belonging to different communities. The values of these quantities for this set of networks are shown in Table 6.1. We found statistical significance, at $p < 0.02$, only with κ and $\langle l \rangle$. Actually, κ and $\langle l \rangle$ are correlated since more heterogeneous networks tend to have shorter average distances due to the shortcuts introduced by hubs [1,11]. The correlation between η and $\langle l \rangle$ actually is not very surprising since one intuitively expects that the shorter the distances, the more accurate the mean-field hypothesis of neglecting long-range correlations becomes. One interesting feature is that the approximation given by Eq. (4.43) in the PQMF theory discards the possibility of triangles, in which the neighbors i and l of j are also connected. So, one could expect a worse performance in networks with a high clustering coefficient, but no statistical correlation with this metric was found ($|r| = 0.51$). In summary, we could not infer which structural properties control the accuracy of the mean-field approaches in regimes of high prevalence.

6.2 Summarizing this chapter

Reliable theoretical approximations are usually required to consider the heterogeneous structure of the contact networks and the dynamical correlations, in which the states of neighboring individuals are statistically correlated. These features are explicitly included in the PQMF theory described in secs. 4.2.3 and 5.2. However, this theory has been applied mainly to analyze the behavior of epidemic processes in the neighborhood of the transition from an endemic to a disease-free state through perturbative analyses where the epidemic prevalence is very low. In this work, we contribute to filling this gap by performing a detailed nonperturbative numerical analysis of the SIS model on synthetic and real networks within a wide range of heterogeneities and assortativities in the regime of high epidemic prevalence.

For synthetic networks generated with the Weber-Porto configuration model (see Sec. 3.1.1), we report that the PQMF theory predicts with great accuracy the regime of high prevalence observed in stochastic simulations in networks with power-law degree distributions for all values of degree exponents investigated ($\gamma = 2.3$, 2.8, and 3.5) and degree correlations (disassortative, neutral, and assortative). In the case of $\gamma > 5/2$, where hubs tend to become separated as the network size increases, we observed that the PQMF theory significantly outperforms the simpler QMF theory where heterogeneity is fully considered but dynamical correlations are neglected, the discrepancy between theories being larger for large γ. The high accuracy of the PQMF theory at high prevalence contrasts with its low performance at asymptotically low prevalence, where the theory is known to deviate from exactly known critical behavior [32], in which $\rho \sim \lambda^{\beta}$ with $\beta > 1$ while the mean-field exponent is $\beta = 1$ (see Sec. 5.2). We argue, however, that this mismatch is constrained to a region very close to $\lambda = \lambda_c \to 0^+$ such that the regime of not too low prevalence can still be accurately described by the PQMF theory.

In a set of real networks, where much more complex structures and correlations can be present, we observed that the PQMF theory always outperforms (sometimes very significantly) the QMF theory but may still present non negligible deviations from simulations in some cases (see Table 6.1). Differently from the low-prevalence regime, where the accuracy of mean-field theories is correlated with the spectral properties of Jacobian matrices, only trivial statistical correlations with simple network metrics could be identified and the problem of predicting when nonperturbative analysis is sufficiently accurate given certain network properties remains open.

Chapter 7

Dissecting localization phenomena of dynamical processes on networks

This chapter is devoted to study localization of epidemic processes on networks. The localization analyses are frequently grounded on eigenvectors of matrices which emerge in theoretical analysis of dynamic processes near to an active to inactive transition, as toughly discussed in chapters 4 and 5. We purpose an investigation of the localization of epidemic activity considering the actual dynamics on networks by an introduction of a normalized activity vector (NAV) that gauges the contribution of each node to the total activity. We analyze the NAV of SIS and contact process (CP) models due to their very distinct nature of activation [30] and critical exponents. The effects of the network structure in the localization of these models are studied considering different topologies such a star graph, random regular network (RRN) with an outlier, annealed and quenched random networks with power-law degree distributions. The application of the method is extended to mean-field theories and their accuracies in determining the localization patterns are investigated. Finally, we show that the order parameter near to the transition is ruled by contributions of lowly active nodes, which are by far the vast majority, while the IPR is determined by a vanishing fraction of the network. Our results highlight that localization patterns cannot be accurately resolved by simple metrics such as the IPR. We expect this method will be an important path for the investigation of localized activity in spreading and other processes on networks.

The results in this chapter were published in Ref. [103].

7.1 The normalized activity vector (NAV)

Localization of dynamical processes on networks can be quantified using a local order parameter that measures the level of activity in a determined subset. In recurrent epidemic processes such as the SIS, nodes are active or inactive with probabilities ρ_i and $1-\rho_i$, respectively. Recalling that the active nodes, which correspond to the infectious ones, define the epidemic prevalence in Eq. (4.30). The activity of a node can assume other forms depending on the dynamical system. For example, for the SIR model, the activity can be defined as the probability that a node will be infected in an outbreak started in a randomly selected node. For synchronization or opinion models the activity will depend explicitly on the neighborhood and should take into account higher order interaction such as pairwise.

We define the activity vector $\boldsymbol{\rho} = (\rho_1, ..., \rho_N)$. In analogy to the PEV, we introduce the NAV $\boldsymbol{\phi} = (\phi_i, ..., \phi_N) \propto \boldsymbol{\rho}$ whose components are given by

$$\phi_i = \frac{\rho_i}{\sqrt{\sum_{i=1}^{N} \rho_i^2}}. \tag{7.1}$$

Two limit cases are the completely localized and delocalized states, in which $\rho_i = N\delta_{ij}\bar{\rho}$ and $\rho_i = \bar{\rho}$, respectively. The corresponding components of the NAVs are $\phi_i = \delta_{ij}$ and $\phi_i = 1/\sqrt{N}$, respectively. The IPR is defined for ϕ_i as

$$Y_4(\boldsymbol{\phi}) = \sum_{i=1}^{N} \phi_i^4 \tag{7.2}$$

for which we derive the limit behaviors of $Y_4 = N^{-1}$ for fully delocalized and $Y_4 = 1$ for fully localized activity.

We analyzed localization in two recurrent activation models, the SIS and CP. Contact process is a variation of the SIS model, in which the infection rate per contact is λ/k_i, in which k_i is the node degree which is transmitting the infection while the

healing processes is exactly the same. A brief review of this dynamics is presented in Appendix B. Despite similarities between the models, their activation processes are of very different nature when evolving on random power-law networks [30]. In SIS the activation is triggered by hubs as discussed in Sec. 4.3 while in CP it is collective as in an usual phase transition [104]. In SIS on random networks with $P(k) \sim k^{-\gamma}$, hubs responsible for the activation can still be either in a densely connected core for $\gamma < 5/2$ or sparsely distributed for $\gamma > 5/2$ [18,30,79]. Thus, patterns of localization are eligible to be different for these models.

CP and SIS are dynamics with absorbing states. We overcome difficulties inherent to these systems using the weak external field method introduced in Sec. 4.4.1. In this method, self-activation (spontaneous creation) is introduced through a uniform and spontaneous rate f such that $f \to 0$ as $N \to \infty$ [94]. One advantage of this approach is that localization can be investigated also in the absorbing phase of the original dynamics. In this chapter we will not adopt $\mu = 1$ throughout the text since it helps dimensional analysis in the extensive analytical treatment developed here. We remark again that $\mu = 1$ could be used as well.

The implementation of stochastic simulations for the SIS model in Sec. 4.4.1 is slightly modified by the introduction of the external field. At each time step, with probability

$$P = \frac{\mu N_{\text{inf}}}{\lambda N_{\text{SI}} + \mu N_{\text{inf}} + fN} \tag{7.3}$$

one infected node is chosen at random and becomes susceptible. As in Sec.4.4.1, N_{inf} is the total number of infected nodes and N_{SI} is the total number of links emanating from them. With probability

$$Q = \frac{fN}{\lambda N_{\text{SI}} + \mu N_{\text{inf}} + fN} \tag{7.4}$$

a node of network is chosen with equal chance and becomes infected if it is susceptible. Finally, with probability $1 - Q - P$ an infected node i is selected with probability proportional to its degree k_i and one of its neighbors j is chosen with equal chance. If

j is susceptible it becomes infected. The time is incremented by

$$\delta t = \frac{-\ln u}{\lambda N_{\rm SI} + \mu N_{\rm inf} + fN} \tag{7.5}$$

in which u is a random number uniformly distributed in the interval $(0, 1)$. In CP simulations we replace $N_{\rm SI}$ by $N_{\rm inf}$ in Eqs. 7.3, (7.4), and (7.5) and chose with equal chance the infected node that transmits the infection while all other steps are identical.

We compute $\boldsymbol{\rho}$ in simulations as the fraction of time that each node i is active during an averaging time $t_{\rm av}$ after a relaxation time of $t_{\rm rlx}$. Mathematically it becomes,

$$\boldsymbol{\rho} = \frac{1}{t_{\rm av}} \int_{t_{\rm rlx}}^{t_{\rm rlx+av}} \boldsymbol{\sigma}(t) dt, \tag{7.6}$$

in which σ_i represents the state of a node i. Note that, in both models this variable assume the values $\sigma_i = 0$ when the node is inactive (susceptible) or $\sigma_i = 1$ when it is active (infected). Values $t_{\rm av} = 10^5 \mu^{-1}$ to $10^8 \mu^{-1}$ and $t_{\rm rlx} = 10^4 \mu^{-1}$ to $10^6 \mu^{-1}$ were considered, being the larger times used for lower activity regimes in which fluctuations are larger. When theoretical approaches are taken into account, we solve them using a fourth-order Runge-Kutta method and $\boldsymbol{\rho}$ is obtained after convergence of the steady-state. In all simulations we used $f = \mu/N$ meaning that at most one new infection is randomly introduced by unity of time.

7.2 Localization phenomena on simple networks

In order to obtain insights about the localization in terms of the NAV, we start our analyses with simple networks.

7.2.1 Star graph

A star graph consists of a central node $i = 0$ connected to $i = 1, ..., K$ neighbors of degree $k = 1$ (leaves). The adjacency matrix is $A_{0i} = A_{i0} = 1$ for $i > 0$ and $A_{ij} = 0$ otherwise. Let us start with the QMF theory for SIS on the star graph defining ρ_0 as the probability that the center is infected and $\rho_1 = \rho_2 = ..., \rho_N$ as the probability that

a leaf is infected. The QMF equations become

$$\frac{d\rho_0}{dt} = -\mu\rho_0 + \lambda K(1-\rho_0)\rho_1 + f(1-\rho_0) \tag{7.7}$$

$$\frac{d\rho_1}{dt} = -\mu\rho_1 + \lambda K(1-\rho_1)\rho_0 + f(1-\rho_1). \tag{7.8}$$

Without self-activation ($f = 0$) the epidemic threshold is $\lambda_c = \mu/\sqrt{K}$, as mentioned in Sec. 4.2.3. The steady-state solution for $f \ll \mu$ and $\lambda \ll \mu$ is given by

$$\rho_0 = \frac{\lambda K\rho_1 - f}{\lambda K\rho_1 - \mu} \tag{7.9}$$

and

$$2\rho_1 = \frac{\lambda}{\mu} - \frac{\mu}{\lambda K} + \frac{f}{\mu} + \sqrt{\left(\frac{\lambda}{\mu} - \frac{\mu}{\lambda K} + \frac{f}{\mu}\right)^2 + \frac{4f}{\lambda K}} \tag{7.10}$$

At $\lambda = \lambda_c = \sqrt{\mu/K}$, $\rho_0 = \sqrt{f/\lambda}$ and $\rho_1 = \sqrt{\lambda f}/\mu$ leading to an IPR, given by $Y_4 = \phi_0^4 + K\phi_1^4$, which asymptotically assumes $Y_4 = 1/4$ confirming localization. For supercritical region $\lambda^2 K \gg \mu$, $\rho_0 \approx 1$ and $\rho_1 = \lambda/\mu$ implying in $Y_4 = 1/K$ in a fully delocalized state. Finally, for subcritical region with $\lambda^2 K \ll \mu^2$ but still $\lambda K \gg \mu$

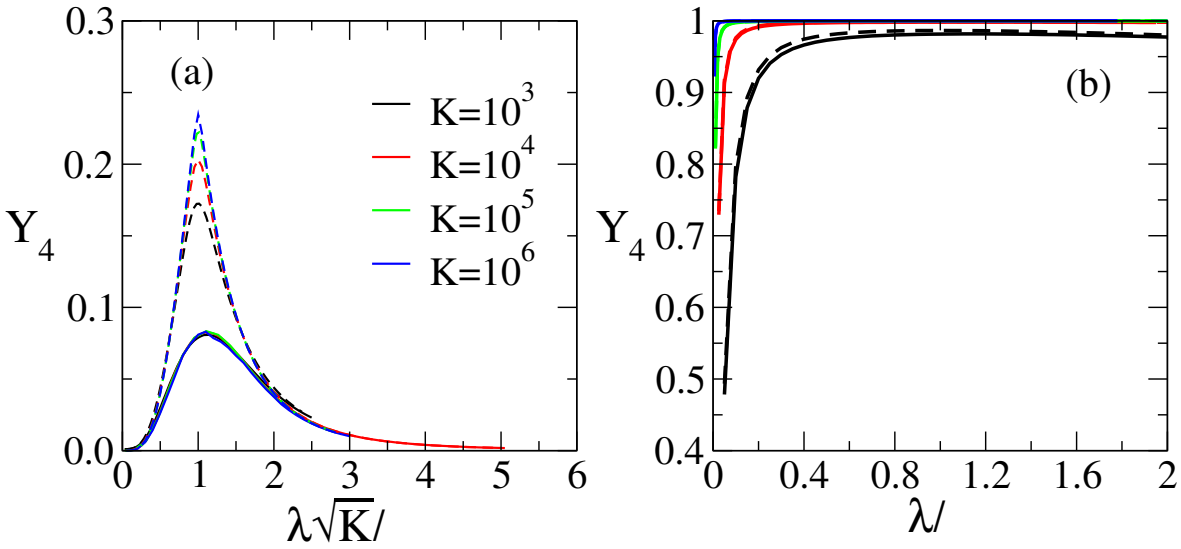


Figure 7.1: Inverse participation ratio analysis for (a) SIS and (b) CP models on star graphs with different number of leaves K. Dashed lines represent the QMF theory solution and solid lines the stochastic simulations. Dashed lines overlap with solid ones in (b) and can not be seen.

and $fK \sim O(1)$, $\rho_0 = f\lambda K/\mu^2$ and $\rho_1 = f/\mu$ leading to $Y_4 = (\lambda^2 K/\mu^2)^2 \ll 1$ in a subextensive localization since it goes to zero slower than K^{-1}.

Figure 7.1(a) compares the IPR as a function of $\lambda\sqrt{K}/\mu$ for the QMF theory of stochastic simulations of the SIS model on star graphs of different sizes K. The maximum localization in the QMF theory occurs at $\lambda\sqrt{K}/\mu = 1$ and the convergence to the asymptotic value $Y_4 = 1/4$ is verified. For simulations, the maximum value is $Y_4 \approx 0.08$ occurring at $\lambda\sqrt{K}/\mu \approx 1.1$ below the threshold of SIS on star graphs $\lambda\sqrt{K}/\mu \approx \sqrt{2}$ for pair QMF theory without self-activation (see Sec. 4.2.3). In both cases, the localization drops very quickly after the transition since the range of λ in which $\lambda\sqrt{K}/\mu$ is finite corresponds to $\lambda << 1$.

The CP dynamics on star graphs have a finite lifespan for $K \to \infty$ for any finite value of (μ, λ) [30] implying that the dynamics without a self-activation is always in the absorbing phase at long times. Developing a QMF theory for CP with self-activation, we have that $\rho_0 = \lambda C(\mu, \lambda, fN)$ and $\rho_1 = C(\mu, \lambda, fN)/N$ where

$$C(\mu, \lambda, \zeta) = \frac{\lambda^2 - \mu^2 + \lambda\zeta + \sqrt{(\lambda^2 - \mu^2 + \lambda\zeta)^2 + \lambda\zeta\mu^2}}{2\zeta\lambda} \tag{7.11}$$

is a finite positive constant when the arguments (μ, λ, ζ) are finite. Therefore, the IPR is asymptotically $Y_4 = 1$, confirmed in Figure 7.1(b) for both theory and simulations, showing that the CP dynamics with a self-activation on a star is fully localized in the center for any finite value of (μ, λ)

7.2.2 Random regular networks with one outlier

Random regular networks (RRN) are simple graphs in which all vertices have the same degree $k_i = m$ and connections are performed at random [26]. They are homogeneous networks where disorder is negligible. In this network, SIS and CP become equivalent by scaling the infection rate as $\lambda^{\text{SIS}} = m\lambda^{\text{CP}}$ and we consider only the former model. The transition for $f = 0$ happens at a finite threshold slightly above the pair approximation given by $\lambda_c^{\text{SIS}} = 1/(m-1)$ [81] and has a homogeneous mean-field like transition [85,99]. The SIS dynamics on a RRN is fully delocalized.

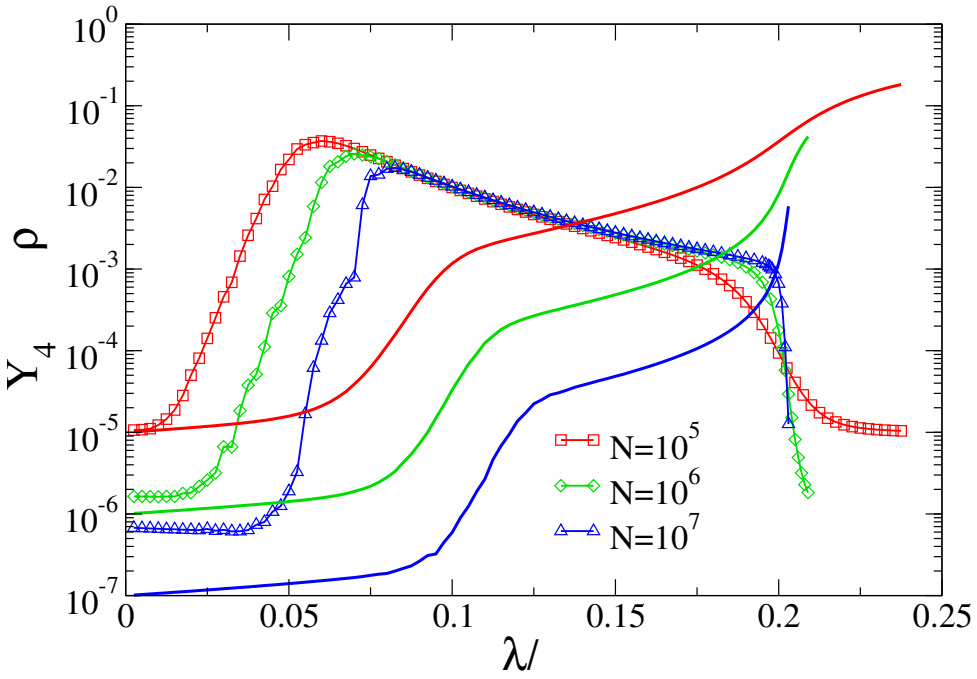


Figure 7.2: Inverse participation ratio Y_4 (lines with symbols) and prevalence ρ (lines) as function of the infection rate for $f = \mu/N$. Network size is indicated in the legends. Degree of vertices is $m = 6$ except one that has degree $K = 10^3$.

Localization can be introduced by the inclusion of a single node of degree $K \gg m$ to form a star subgraph immersed in the RRN [105]. The system can, in principle, present two activations being the first one of the star subgraph and other one of the remainder of the network at $\lambda/\mu \approx 1/(m-1)$. Figure 7.2 shows the epidemic prevalence and IPR as function of the infection rate for a RRN with degree $m = 6$ plus a hub of degree $K = 10^3$. We see two steep shoulders indicating the activations of the star subgraph and the RRN component. Remember that the transition is smoothed by the self-activation rate and would be singular only in the thermodynamic limit. Also observe the lower bound $\rho = 1/N$ before the star activation. The IPR increases rapidly, reaching a maximum slightly before the first shoulder and then decays slowly up to $\lambda = \lambda_c^{\mathrm{RRN}}$, after which drops suddenly to a value $Y_4 \sim 1/N$. Therefore, one can identify an extended localized phase in the range $\lambda_c^{\mathrm{star}} < \lambda < \lambda_c^{\mathrm{RRN}}$, ending with a sudden delocalization. A very similar system, a Bethe lattice with a hub, was investigated in the framework of QMF theory [44], in which it was shown that the PEV is localized while the eigenvector of the second LEV is delocalized.

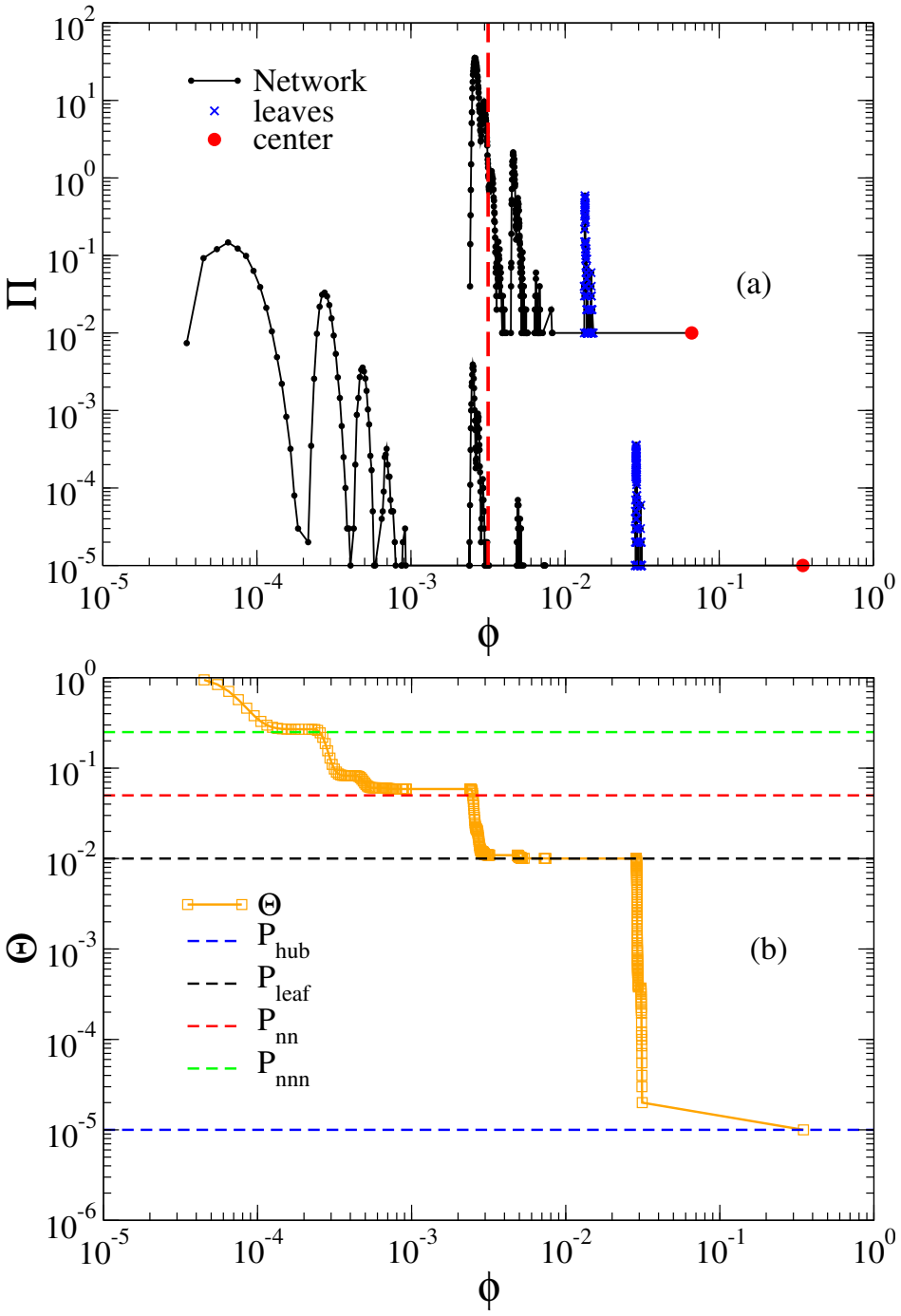


Figure 7.3: Analyses of the NAVCPD for SIS on a RRN of degree $m = 6$ plus a hub of degree $K = 10^3$. The network size is $N = 10^5$ and the binning parameter used to compute the distribution is $\delta\phi = 1/N$. (a) Probability distributions for $\lambda/\mu = 0.0875$ (bottom curve), which is near to the activation of the star subgraph, and for $\lambda/\mu = 0.2025$ (top curve), which near to the activation of the RRN component and multiplied by a factor 10^3 for sake of visibility, are shown. The components of the center and leaves of the star subgraph are also indicated. Dashed line represents $\phi = 1/\sqrt{N}$. (b) Complementary cumulative probability for $\lambda = 0.0875$. Dashed lines are probabilities that a randomly selected node belongs to the following subgraphs: hub, leaves, nearest and next-nearest neighbors of the leaves from the bottom to the top, respectively.

While insightful and providing the localization intensity, the IPR does not indicate where and how localization takes place. We can extract more information considering the NAV component probability distribution (NAVCPD) defined as the fraction of vertices $\mathbf{\Pi}(\phi)$ whose components are in the interval $\phi_i \in [\phi, \phi + \delta\phi]$. The NAVCPD for SIS dynamics near to the activation of the star subgraph are shown in Figure 7.3(a). The distribution shows high localization in the hub ($\phi \approx 0.35$) and leaves connected to it ($\phi \approx 0.029$). However, the distribution also presents other spaced peaks representing, in principle, intermediary localization in successive shells of neighbors centered on the hub. We tested this hypothesis calculating the complementary cumulative probability (tail distribution) $\mathbf{\Theta}(\phi)$ defined as the probability that a node selected at random has NAV component larger or equal than ϕ. Figure 7.3(b) shows $\mathbf{\Theta}(\phi)$ corresponding to the bottom distribution of Figure 7.3(a). We can clearly see the plateaus corresponding to the probability that a randomly select node is a leaf, $P_{\text{leaf}} = K/N$, and their nearest, $P_{\text{nn}} \approx (m-1)K/N$, and next-nearest, $P_{\text{nn}} \approx (m-1)^2 K/N$, neighbors. Observe that each plateau in $\mathbf{\Theta}(\phi)$ corresponds to a peak in $\mathbf{\Pi}(\phi)$. Interestingly, other localizations are also detected in the NAVCPD, as indicated by other peaks and plateaus not associated with the aforementioned subgraphs. Finally, near to the activation of the RRN subgraph, top curve of Figure 7.3(a), the NAV is concentrated around the delocalized value $\phi = 1/\sqrt{N}$, but the distribution still presents localization patterns especially in leaves and center.

7.3 Localization on synthetic random power-law networks

We now consider synthetic uncorrelated networks with power-law degree distributions. We consider both annealed and quenched networks. In the former, connections are rewired at random with a rate much faster than the typical rates of the dynamical processes evolving on them such that dynamical correlations are completely suppressed whereas in the latter edges are frozen and dynamical correlations are relevant [14].

7.3.1 Annealed networks

Differently of Sec. 4.2.1, in which we are interested in determining the epidemic threshold and the behavior of the prevalence near it, we use the HMF theory to study localization reckoning the NAV. The classical HMF theory becomes an exact prescription of annealed networks if fluctuations are negligible, as previously mentioned in Sec. 4.2.1. The HMF equation for the probability ρ_k that a node of degree k is infected including the external source is given by

$$\frac{d\rho_k}{dt} = -\mu\rho_k + \lambda k(1-\rho_k)\Omega_k + f(1-\rho_k), \tag{7.12}$$

where Ω_k is the probability that a randomly chosen neighbor is infected. For SIS we have $\Omega_k = \sum_{k'} P(k'|k)\rho_{k'}$ [33–35] and for CP $\Omega_k = \sum_{k'} P(k'|k)\rho_{k'}/k'$ [106] in which $P(k'|k)$ is the probability that a node of degree k is connected to node of k'. The steady-state solution is

$$\rho_k = \frac{\lambda k\Omega_k + f}{\mu + \lambda k\Omega_k + f}. \tag{7.13}$$

For uncorrelated networks we have $P(k'|k) = k'P(k')/\langle k\rangle$ and $\Omega_k = \Omega$ independent of the degree. The epidemic threshold for SIS with $f = 0$ is $\lambda_c/\mu = \langle k\rangle/\langle k^2\rangle$ and goes to zero only for $\gamma < 3$ (see Sec. 4.2.1) while for CP we have $\lambda_c = \mu$ (see Appendix B). At $\lambda = \lambda_c$ we have $(\lambda\Omega)/\mu \approx \rho/\langle k\rangle = \sum_k P(k)\rho_k/\langle k\rangle$ in both SIS and CP models. This result can be found handling Eq. (7.13) in terms of hypergeometric functions and can be found in [34] for SIS and [107] for CP, for example. The NAV in terms of ρ_k becomes

$$\phi_i = \frac{\rho_{k_i}}{\sqrt{N\langle k^2\rangle}}, \tag{7.14}$$

while the corresponding IPR is

$$Y_4 = \frac{\langle\rho_k^4\rangle}{N\langle\rho_k^2\rangle^2}. \tag{7.15}$$

The localization for $\lambda = \lambda_c$ is correlated with the node degree in annealed networks according to Eqs. (7.13) and (7.14): the higher the degree the more localized its activity it. According to the denominator of Eq. (7.13), the asymptotic behavior

depend on $\langle k_{\max}\rangle\rho$ in the regime $f \ll \rho$ which we are considering. If $k_{\max}\rho \ll 1$ the NAV becomes

$$\phi_i \simeq \frac{k_i}{\sqrt{N\langle k^2\rangle}}, \tag{7.16}$$

which leads to an IPR

$$Y_4 = \frac{\langle k^4\rangle}{N\langle k^2\rangle^2}. \tag{7.17}$$

Equations (7.16) and (7.17) correspond to the PVE of the adjacency matrix of an annealed network given by $A_{ij} = k_i k_j/(\langle k\rangle N)$ [74] presented, for example, in Refs. [44, 79]. For example, the critical prevalence of the CP in the QS state, where the dynamics returns to a previously visited configuration when the absorbing state is reached (see Sec. 4.4.1), goes to zero as $\rho_{\rm qs} \sim (gN)^{-1/2}$ [108] in which $g = \langle k^2\rangle/\langle k\rangle^2$. We numerically checked that the same result holds for CP and SIS on annealed networks using self-activation at the effective transition point given by the maximal of the dynamical susceptibility χ (see Appendix B).

For power-law degree distributions, we have $\langle k^n\rangle \sim k_{\max}^{n+1-\gamma}$ for $\gamma < n+1$ and $\langle k^n\rangle \sim$ const., otherwise, implying that $\rho k_{\max} \sim N^{(\gamma-1-\omega)/2\omega}$ for $\gamma < 3$ and $\rho k_{\max} \sim N^{(2-\omega)/2\omega}$ if $\gamma > 3$ in which we assumed a general scaling for the maximal degree $k_{\max} \sim N^{1/\omega}$ with $\omega > 1$. We analyze natural $\omega = \gamma - 1$ and structural $\omega = \max(2, \gamma-1)$ cutoffs [68,109]. As discussed in Sec 3.1.1, the first emerges naturally imposed by the degree distribution in finite-size networks, while the second allows to generate uncorrelated networks avoiding self- and multiple- connections for any $\gamma > 2$ [68,110].

Assuming $\rho \sim (gN)^{-1/2}$ for a structural cutoff, we obtain $\rho k_{\max} \ll 1$ for $\gamma > 2$ leading to

$$Y_4 \sim \begin{cases} N^{-(\omega+1-\gamma)/\omega} & 2 < \gamma < 3 \\ N^{-(\gamma+\omega-5)/\omega} & 3 < \gamma < 5 \\ N^{-1} & \gamma > 5 \end{cases}, \tag{7.18}$$

implying that IPR vanishes for all values of $\gamma > 2$. Strong corrections to the pure power-law scaling are present for γ near to 3 and 5 (It is shown in appendix B). Strict delocalization with $Y_4 \sim N^{-1}$ is found only for $\gamma > 5$ whereas subextensive localization with $Y_4 \sim N^{-\nu}$ and $\nu < 1$ [79] is found otherwise. Interestingly, SIS and

CP models have the same localization structure near to the transition whereas their critical behaviors for infinite networks with $N = \infty$ and $k_c = \infty$ present different exponents for scale-free networks [34,107]. For natural cutoff $\omega = \gamma - 1$ and $\gamma < 3$, again $\rho \sim (gN)^{-1/2}$, we have $\rho k_{\max} \sim \mathcal{O}(1)$ and

$$\langle \rho_k^n \rangle \simeq \int_{k_0}^{k_{\max}} \rho_k^n P(k) dk = F\left(n, \gamma - 1, \gamma; -\frac{\langle k \rangle}{\rho k_0}\right) - F\left(n, \gamma - 1, \gamma; -\frac{\langle k \rangle}{\rho k_{\max}}\right) \left(\frac{k_0}{k_{\max}}\right)^{\gamma - 1} \quad (7.19)$$

in which $F(a, b, c; x)$ is the Gauss hypergeometric function and $n = 2$ or 4. Taking the asymptotic limit with $\rho k_0 \ll 1$ and keeping only the leading contribution in the first term we obtain

$$\langle \rho_k^n \rangle \simeq \frac{\Gamma(n + 1 - \gamma)\Gamma(\gamma)}{\Gamma(n)} \left(\frac{\rho k_0}{\langle k \rangle}\right)^{\gamma - 1} - \left(\frac{k_0}{k_{\max}}\right)^{\gamma - 1} F\left(n, \gamma - 1, \gamma; -\frac{\langle k \rangle}{\rho k_{\max}}\right), \quad (7.20)$$

in which $\Gamma(x)$ is the Gamma function. Since $\rho k_{\max} \sim \mathcal{O}(1)$ so does $F(n, \gamma - 1, \gamma; -\frac{\langle k \rangle}{\rho k_{\max}})$ implying that both terms in Eq. (7.20) are the same order $\rho^{\gamma - 1}$, since $\rho \sim 1/k_{\max}$, and

$$\langle \rho_k^n \rangle \simeq c_n \rho^{\gamma - 1} \sim 1/N, \quad (7.21)$$

in which c_n is a constant that depends on both epidemic model and degree distribution. Again, using Eq. (7.15), we have that the IPR becomes finite indicating nonextensive localization in a finite number of nodes. Notice that while a subextensive fraction of the network is active since $\rho \sim N^{-1/(\gamma - 1)}$ and, consequently, an infinite number of vertices are simultaneously infected, the localization identified by the IPR is only in a finite number of nodes. This issue will be discussed further in the sequence of this chapter.

We performed simulations of both SIS and CP with self-infection on annealed networks with power-law degree distributions using natural and structural cutoffs. Figure 7.4 shows the IPR analyses for $\gamma = 2.3$. Other values of γ are shown in comparisons with quenched networks; see Sec. 7.3.2. The dependence of the IPR on infection rate is presented in Figure 7.4(a). The CP presents a faster drop to zero at $\lambda/\lambda_c \gtrsim 1$ in comparison with SIS model[1]. Notice that the abscissa is divided by λ_c that goes to 0

[1]The position of the susceptibility peak is slightly shifted from the mean-field prediction for SIS

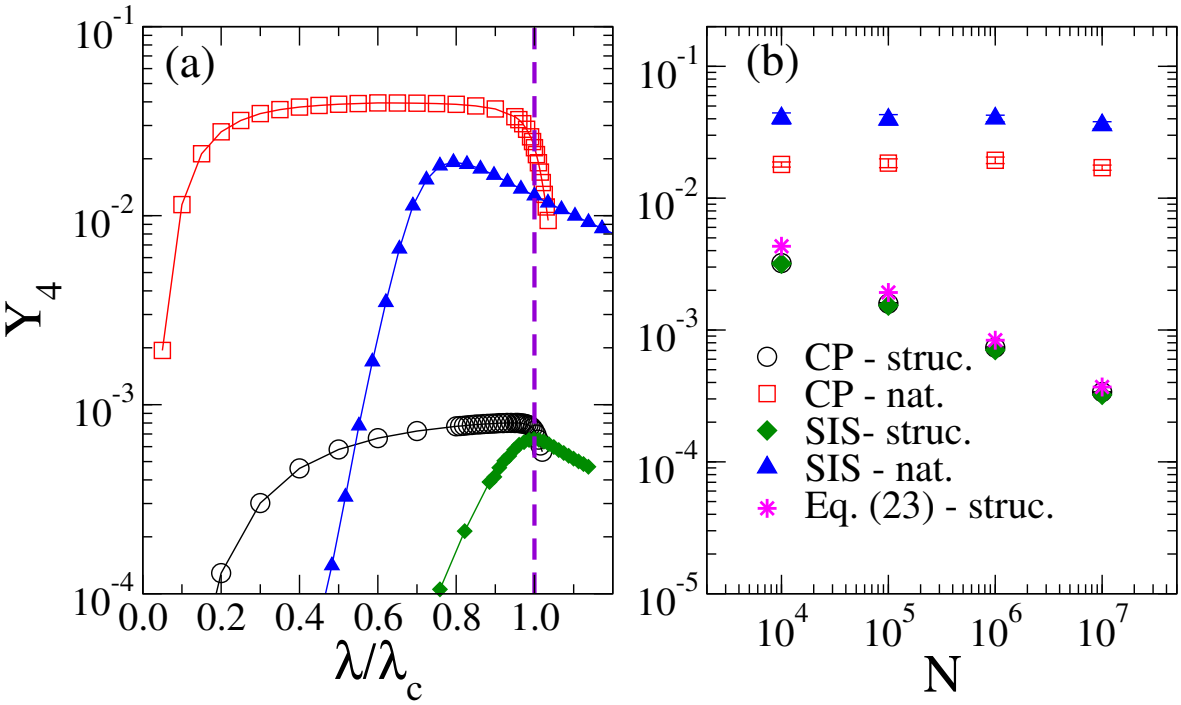


Figure 7.4: IPR analysis for SIS and CP dynamics on annealed networks with a power-law degree distribution of exponent $\gamma = 2.3$ and two types of degree cutoff: natural ($k_{max} \sim N^{1/(\gamma-1)}$) and structural ($k_{max} \sim \sqrt{N}$). IPR dependence (a) with infection rate for fixed size $N = 10^6$ and (b) with size for $\lambda = \lambda_c$ are shown. Dashed line represents $\lambda = \lambda_c$ in (a). IPR for structural cutoff computed with Eq. (7.17) is also shown in (b).

and μ for SIS and CP, respectively. At the transition, Figure 7.4(b), the predictions of the mean-field theory are confirmed. In the case of a structural cutoff, both SIS and CP present the same IPR value approaching zero according to Eq. (7.17). For $\gamma > 5/2$, the scaling given by Eq. (7.18) is confirmed but the prefactor is larger than the prediction of Eq. (7.17) as shown in Figure 7.7. For the natural cutoff, the IPR becomes finite confirming the localization predicted by the mean-field theory. Finally, the NAVCPD presents peaks at each node degree present in the network as will be shown in the comparison with the quenched case; see Sec. 7.3.2.

An interesting aspect calls our attention in this analysis. Activity in both CP and SIS dynamics in HMF or annealed approaches are localized at the transition according to the IPR analysis but an extensive, real endemic phase emerges above λ_c [34,107]. So, we present examples where localization in the critical point does not discredit a mean-field theory for describing an endemic phase transition. A natural, but still

model with the natural cutoff. For CP, this discrepancy does not happen.

unanswered question, is whether such a reasoning applies to the QMF theory as well.

A central issue for dynamic process with localization is to understand which subset of the network rules the epidemic prevalence and it was subject of intense studies [29,39,44,78,87,93,111] (See Secs. 4.2.2 and 4.3). A misleading interpretation is that the epidemic prevalence is given by the most localized region indicated by the IPR as we are going to disentangle from now on. In other words, we are going to show that a finite IPR, which is dominated by a few nodes, does not mean that the epidemic prevalence is also dominated by these few nodes. Let the fractional average of a quantity $F(\phi)$ be defined as

$$\zeta\{F(\phi)\} = \frac{\int_0^\phi F(\phi')\mathbf{\Pi}(\phi')d\phi'}{\langle F(\phi)\rangle}, \tag{7.22}$$

which yields the weight due to $\phi' < \phi$ in the average. For example, if $F = 1$ we obtain the cumulative probability distribution. Since $\phi_i \propto \rho_i$ by definition, we have $\zeta\{\rho^n\} = \zeta\{\phi^n\}$ and $\zeta\{Y_4\} = \zeta\{\phi^4\}$. The fractional average increases monotonically from 0 to 1 as ϕ varies from its minimal to maximal values. As an illustrative example, a fully delocalized state, in which $\mathbf{\Pi} = \delta(\phi - 1/\sqrt{N})$, has fractional averages given by a Heaviside function $\zeta = \mathbf{\Theta}(\phi - 1/\sqrt{N})$.

For HMF theory with a structural cutoff, the fractional average can be analytically computed using Eq. (7.16) to determine $\mathbf{\Pi}(\phi)$ in terms of $P(k)$ resulting

$$\zeta\{\rho\} = \frac{1 - (\phi/\phi_0)^{2-\gamma}}{1 - (\phi_{\max}/\phi_0)^{2-\gamma}} \tag{7.23}$$

and

$$\zeta\{Y_4\} = \frac{1 - (\phi/\phi_0)^{5-\gamma}}{1 - (\phi_{\max}/\phi_0)^{5-\gamma}}, \tag{7.24}$$

where ϕ_0 and $\phi_{\max}$ are given by Eq. (7.16) with $k_i = k_0$ and $k_i = k_{\max}$, the smallest and largest degree in the network respectively. Figure 7.5 presents the fractional averages computed for HMF theory with different degree exponents and $k_c = 2\sqrt{N}$ ($k_c = N$) for $\gamma < 3$ ($\gamma > 3$) using Eqs. (7.23) and (7.24). The fractional averages for the IPRs become appreciable for ϕ near to $\phi_{\max}$ when $\zeta\{Y_4\} \approx 1$, while for the prevalence it becomes

appreciable just slightly above ϕ_0 and is close to 1 when $\zeta\{Y_4\}$ is still departing from zero.

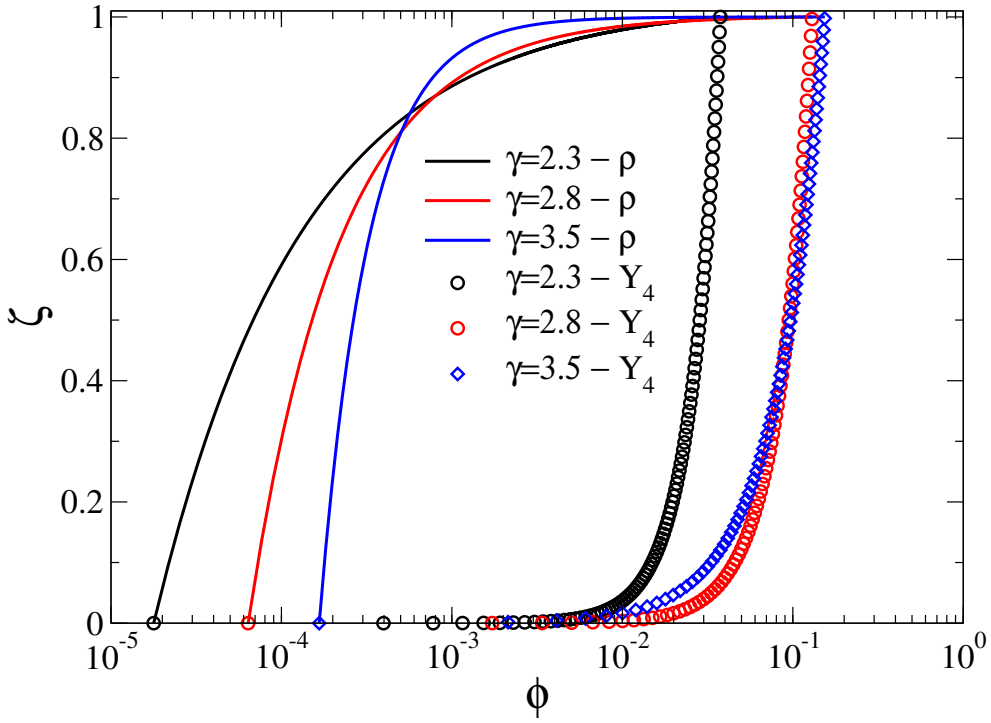


Figure 7.5: Fractional average analyses for prevalence ρ (lines) and IPR Y_4 (symbols) for SIS using Eqs.(7.23) and (7.24) for size $N = 10^7$, $k_c = 2\sqrt{N}$, and different degree exponents.

The gap between fractional averages of prevalence and IPR increases with size implying that the epidemic prevalence is given by the delocalized component of the network. To show this, consider the value ϕ^* for which the fractional average corresponds to a fraction $(1 - c)$ of total average, in which c is small but finite (typically $c = 0.1$), yielding an estimate for the range of the leading contributions to the averages. The corresponding ratio of ϕ^* for ρ and Y_4 for $\gamma < 5$ is

$$\frac{\phi^*\{\rho\}}{\phi^*\{Y_4\}} \simeq \frac{k_0}{k_{\max}} \frac{c^{-\frac{1}{(\gamma-2)}}}{(1-c)^{\frac{1}{5-\gamma}}} \ll 1. \tag{7.25}$$

This result should be understood as follows. While the IPR is dominated by a small, subextensive or nonextensive part of the system, relevant contributions for the order parameter come from the rest of the network. Moreover, as the network size increases the most active set contributes less for the overall activity and more for the IPR since

$k_{\max}$ diverges with N. The interpretation of localization near to a transition is made precise: While a small portion of the network can have extremely high activity that rules the IPR analysis, its contribution to the epidemic prevalence is negligible; the latter is given by the rest of the network that has lower level of activity but is much more numerous.

7.3.2 Quenched networks

The simpler analytical toolbox based on HMF theory used for annealed networks is not available for the quenched case. Therefore, we use stochastic simulation and numerical integration of the QMF equations [103] (See Secs. 4.4 and 6.1). For the latter, we include the self-activation term in Eq. (4.25) to obtain

$$\frac{d\rho_i}{dt} = -\mu\rho_i + \lambda(1-\rho_i)\sum_{j=1}^{N} A_{ij}\rho_j + f(1-\rho_i), \tag{7.26}$$

which was solved numerically using a fourth order Runge-Kutta method. The QMF equation for CP can be obtained from Eq. (7.26) replacing A_{ij} by A_{ij}/k_j, as discussed in Appendix B.

We used the uncorrelated configuration model (UCM) presented in Sec. 3.1.1, with a structural cutoff $k_c = 2\sqrt{N}$ to generate quenched networks. Figure 7.6 shows the NAVCPD for simulations of the SIS and CP dynamics at their corresponding transition points running on UCM networks with $\gamma = 2.3$. Both models present very similar values of the IPR, consistent with the decays for annealed networks given by Eq. (7.17); see inset of Figure 7.4(b). While the distribution for CP presents peaks corresponding to $\phi_i \propto k_i$ and matching almost exactly the annealed counterpart, the distribution for SIS matches almost perfectly the QMF theory, agreeing with the annealed case only in the tails. The QMF theory for CP at the transition also presents the peaks $\phi_i \propto k_i$. Therefore, the IPR alone fails in distinguishing differences in the localization of these dynamic processes. As discussed in Sec. 7.1 and in Appendix B, SIS and CP on quenched networks are characterized by different activation mechanisms [30] and differences in the localization are expected. The NAVCPD for critical SIS on quenched

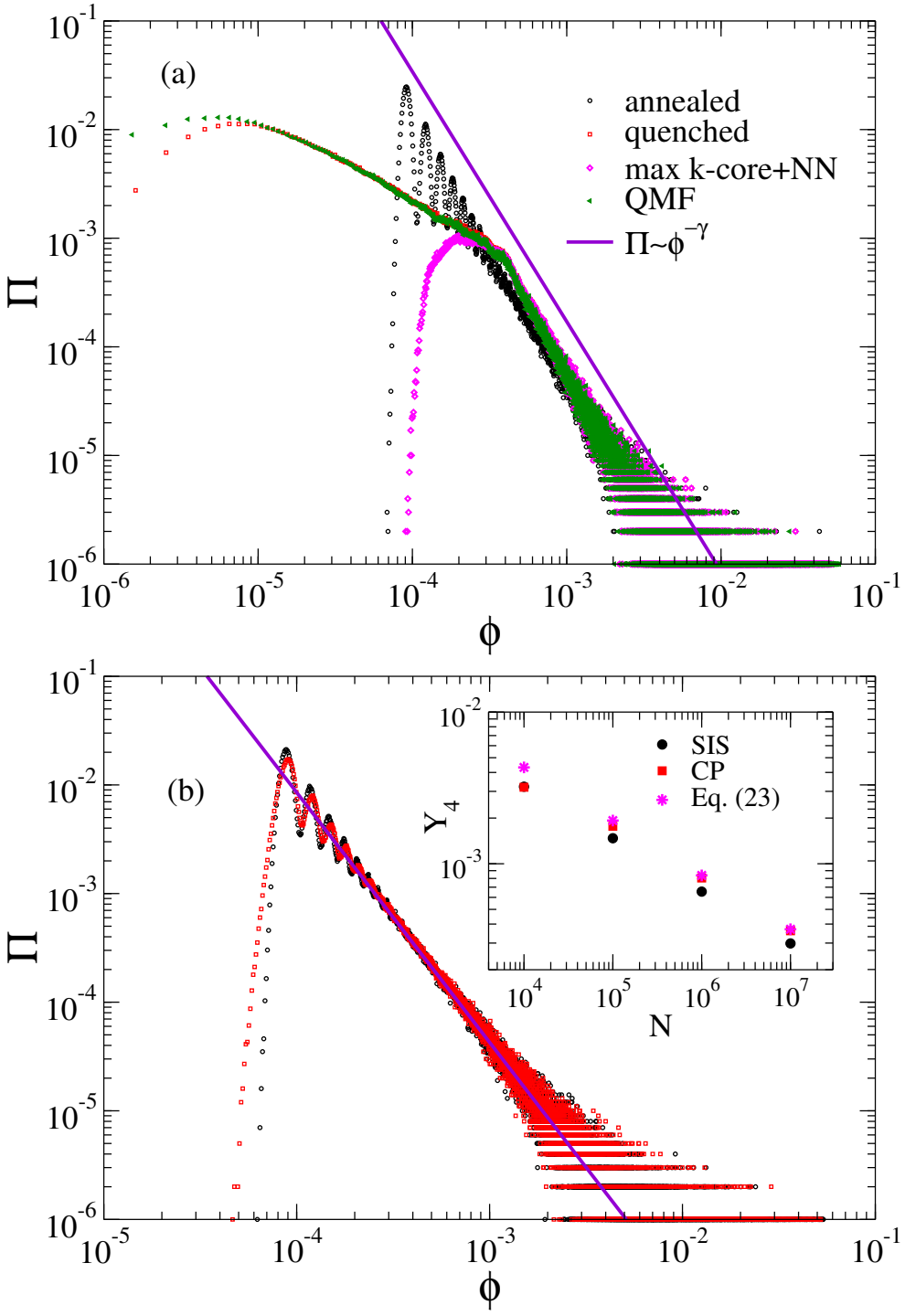


Figure 7.6: Analyses of the NAVCPD for simulations of (a) SIS and (b) CP models on a single sample of UCM or annealed networks of size $N = 10^6$ and degree exponent $\gamma = 2.3$. The curves were obtained at the epidemic threshold of the corresponding annealed ($\lambda_c^{CP} = 1$ and $\lambda_c^{SIS} = 0.0081$) and quenched ($\lambda_c^{CP} = 1.12$ and $\lambda_c^{SIS} = 0.0082$) networks, the latter for both QMF theory and simulations. The NAVCPD computed only on the subgraph containing the vertices of the maximum k-core plus their nearest-neighbors (NN) is also shown (magenta symbols). Inset: Finite-size scaling of the IPR for SIS and CP models on UCM networks at the transition point. Both cases consider an upper cutoff $k_c = 2\sqrt{N}$ for degree distribution.

networks with $\gamma = 2.3$ presents two asymptotic regimes: It is approximately $\mathbf{\Pi} \sim \phi^{-1}$ for low localization while the tail scales as does the degree distribution $\mathbf{\Pi} \sim \phi^{-2.3}$. We numerically determined that the crossover point between these regimes is very close to

$$\phi_* \approx \langle \phi \rangle = \frac{1}{\sqrt{N}} \left(1 - \frac{\langle \rho \rangle}{\sqrt{\langle \rho^2 \rangle}} \right) \sim N^{-1/2}, \tag{7.27}$$

in which the equality can be obtained from Eq. (7.1). This crossover is related to the nature of the activation of the SIS process for this range of γ, which is triggered in a densely connected component of the network identified by the maximum index of a k-core decomposition [18]. As presented in Sec 2.1, the k-core decomposition is the removal of vertices and edges connected to them such that only vertices of degree $k \geq q$ remain in the network, where q is the k-core index [67]. The maximum k-core corresponds to the last step before all vertices are removed. Figure 7.6(a) shows the NAVCPD computed for SIS only on a subgraph containing the vertices belonging to the maximum k-core plus their nearest neighbors. The distribution in this subgraph matches very well the overall distribution for $\phi > \phi_*$, dropping quickly for $\phi < \phi_*$, which corroborates our assertion. Note that the maximal k-core itself corresponds only to the narrow end of the tail. This subset corresponds to a vanishing fraction of the network as $N \to \infty$ [112] that dominates the contributions for IPR; see discussion on Figure 7.8.

Some important consequences come from Figure 7.6 and analyses of CP for other values of the degree exponents. A first one is the corroboration that CP on quenched networks is a HMF-like process [113,114], an issue intensely debated in the 2010s [106, 107,115–117], due to the almost perfect match of NAVCPD for simulations on quenched and annealed networks. For SIS, a remarkable agreement between NAVCPD for QMF theory and simulations occurs (a comparison between CP and the NAVCPD for QMF theory is shown in Appendix B) showing that the QMF is able to capture almost perfectly the localization pattern observed in simulations for the investigated range size (up to $N = 10^7$), even though the critical exponents of the prevalence are different with $\beta > \beta_{\text{QMF}} = 1$ [32,48] (We observed this result numerically in Sec. 6.1.1). Noti -

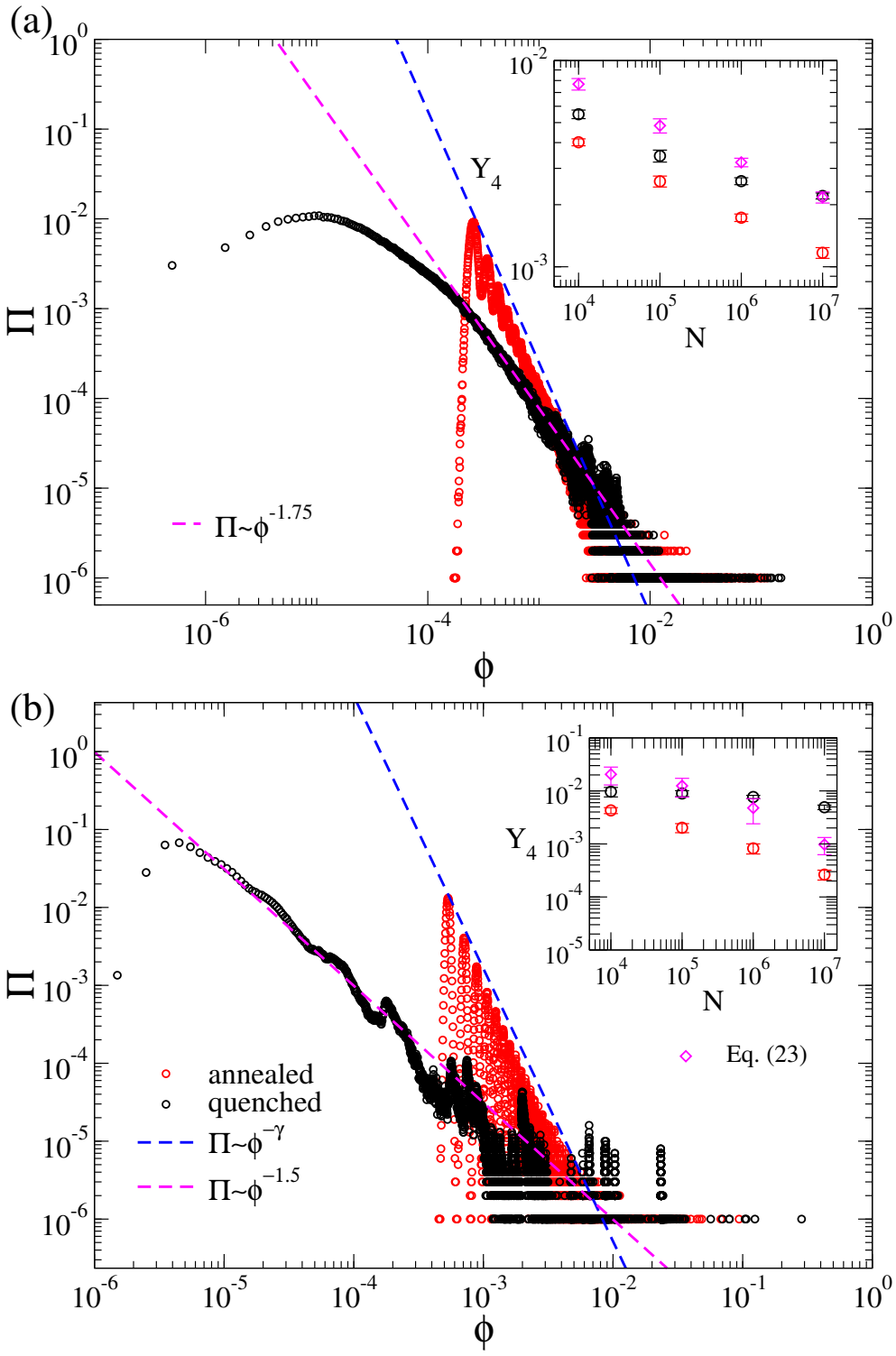


Figure 7.7: Analyses of the NAVCPD for SIS model on UCM and annealed networks (single realization of size $N = 10^6$) with degree exponent (a) $\gamma = 2.8$ and (b) 3.5 adopting $k_c = 2\sqrt{N}$ and $k_c = N$, respectively. In the latter, the chosen k_c leads to a highly fluctuating and $k_{max} \sim N^{1/(\gamma-1)}$. The inset shows the finite-size scaling for the IPR estimated at the maximal value of the dynamical susceptibility. The average was performed for 10 and 50 independent network realizations of quenched and annealed network, respectively. IPR computed using Eq. 7.17 is also shown.

ce that a similar situation happens for CP and SIS on annealed networks as discussed in Sec. 7.3.1. Obviously, one cannot discard the possibility of a different scenario for much larger, computationally inaccessible sizes. Finally, the crossover in the NAVCPD of SIS also indicates that the HMF behavior holds for an extended maximum k-core that encloses its nearest-neighbors.

For $\gamma > 5/2$ striking divergences between SIS dynamics on annealed and quenched networks emerge [18,26,32], the more evident for the larger γ. Figure 7.7(b) shows the NAVCPD analyses of the critical SIS for $\gamma = 3.5$ on both quenched and annealed networks with a same degree distribution. For annealed networks, the behavior is similar to the case $\gamma = 2.3$ with peaks proportionally related to the node degree. The corresponding IPR decays according to Eq. (7.17) as $Y_4 \sim N^{-0.4}$, but the prefactor is not accurate as it was for $\gamma = 2.3$; see Figure 7.6(b).

Simulations on annealed networks are less localized than the prediction of HMF theory due to the fluctuations absent in theory which contribute to spread the activity reducing, therefore, the localization. In the case of quenched UCM networks with $\gamma = 3.5$, the IPR seems to converge to a finite value but it has not achieved its asymptotic value in the investigated size range with $Y_4 \approx 0.01$ for $N = 10^7$. The NAVCPD for quenched networks differs substantially from their annealed counterparts. No correlation between NAV and degree distribution is evident. Also, it does not quantitatively match with QMF theory either, but it seems to capture the trend; see discussion on Figure 7.9. The NAVCPD for the quenched network presents very heavy tails, decaying approximately as $\mathbf{\Pi} \sim \phi^{-\eta}$ with $\eta \lesssim 2$ (the actual η value is beyond our scope), being therefore ruled by outliers that determine the IPR value. The case $\gamma = 2.8$ is similar to $\gamma = 3.5$ even for the network size adopted in Figure 7.7(a), in which effects of finite-size are expected (similar effects are observed for the PEV, as discussed in Sec. 5.3.2).

The fractional averages of prevalence and IPR for SIS at $\lambda = \lambda_c$ are presented in Figure 7.8 for single realizations of UCM networks with $N = 10^7$ vertices and three values of the degree exponent. As in the annealed case, we have that vertices of lowest ϕ values contribute significantly for prevalence while the IPR is determined by the most

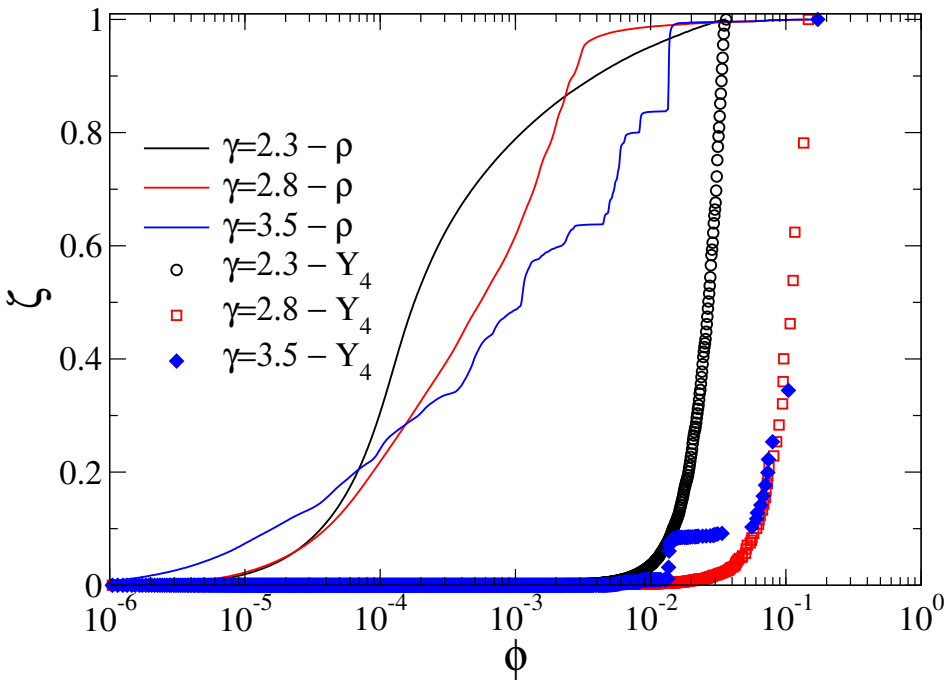


Figure 7.8: Fractional average analyses for prevalence ρ (lines) and IPR Y_4 (symbols) for SIS at the transition point with self-infection run on single realizations of UCM networks of size $N = 10^7$ and different degree exponents.

active ones; the latter contribute negligibly for the prevalence value. Curves for $\gamma = 2.3$ are smooth and qualitatively similar to the HMF case, with significant differences only for prevalence at very low ϕ. In turn, curves for larger values of γ contrast with the HMF case, being more evident for larger γ, in which the fractional IPRs vary abruptly for large ϕ. In Figure 7.8, for example, the node of highest ϕ represents 22% and 66% of the total IPR for networks with $\gamma = 2.8$ and 3.5, respectively. Curve $\zeta\{\rho\}$ for $\gamma = 3.5$ presents several jumps and plateaus. The same happens for the corresponding IPR, which can be seen in a (not shown) double-logarithm scale. This is an explicit observation of localized activity that is known to drive the epidemic activation for SIS at UCM networks with $\gamma > 3$ [29–31].

Important insights and elucidations are extracted from Figures 7.6, 7.7, and 7.8, helping to rationalize whether metastability of weakly interacting subgraphs leads or not to an actual endemic phase [29,39,44,78,87,93,118]. The SIS dynamics on quenched networks at the epidemic threshold is indeed localized for all values of γ (subextensively for $\gamma < 5/2$ and on a finite set for $\gamma > 5/2$) as does the PVE of the adjacency matrix

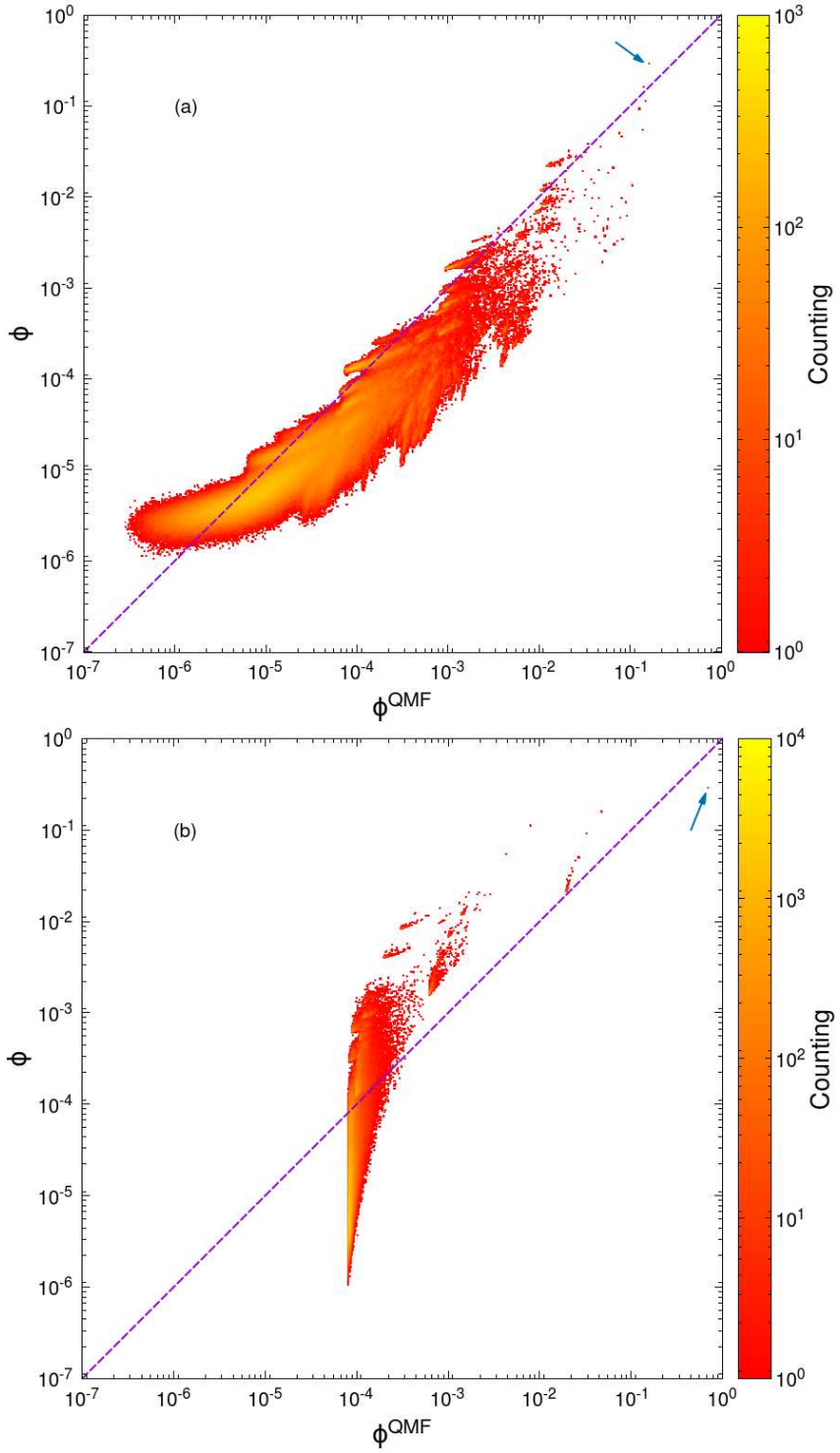


Figure 7.9: Binned scatter plots (grid with 500×500 bins in a logarithm scale) for NAV components at the transition point obtained in stochastic simulation (ordinate) and QMF theory (abscissa) on UCM networks ($k_c = N$) of degree exponent 3.5. Color bars represent the amount of points within a bin. Simulations for $\lambda = \lambda_c = 0.0775$, given by the dynamical susceptibility, are compared with (a) integration of the QMF equations for this same value of λ and (b) for critical QMF theory at $\lambda = \lambda_c^{\mathrm{QMF}} = 1/\Lambda_1 = 0.0225$. Networks size is $N = 10^6$. Dashed line is the identity $\boldsymbol{\phi} = \boldsymbol{\phi}^{\mathrm{QMF}}$.

[79], in the sense that a vanishing minority of vertices presents extremely high epidemic activity, ruling the IPR analysis. However, the corresponding epidemic process can still be undergoing a phase transition, in which extensive components are participating in the processes.

Figure 7.9(a) presents log-binned scatter plots for NAV components obtained in SIS simulations ($\boldsymbol{\phi}$) and integration of the QMF equations ($\boldsymbol{\phi}^{\mathrm{QMF}}$) for an infection rate $\lambda = \lambda_c$ determined by the maximum of χ. In contrast with the case $\gamma = 2.3$, for which an almost perfect match is observed (see Figure 7.6), the scatter plots for $\gamma = 3.5$ present a very disperse distribution of points. The integration captures the trend of the simulation presenting, however, most vertices with NAV components higher than in stochastic simulations, which is indicated by most points below the identity line. This is due to the QMF equation being actually supercritical since $\lambda_c > \lambda_c^{\mathrm{QMF}}$. In Figure 7.9(b), we compare the situation in which both QMF theory and stochastic simulations are in their corresponding epidemic thresholds ($\lambda = \lambda_c$ for simulations and $\lambda = \lambda_c^{\mathrm{QMF}} = 1/\Lambda^{(1)}$ for QMF theory). Here, the converse of Figure 7.9(a) is observed: Activity in QMF theory is slightly correlated with stochastic simulations only for the few and most active vertices whereas the vast majority of the network presents no correlation.

In order to illustrate the distinct levels of localization in these approaches, consider the single node contributions to the IPR in the network sample shown in Figure 7.9. The node indicated by arrows is the one of higher activity in both simulations and QMF theory. However, the gap between the first, ϕ_1, and second, ϕ_2, most active vertices is much larger in critical QMF than in simulations, given by $\phi_1/\phi_2 \approx 15$ and 1.8, respectively. In both cases, the IPR is led by the most localized node but much more in the QMF theory. Indeed, this single node contributes with $(\phi_1^{\mathrm{QMF}})^4 = 0.2380$ to the IPR $Y_4(\boldsymbol{\phi}^{\mathrm{QMF}}) = 0.2384$ for critical QMF theory, and with $\phi_1^4 = 0.0074$ of $Y_4(\boldsymbol{\phi}) = 0.0087$ for stochastic simulations. For supercritical QMF analyzed in Figure 7.9(a), we have $(\phi_1^{\mathrm{QMF}})^4 = 0.0007$ of $Y_4(\boldsymbol{\phi}^{\mathrm{QMF}}) = 0.0029$. Note that the IPR of critical QMF is very close to the value $Y_4 = 1/4$ of a star graph; see Section 7.2.1.

From the perspective of theory accuracy, these results show that QMF the-

ory should be used with extreme caution to determine the most active vertices in the network since it may deviate substantially from the actual picture as shown in Figure 7.9(a). An alternative is to use pair QMF theory which reckons dynamical correlations considering pairwise equations as discussed in Secs. 4.2.3 and 5.2. Indeed, a pair QMF theory was recently used to identify the most central spreaders that should be immunized to efficiently stop a SIS dynamics [82]. However, as shown in chapter 5, even pair QMF theory has limitations for SIS dynamics on network with large γ and should also be used with caution.

7.4 Summarizing this chapter

In this chapter, we develop a localization analysis applicable to any type of dynamical process, in which an activity can be gauged in terms of a local order parameter that was considered at a node level but lower resolution motifs are eligible as well. The method is generic and can be applied to both mean-field and stochastic simulations whether the system is near to a transition or not.

We applied the methodology to the SIS and CP, two similar conceptually dynamical processes with very distinct activation mechanisms. Handling their annealed versions on networks with power-law degree distributions, which can be done analytically due to the intrinsic mean-field behavior for this class of networks [14], we show that SIS and CP present exactly the same localization patterns near to the epidemic threshold ăfor any values of the degree exponent γ, despite these models being characterized by different critical exponents for the epidemic prevalence when $\gamma < 4$ [12,106]. Also, we show that both CP and SIS dynamics at their respective thresholds are localized in a subextensive subset (increasing sublinearly with network size) for $\gamma < 5$ and concomitantly give rise to an endemic (delocalized) phase immediately above the threshold.

We also analyzed localization of both dynamics on quenched networks. Firstly, we corroborate the mean-field nature of the CP on networks, an issue widely debated formerly [106,107,115–117] in terms of mean-field exponents, showing that annealed

and quenched versions of a same degree distribution have exactly the same localization patterns depending only on the node degree. We stress again that the localization pattern alone cannot give the ultimate response and only complements other evidences reported elsewhere [113,114].

The ăresults for SIS on quenched networks are richer and more complex. For $\gamma < 5/2$, the localization patterns at the threshold ămatch almost perfectly the QMF theory,ăagreeing with the annealed case only for ăa subset essentially composed by the ăvertices of the maximum k-core plus its nearest neighbors, still a subextensive set. Again, one knows that critical exponent of the epidemic prevalence of the QMF theory ă$\beta_{\mathrm{QMF}} = 1$ is not the correct one since $\beta > 1$ for any value of $\gamma > 2$ [32]. Despite this, ăour numerical results suggest that ăQMF is still able to capture almost exactly the localization structure. For $\gamma > 5/2$, the distribution of activity localization is more complex presenting outliers and heavy tailed distributions that lead to epidemic localization in a finite subset as predicted by QMF theory without presenting an accurate match with simulations. Indeed, comparing critical behavior of both simulations and QMF theory, correlation is found only for a tiny part of the nodes in which activity is highly localized.

Defining the concept of fractional averages we show that even in localized states the order parameter is ruled by the vast majority of vertices in which the activity is smaller in contrast with the IPR which is ruled by a small fraction (subextensive) of the network. A few nodes are active most of time while others are active only eventually due to stochastic interactions with the former. However, the latter being infinitely larger in number will lead any intensive measure as the order parameter. These finds are foundational for the correct interpretation of activation mechanism of an endemic phase.

Our results show that localization patterns are much more complex and revealing than simple metrics as the IPR can tell. Using spectral properties of, for example, adjacency matrix derived within mean-field framework can hide important features of the dynamical process and also lead to misleading ăor incomplete conclusions on the physical mechanisms involved in.

Chapter 8

Concluding remarks, forthcoming research and perspectives

The protein-protein [2] and face-to-face contacts [6], the interaction among members of the social microblogging Twitter [7–10] are all examples of interacting systems that can be modeled using complex networks [1,11,54]. A large number of networks representing real systems shows a heavy-tailed degree distribution [1,11,20], usually with strong levels of correlations [1,11–13]. One example is the Internet, in which a disassortative mixing is present [12]. Dynamical processes on complex networks have been studied for the absence of degree correlation which allows simplifications in the theoretical approaches. Reliable theoretical approximations are usually required to consider the heterogeneous structure of the contact networks and the dynamical correlations, which can be done considering mean-field theories [14].

In this thesis, a broad study of the epidemic model SIS on networks with power-law degree distribution and degree correlations was performed. We verified that in the limit of size $N \to \infty$, the threshold vanishes regardless of the correlation pattern and heterogeneity. Compared to the uncorrelated case, assortative networks have a smaller threshold while it is larger for disassortative mixing. The accuracy of QMF and PQMF theories in describing the SIS threshold is influenced by the degree correlations with PQMF outperforming the QMF theory.

The PQMF theory can be described by spectral properties of the Jacobian

matrix associated to this theory, as in the QMF case. We observed that the accuracy of the theories in predicting the epidemic threshold is positively correlated with the amplitude of the spectral gap and is inversely related to the level of localization of the PVE of the matrices associated with each theory. This general observation was corroborated by an analysis of a large set of real correlated networks, covering a wide range of sizes and topological features.

The steady-state prevalence in stochastic simulations as a function of the infection rate λ for $\gamma < 5/2$ presents a linear regime predicted by the QMF theory when λ is not very close to the epidemic threshold, whereas the exponent $\beta = 1/(3-\gamma)$ [32] is observed asymptotically close to the transition, where fluctuations are more important. However, the latter behavior is squeezed around the region $\lambda = \lambda_c \to 0$ as the network size increases.

A nonperturbative approach for the regime of high epidemic incidence is possible through numerical integration of both QMF and PQMF dynamical equations. We report that PQMF theory predicts very accurately the regime of high prevalence observed in stochastic simulations. In the case of $\gamma > 5/2$, where hubs tend to be separated apart as network size increases, we observed that the PQMF theory significantly outperforms the simpler QMF theory, being the discrepancy between theories larger for large γ.

In a set of real networks, where much more complex structures and correlation patterns can be present, we observed that PQMF always outperforms (sometimes very significantly) the QMF theory, but may still present non-negligible deviations from simulations in some cases. Differently from the low prevalence regime where the accuracy of mean-field theories is correlated with spectral properties of Jacobian matrices related to these theories, only trivial statistical correlations with simple network metrics could be identified.

As a perspective, we expect to extend our nonperturbative analyses to other theoretical approaches. One of them is the recurrent dynamic message-passing (rDMP) [49] where the mutual reinfection between pairs of connected nodes (backtracking) is disregarded. In this approach, the epidemic threshold and the behavior

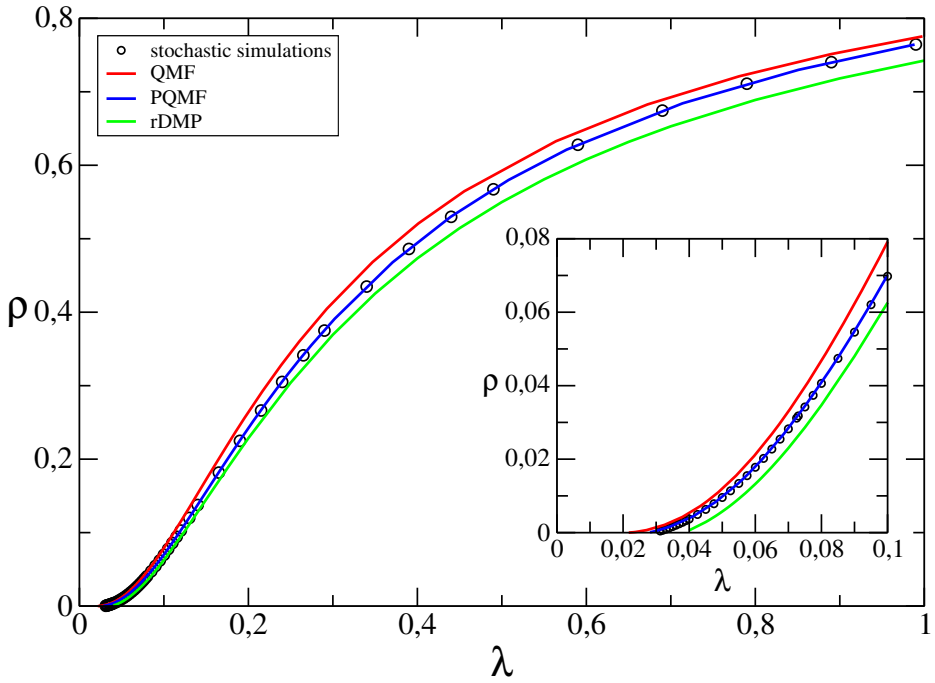


Figure 8.1: Comparison among QMF, PQMF and rDMP approaches in predicting the regime of high epidemic prevalence in a uncorrelated network with power-law degree distribution network. Symbols represent stochastic simulation while red, blue and green solid lines are the numerical integration of the QMF, PQMF and rDMP approaches, respectively. We adopt $\gamma = 2.8$ and $N = 10^6$ nodes.

near to it is described by the $2N \times 2N$ matrix named as nonbacktracking or Hashimoto matrix [49,90,119], given by

$$\mathbf{B} = \begin{pmatrix} \mathbf{A} & \mathbf{I} - \mathbf{D} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}, \tag{8.1}$$

in which $\mathbf{A}$ is the adjacency, $\mathbf{I}$ is the identity, $\mathbf{D}$ is a diagonal $D_{ij} = k_i\delta_{ij}$ and $\mathbf{0}$ is the null matrices, respectively. The successive reinfection process is essential in the activation mechanism of the SIS model as discussed in Sec. 4.3. On networks with uncorrelated power-law degree distribution, the influence of the backtracking in the SIS model will be appreciable for $\gamma > 5/2$, in which the backtracking mechanism triggers the epidemic process. It is indicated by the non-null epidemic threshold predicted by the rDMP theory for $\gamma > 3$ [90]. Considering a network with $\gamma = 2.8$ and $N = 10^6$, as shown in Figure 8.1, our preliminary results extending the analysis of the SIS model for the regime of high epidemic prevalence corroborate with the relevance of back-

tracking for this dynamics. The expected epidemic threshold and the prevalence are shown in Figure 8.1 . The rDMP underestimates the prevalence for all infection rates investigated. The PQMF theory appears as the most accurate theoretical approach in predicting the stationary prevalence in odds with the claims of Ref. [49].

We expect stronger effects for larger network sizes, $\gamma > 3$, and in the presence of disassortative degree correlation since the mechanism of activation described in Sec. 4.3 is more relevant. In addition, we are interested in investigating other recurrent dynamics such as the susceptible-infected-recovered-susceptible (SIRS) and the susceptible-exposed-infected-recovered-susceptible (SEIRS) where the effects of backtracking can be minimized by the recovered state since a reinfection is temporarily unavailable.

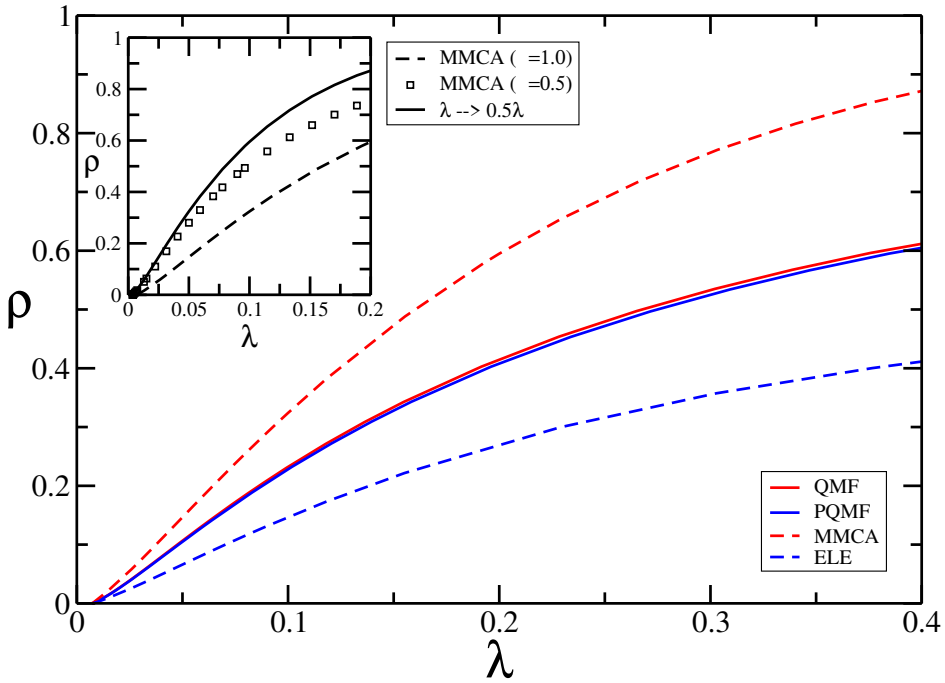


Figure 8.2: Comparison between discrete- and continuous-time approaches. The former are represented by the MMCA (red dashed line) and ELE (blue dashed line) while the latter by QMF (red line) and PQMF (blue line) theories adopting $\mu = 1$. Inset: Dependence of the discrete theories with the choice of the healing rate. The transformation $\lambda \to \lambda\mu$ (line) does not map the curves corresponding to MMCA theory with $\mu = 1.0$ (dashed line) onto that with $\mu = 0.5$ (symbols). These analyses were performed on uncorrelated power-law degree distribution network with $\gamma = 2.3$ and $N = 10^6$.

Since there are large discrepancies between discrete- and continuous-time implementation of the SIS dynamics [98], another perspective is an extension of the non-

perturbative analysis comparing the theoretical discrete-time approaches represented by the microscopic Markov chain approach (MMCA) [21] and the epidemic linking equations (ELE) [82] with their continuous-time equivalents, QMF and PQMF theories respectively. Some differences between these theoretical approaches are shown in Figure 8.2. In the regime of high epidemic incidence, the MMCA theory overestimates the prevalence of the QMF while ELE underestimates it in the PQMF theory. The effects of dynamical correlations are stronger in discrete-time approaches. As shown in the inset of Figure 8.2, the prevalence as a function of the control parameter to a discrete-time approach for $\mu = 1$ can not be mapped in other curves corresponding to other values of μ under the transformation $\lambda \to \mu\lambda$ differently of the continuous case. Effects of the choice for the infection and healing parameter are observed in discrete-time implementation of the dynamics [98]. Determining the regimes of equivalence between continuous- and discrete-time theoretical approaches and a spectral analysis of the former are some points to be considered in near future.

Finally, we develop a method to analyze localization applicable to any type of dynamical process where activity can be gauged in terms of a local order parameter that was considered at a node level but lower resolution motifs are eligible as well. Applying this methodology to the SIS and CP, two similar conceptually dynamical processes with very distinct activation mechanisms and different critical exponents for the epidemic prevalence when $\gamma < 4$, the same localization patterns near to the epidemic threshold on annealed networks.

We also analyzed localization of SIS and CP dynamics on quenched networks. We corroborate the mean-field nature of the CP on networks showing that annealed and quenched versions of a same degree distribution have exactly the same localization patterns depending only on the node degree. The results for SIS on quenched networks are richer and more complex. For small $\gamma < 5/2$, the localization patterns at the threshold match almost perfectly the QMF theory, agreeing with the annealed case only for a subset essentially composed by the nodes of the maximum k-core plus its nearest neighbors, still a subextensive set. For $\gamma > 5/2$, the distribution of activity localization is more complex presenting outliers and heavy tailed distributions that

lead to epidemic localization in a finite subset as predicted by the QMF theory without presenting an accurate match with simulations.

We observed that even in localized states the order parameter is ruled by the vast majority of nodes where the activity is smaller in contrast with the IPR which is ruled by a fewer (subextensive), small fraction of the network. Our results show that the localization patterns are much more complex and revealing than simple metrics such as the IPR can tell. Using spectral properties of matrices derived within mean-field frameworks can hide important features of the dynamical process and also lead to misleading or incomplete conclusions on the physical mechanisms involved in.

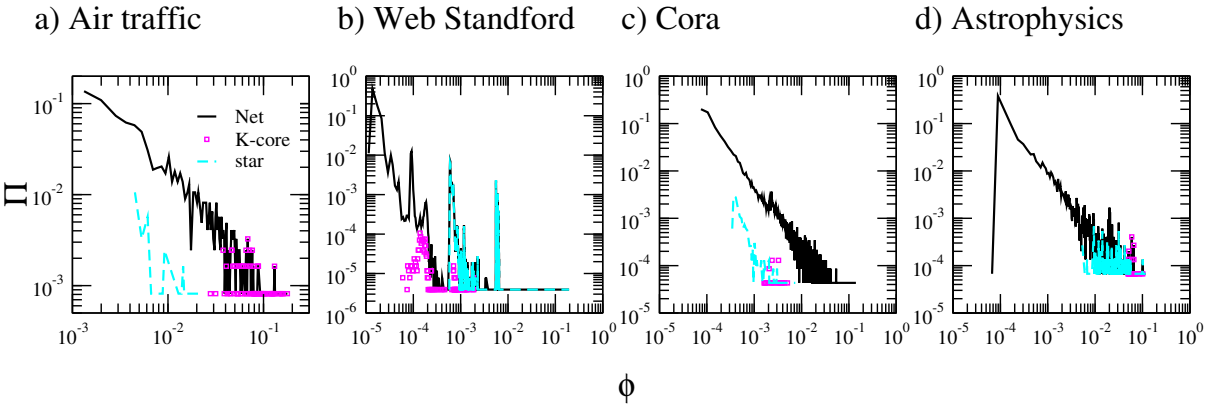


Figure 8.3: Analysis of NAVPD for the SIS model on real network (black line), maximum k-core (symbols) and star graph formed by the largest hub and its nearest-neighbors (dashed line): a) Air Traffic, b) Web Stanford, c) Cora and d) Astrophysics networks.

In real networks, the localization analysis exploring the concept of IPR is more difficult since one can not determine its behavior as a function of the network size. As a perspective, we intent to extend the methodology developed for synthetic networks to the real ones. Considering the SIS dynamics, preliminary results comparing the NAVPD at the epidemic threshold determined by the maximum of the susceptibility are shown for Air Traffic, Web Stanford, Cora and Astrophysics networks in Figure 8.3. The NAVPD presents a power-law behavior for all networks indicating strong localization effects. In a comparison among the NAVPD for the network, maximum k-core, and the largest hub plus its nearest-neighbors (star), nodes presenting higher activity can be localized in the maximum k-core (Air Traffic), in the star (Web Stanford), in both of them when the largest hub is inside of the maximum k-core (astrophysics) and

in none of them (cora). Similar results were observed in an analysis of the weights of these structures in the PEV of the adjacency matrix [78] for these networks. Other set of nodes determined based on other centralities can be considered. In addition, we expected to study localization in other network topologies such as multilayer [120] and temporal [121] networks and other dynamical processes such as opinion dynamics [122], synchronization [123] and epidemic spreading with biased mobility, in which localization can be strong [52].

Appendix A

Real networks

We consider 99 real networks with diverse structural properties, based on the lists of Refs. [63,124]. Here we investigate their giant connected components, after symmetrizing all edges (weighted and/or directed) and avoiding multiple and self connections. The list of networks with some topological properties and epidemic (SIS) parameters is shown in Table A.1. For detailed information about the original references for all the networks, please check Refs. [63,124].

Network	N	$\langle k \rangle$	P	$\mathrm{IPR_A}$	$\mathrm{IPR_B}$	$\Delta\Lambda_{\mathrm{A}}^{1,2}$	λ_c	λ_c^{QMF}	$\lambda_c^{\mathrm{PQMF}}$
Karate club	34	4.59	−0.476	0.0730	0.0649	1.75	0.235	0.149	0.181
Radoslaw Email	167	38.9	−0.295	0.0133	0.0132	45.2	0.0191	0.0165	0.0168
Spanish B	12,643	8.70	−0.290	0.0246	0.0174	47.5	0.0105	0.00897	0.00951
Spanish A	11,558	7.45	−0.282	0.0190	0.0150	57.5	0.0113	0.00985	0.0103
US Air Transportation	500	11.9	−0.268	0.0176	0.0173	29.3	0.0251	0.0208	0.0214
Little Rock Lake	183	26.6	−0.266	0.0148	0.0145	14.6	0.0291	0.0242	0.0249
Japanese	2,698	5.93	−0.259	0.0296	0.0214	21.7	0.0281	0.0233	0.0250
English	7,377	12.0	−0.237	0.0120	0.0103	65.3	0.0101	0.00914	0.0094
French	8,308	5.74	−0.233	0.0351	0.0200	26.3	0.0197	0.0165	0.0179
Jung	6,120	16.4	−0.233	0.0478	0.0335	46.9	0.00810	0.00703	0.00743
JDK	6,434	16.7	−0.223	0.0484	0.0341	47.6	0.00810	0.00696	0.00737
Political blogs	1,222	27.4	−0.221	0.00701	0.00687	14.1	0.0153	0.0135	0.0137
Internet	22,963	4.22	−0.198	0.0146	0.0116	18.4	0.0165	0.0140	0.0148
AS Caida	26,475	4.03	−0.195	0.0240	0.0140	18.5	0.0173	0.0144	0.0157
EU email	224,832	3.02	−0.189	0.00340	0.00328	15.2	0.0107	0.00975	0.0101
UC Irvine	1,893	14.6	−0.188	0.00643	0.00608	28.6	0.0233	0.0208	0.0214
Linux, mailing list	24,567	12.9	−0.185	0.00395	0.00386	147	0.00490	0.00448	0.00452
AS Oregon	6,474	3.88	−0.182	0.0868	0.0429	19.1	0.0281	0.0216	0.0249
Linux, soft.	30,817	13.8	−0.175	0.0256	0.0197	94.1	0.00670	0.00585	0.00616
Gnutella, Aug. 25, 2002	22,663	4.83	−0.173	0.000815	0.000464	1.79	0.108	0.0916	0.104

Les Miserables	77	6.60	−0.165	0.0492	0.0482	3.05	0.123	0.0833	0.0919
Petster-cats	148,826	73.2	−0.164	0.00687	0.00635	405	0.000900	0.000847	0.000855
C. Elegans, neural	297	14.5	−0.163	0.0189	0.0176	10.1	0.0511	0.0410	0.0434
Libimseti	220,970	156	−0.139	0.000406	0.000398	348	0.00110	0.00106	0.00106
David Copperfield	112	7.59	−0.129	0.0473	0.0397	7.57	0.103	0.0760	0.0844
Political books	105	8.40	−0.128	0.0444	0.0419	0.313	0.133	0.0838	0.0927
Google	15,763	18.9	−0.122	0.0430	0.0303	65.1	0.00670	0.00575	0.00608
Social 3	32	5.00	−0.119	0.0665	0.0568	2.16	0.265	0.167	0.205
Euron	33,696	10.7	−0.116	0.00379	0.00361	43.9	0.00910	0.00844	0.00859
Web Stanford	255,265	15.2	−0.116	0.0245	0.0230	117	0.00250	0.00223	0.0023
Bay Wet	128	32.4	−0.112	0.0151	0.0147	25.6	0.0301	0.0252	0.0259
Bay Dry	128	32.9	−0.104	0.0148	0.0145	25.7	0.0301	0.0249	0.0256
Gnutella, Aug. 30, 2002	36,646	4.82	−0.104	0.000672	0.000604	2.19	0.0897	0.0773	0.0856
Gnutella, Aug. 31, 2002	62,561	4.73	−0.0927	0.000921	0.000731	1.85	0.0881	0.0759	0.0844
Petster-hamster	1,788	14.0	−0.0889	0.0100	0.00938	21.6	0.0249	0.0217	0.0223
Petster-dogs	426,485	40.1	−0.0884	0.00176	0.00157	300	0.00140	0.00135	0.00136
Network Science	379	4.82	−0.0817	0.0794	0.0705	2.21	0.230	0.0964	0.110
AS Skitter	1,694,616	13.1	−0.0814	0.00746	0.00709	251	0.00160	0.00149	0.00151
Slashdot zoo	79,116	11.8	−0.0746	0.00229	0.00218	42.8	0.00830	0.00767	0.00778
Wikipedia, edits	113,123	35.8	−0.0651	0.00295	0.00266	169	0.0027	0.00253	0.00255
CiteSeer	365,154	9.43	−0.0632	0.0177	0.0110	4.54	0.0202	0.0172	0.0183
Cora	23,166	7.70	−0.0553	0.0100	0.00898	3.66	0.0381	0.0317	0.0334
Thesaurus	23,132	25.7	−0.0477	0.0017	0.00156	44.7	0.0105	0.0100	0.0102
DBLP, citations	12,495	7.93	−0.0461	0.0282	0.0174	12.0	0.0277	0.0234	0.0251
Dolphins	62	5.13	−0.0436	0.0526	0.0493	1.26	0.231	0.139	0.164
DBpedia	3,915,921	6.42	−0.0427	0.201	0.0840	153	0.00190	0.00140	0.00180
Wikipedia, pages	2,070,367	40.9	−0.0418	0.00477	0.00294	194	0.00130	0.00124	0.00127
Epinions	75,877	10.7	−0.0406	0.00219	0.00211	79.8	0.00570	0.00543	0.00548
Slashdot	51,083	4.56	−0.0347	0.144	0.0347	20.2	0.0219	0.0170	0.0201
Hep-Th, citations	27,400	25.7	−0.0305	0.00931	0.00705	20.3	0.00990	0.00899	0.00922
S 838	512	3.20	−0.0300	0.179	0.0340	0.889	0.382	0.200	0.297
Gowalla	196,591	9.67	−0.0293	0.0180	0.00764	60.0	0.00650	0.00585	0.00609
Amazon, Mar. 12, 2003	400,727	11.7	−0.0203	0.118	0.0381	7.97	0.0273	0.0178	0.0227
Amazon, Jun. 6, 2003	403,364	12.1	−0.0176	0.0891	0.0279	7.87	0.0252	0.0175	0.0219
Amazon, May. 5, 2003	410,236	11.9	−0.0169	0.0843	0.0309	7.79	0.0249	0.0172	0.0214
Air traffic	1,226	3.93	−0.0152	0.0191	0.0154	1.38	0.152	0.109	0.127
Gnutella, Aug. 4, 2002	10,876	7.35	−0.0132	0.00469	0.00377	4.26	0.0685	0.0586	0.0637
Gnutella, Aug. 24, 2002	26,498	4.93	−0.00778	0.214	0.0800	8.68	0.0865	0.0511	0.0700
Hep-Ph, citations	34,401	24.5	−0.00644	0.00421	0.00323	3.61	0.0143	0.0131	0.0133
S 420	252	3.17	−0.00591	0.0542	0.0172	0.398	0.400	0.229	0.320
Amazon, May. 2, 2003	262,111	6.87	−0.00248	0.106	0.0114	0.771	0.0605	0.0425	0.0508
S 208	122	3.10	−0.00201	0.0419	0.0301	0.475	0.437	0.244	0.338
Digg	29,652	5.72	0.00265	0.00457	0.00346	12.5	0.0369	0.0324	0.0348
US Power grid	4,941	2.67	0.00346	0.0409	0.0386	0.874	0.396	0.134	0.157
Gnutella, Aug. 5, 2002	8,842	7.20	0.0146	0.00931	0.00855	7.68	0.0505	0.0425	0.0453
Jazz	198	27.7	0.0202	0.0143	0.0141	12.6	0.0301	0.0250	0.0257

Gnutella, Aug. 9, 2002	8,104	6.42	0.0331	0.00782	0.00737	13.1	0.0409	0.0351	0.0370
Gnutella, Aug. 8, 2002	6,299	6.60	0.0355	0.00795	0.00752	13.7	0.0413	0.0352	0.0371
LiveJournal	5,189,808	18.8	0.0394	0.00157	0.00157	42.7	0.00240	0.00186	0.00186
High school, 2012	180	24.7	0.0464	0.0102	0.0101	4.49	0.0401	0.0332	0.0344
Open flights	2,905	10.8	0.0489	0.00963	0.00942	20.6	0.0181	0.0159	0.0162
Gnutella, Aug. 6, 2002	8,717	7.23	0.0516	0.0103	0.00957	3.20	0.0545	0.0447	0.0478
URV email	1,133	9.62	0.0782	0.00956	0.00865	3.78	0.0581	0.0482	0.0512
High school, 2011	126	27.1	0.0829	0.0173	0.0171	11.6	0.0361	0.0294	0.0304
DBLP, collaborations	1,137,114	8.83	0.0964	0.00797	0.00840	0.0594	0.0113	0.00847	0.00855
MathSciNet	332,689	4.93	0.103	0.0110	0.0103	1.56	0.0347	0.0277	0.0291
Social 1	67	4.24	0.103	0.0486	0.0418	0.975	0.292	0.179	0.223
Cond-Mat, 1993-2003	21,363	8.55	0.125	0.0103	0.00947	7.41	0.0309	0.0264	0.0275
Protein 1	95	4.48	0.129	0.0723	0.0670	0.314	0.384	0.187	0.232
Cond-Mat, 1995-1999	13,861	6.44	0.157	0.0163	0.0146	3.34	0.0509	0.0400	0.0424
College football	115	10.7	0.162	0.00977	0.00967	1.50	0.124	0.0928	0.102
Cond-Mat, 1995-2003	27,519	8.44	0.166	0.00917	0.00847	6.09	0.0293	0.0248	0.0258
US Patents	3,764,117	8.77	0.168	0.0103	0.0100	8.05	0.0113	0.00885	0.00899
Facebook links	63,392	25.8	0.177	0.00143	0.00140	25.8	0.00810	0.00754	0.00762
Cond-Mat, 1995-2005	36,458	9.42	0.177	0.00814	0.00761	12.7	0.0223	0.0195	0.0201
Hep-Th, 1995-1999	5,835	4.74	0.185	0.0523	0.0523	3.70	0.0913	0.0554	0.0587
AstroPhys, 1993-2003	17,903	22.0	0.201	0.00447	0.00432	18.9	0.0117	0.0106	0.0108
Protein 2	53	4.64	0.209	0.0536	0.0500	0.722	0.305	0.172	0.210
Facebook wall	43,953	8.30	0.216	0.00229	0.00214	7.86	0.0277	0.0252	0.0261
Dublin	410	13.5	0.226	0.0263	0.0261	3.62	0.0601	0.0428	0.0448
Actor coll. net.	374,511	80.2	0.226	0.000600	0.000599	429	0.00120	0.00118	0.00118
Astrophysics	14,845	16.1	0.228	0.00504	0.00494	5.64	0.0155	0.0135	0.0138
PGP	10,680	4.55	0.238	0.0166	0.0163	4.25	0.0301	0.0236	0.0243
Hep-Th, 1993-2003	8,638	5.74	0.239	0.0312	0.0312	8.03	0.0669	0.0322	0.0333
Reactome	5,973	48.8	0.241	0.00414	0.00413	27.2	0.00550	0.00481	0.00483
Flickr	105,722	43.8	0.247	0.00105	0.00105	101	0.00170	0.00162	0.00163
E. Coli, transcription	97	4.37	0.412	0.0854	0.0807	0.327	0.328	0.153	0.184
Hep-Ph, 1993-2003	11,204	21.0	0.630	0.00389	0.00389	153	0.00450	0.00408	0.0041
GR-QC, 1993-2003	4,158	6.46	0.639	0.0209	0.0209	7.49	0.0273	0.0219	0.0225

Table A.1: Properties of the set with 99 networks of distinct types. We show the network size N, the average degree $\langle k \rangle$, the Pearson coefficient P, the IPRs of both A_{ij} and critical B_{ij} matrices, the spectral gap of A_{ij}, the thresholds of simulations (λ_c), QMF (λ_c^{QMF}) and PQMF ($\lambda_c^{\mathrm{PQMF}}$) theories.

Appendix B

Complementary analyses for localization

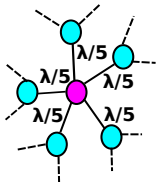
B.1 Contact process (CP)

The contact process (CP) is a model in which an infected individual heals with a constant rate μ and infects its nearest neighbors with a rate λ/k, in which k is the degree of the infected node. In Figure B.1(a) the processes of infection and healing considering an infected node of degree $k = 5$ is illustrated. Each susceptible neighbor of this node is infected with rate $\lambda/5$ while this infected node heals with rate μ. The CP presents a transition from an inactive phase to an active phase at the epidemic threshold, Figure B.1(b). Besides the similarities with the SIS dynamics, this dynamics presents some very distinct features. One of them is the activation on random power-law networks which is triggered by hubs in the SIS (see Sec. 4.3) and is collective for CP [30]. Other differences can be seen considering mean-field theories.

B.1.1 HMF theory

The dynamical equation to the density of infected nodes of degree k for the CP, ρ_k, is

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


λ/5
λ/5
λ/5
λ/5
λ/5

critical exponent $\beta = 1/(\gamma - 2)$, for $2 < \gamma < 3$ and $\beta = 1$ for $\gamma > 3$ [74]. These results show that the critical behavior of the CP predicted by the HMF theory is different of the SIS model.

The critical prevalence of the CP in the quasi-stationary (QS) state as a function of the network size, goes to zero as $\rho_{qs} \sim (gN)^{-1/2}$ [108] where $g = \langle k^2 \rangle / \langle k \rangle^2$. This

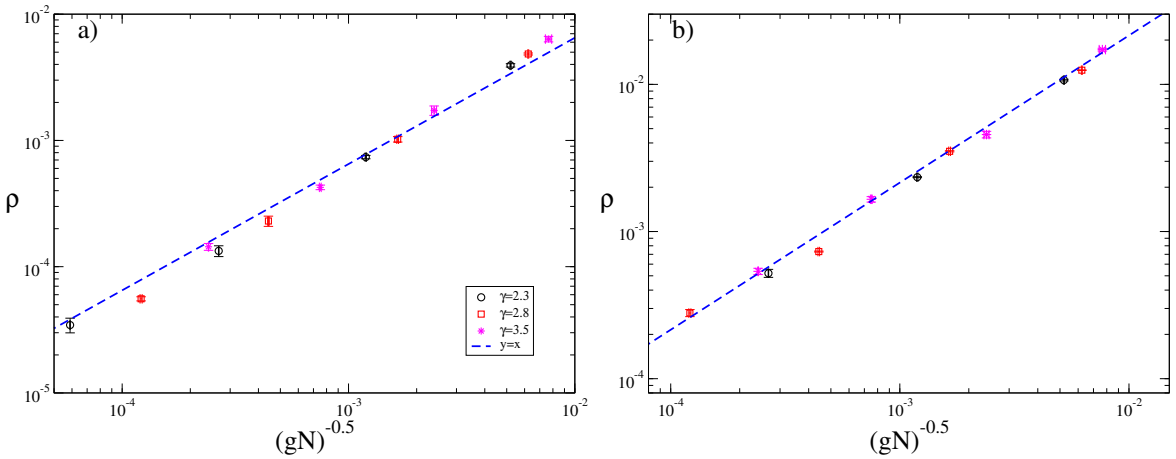


Figure B.2: Scaling of the critical prevalence of the SIS and CP with self-activation as a function of $(gN)^{-1/2}$ for different values of γ and netowrk sizes $N = 10^4$ to 10^7.

relation holds for CP and SIS on annealed networks using self-activation at the effective transition point given by the maximal of the dynamical susceptibility χ [26], as shown in Figure B.2, in which we considered UCM networks of size up to $N = 10^7$. Due to finite-size effects a perfect match is expected only in the thermodynamic limit [74] .

B.1.2 QMF theory

The QMF equation for CP including a self-activation is given by,

$$\frac{d\rho_i}{dt} = -\mu\rho_i + \lambda(1-\rho_i)\sum_{j=1}^{N}\frac{A_{ij}}{k_j}\rho_j + f(1-\rho_i). \tag{B.4}$$

The first term corresponds to the healing and the second the infection process. Finally, the third term is the spontaneous infection by self-activation.

In the absence of self-activation ($f = 0$), a linear stability analysis can be performed around the trivial solution $\rho_i = 0$. It results in an epidemic threshold

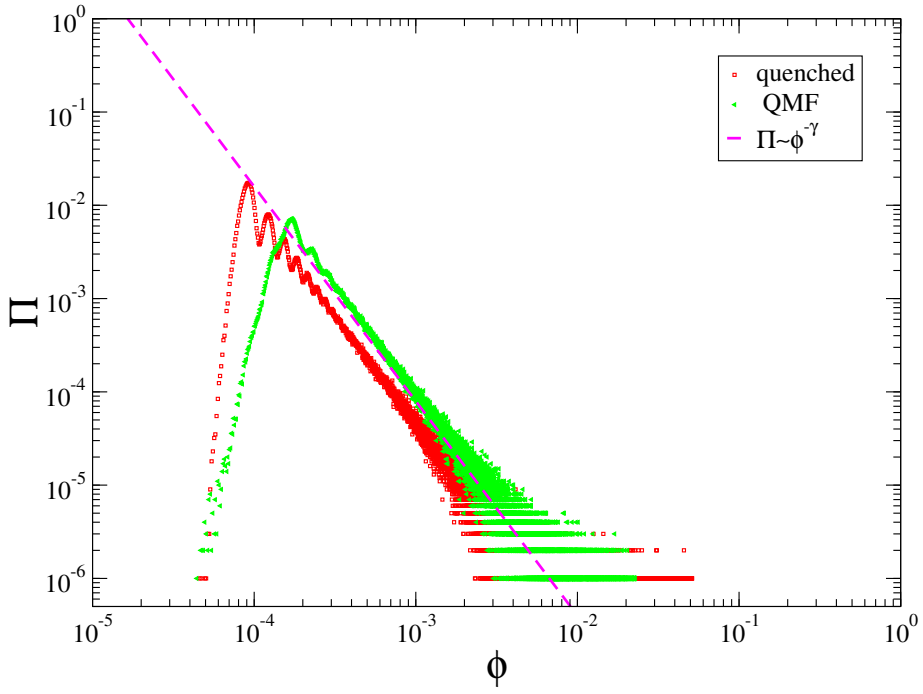


Figure B.3: NAVCPD for stochastic simulations and QMF theory for CP model on a single sample of the UCM network at the epidemic threshold determined by the susceptibility ($\lambda = 1.12$). The network size is $N = 10^6$, $\gamma = 2.3$ and $k_c = 2\sqrt{N}$.

$\lambda_c = \mu/\Lambda_1$, in which Λ_1 is the largest eigenvalue of the matrix A_{ij}/k_j. Note that $v_i = k$ is a positive eigenvector of this matrix with eigenvalue $\Lambda_1 = 1$. Thus the Perron-Frobenius theorem guarantees that its the largest one and $\lambda_c = \mu$ for any degree correlation or distribution.

We compare the NAVCPD distributions obtained by integration of Eq. (B.4) and stochastic simulations at the epidemic threshold determined by the maximum of the susceptibility on UCM networks presenting a power-law degree distribution $P(k) \sim k^{-\gamma}$. In both cases the NAVCPD presents peaks in which $\phi_i \propto k_i$ and scales as does the degree distribution $\Pi \sim \phi^{-\gamma}$, Figure B.3. The peaks are less pronounced for the QMF theory. It is due to the value of $\lambda > \mu$ considered to correspond to a supercritical regime for the QMF theory.

B.2 Finite-size effects in the IPR of HMF theory

We complement the analyses performed for the IPR of the HMF theory on uncorrelated networks with power-law degree distribution developed in Sec. 7.3.1. Considering that $P(k) = Ak^{-\gamma}$, in which A is the normalization constant, $\langle k^n \rangle$ is given by

$$\langle k^n \rangle = \int_{k_0}^{k_{\max}} k^n P(k) dk = \frac{A k_0^{-\gamma+n+1}}{(n+1-\gamma)} \left[\left(\frac{k_{\max}}{k_0} \right)^{-\gamma+n+1} - 1 \right]. \tag{B.5}$$

Considering Eq. (7.17), in which

$$Y_4 = \frac{\langle k^4 \rangle}{N \langle k^2 \rangle^2}, \tag{B.6}$$

for the structural cutoff and using Eq. (B.5), the IPR becomes

$$Y_4 = \frac{1}{N} \left(\frac{(3-\gamma)^2}{5-\gamma} \right) \left(\frac{k_0^{\gamma-1}}{A} \right) \frac{\left[\left(\frac{k_{\max}}{k_0} \right)^{5-\gamma} - 1 \right]}{\left[\left(\frac{k_{\max}}{k_0} \right)^{3-\gamma} - 1 \right]^2}. \tag{B.7}$$

To analyze the finite-size effects, we consider distinct ranges of γ.

For $2 < \gamma < 3$, expanding in powers of $k_{\max}/k_0$, we obtain

$$Y_4 \sim \frac{1}{A} \frac{(3-\gamma)^2}{5-\gamma} \left(\frac{k_{\max}^{\gamma-1}}{N} \right) \left[1 + 2 \left(\frac{k_0}{k_{\max}} \right)^{3-\gamma} + ... \right]. \tag{B.8}$$

Considering a structural cutoff $k_c = 2\sqrt{N}$ for this range of γ, we have that $k_{\max} \approxeq k_c$. Thus,

$$Y_4 \sim N^{-\frac{(3-\gamma)}{2}} \left[1 + 2 \left(\frac{k_0}{2} \right)^{3-\gamma} N^{-\frac{(3-\gamma)}{2}} + O(N^{-\frac{(3-\gamma)}{2}}) \right] \tag{B.9}$$

which implies in corrections of order $N^{-\frac{(3-\gamma)}{2}}$. These corrections become more relevant for γ close to 3. The results obtained in stochastic simulations scale as Eq. (7.17) for the range of size studied (up to $N = 10^7$) but it does not correspond to Eq. (7.18). We remark that the corrections obtained in Eq. (B.9) are relevant and the finite-size

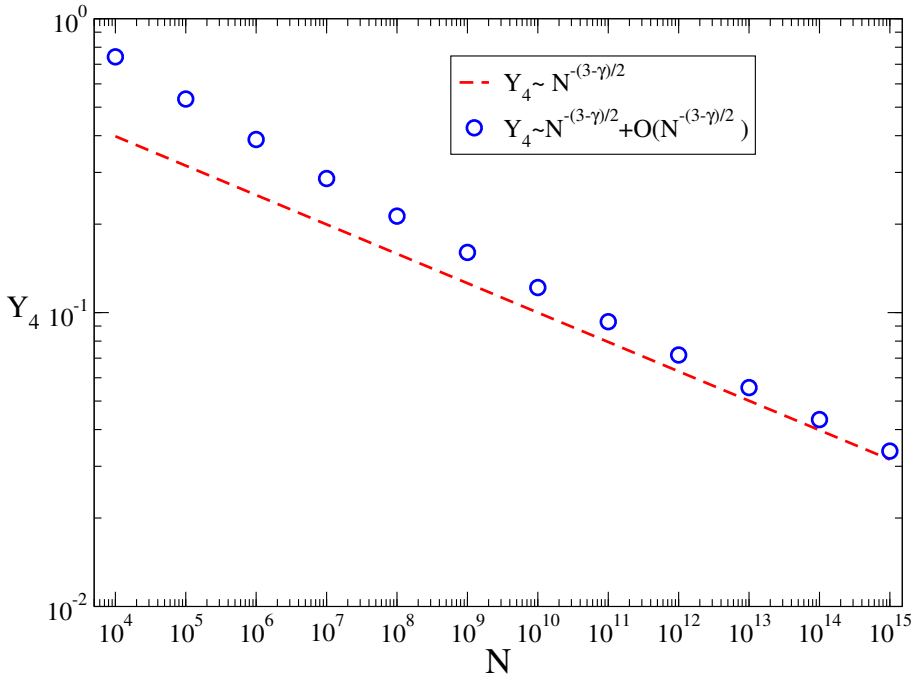


Figure B.4: Comparison between the scaling predicted in Eq. (7.18) (dashed line) and in Eq. (B.9) (symbols) for $\gamma = 2.8$. For sizes larger than $N = 10^{12}$, the scaling predicted by Eq. (B.9) converges to that in Eq. (7.18).

effects introduce a delay in the convergence to the scaling of Eq. (7.18), as shown in Figure B.4.

Following similar steps for $3 < \gamma < 5$,

$$Y_4 \sim \frac{1}{NA}\frac{(3-\gamma)^2}{5-\gamma}\left(\frac{k_{\max}}{k_0}\right)^{5-\gamma}\left[1 + 2\left(\frac{k_{\max}}{k_0}\right)^{3-\gamma} + ...\right] \tag{B.10}$$

For a cutoff $k_c = N$, we have $k_{\max} \sim N^{\frac{1}{\gamma-1}}$ [74] and the leading correction in the scaling present in Eq. (7.18) is the order $N^{\frac{(3-\gamma)}{\gamma-1}}$. For $\gamma = 3.5$, the scaling obtained for the range of the networks size studied is compatible with that predicted in Eq. (7.17) as shown in Figure 7.7. Fluctuations around this scaling are observed and they are consequence of the large fluctuations in the average value of $k_{\max}$ [74], as shown in Figure B.5.

For $\gamma > 5$, the scaling predicted in Eq. (7.18) for the IPR as a function of the size indicates a delocalized state. This result is verified for $\gamma = 6$ in Figure B.6 .

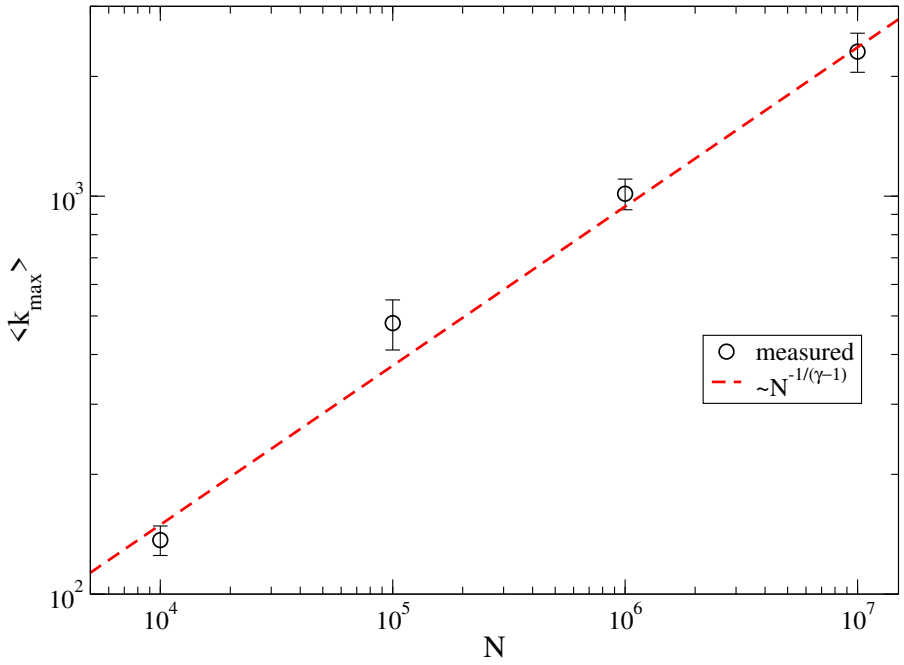


Figure B.5: Average largest degree as a function of the network size for a power-law degree distribution with $\gamma = 3.5$. Dashed line indicates the expected value while the symbols are the values obtained in simulations. Averages were performed considering 50 independent networks.

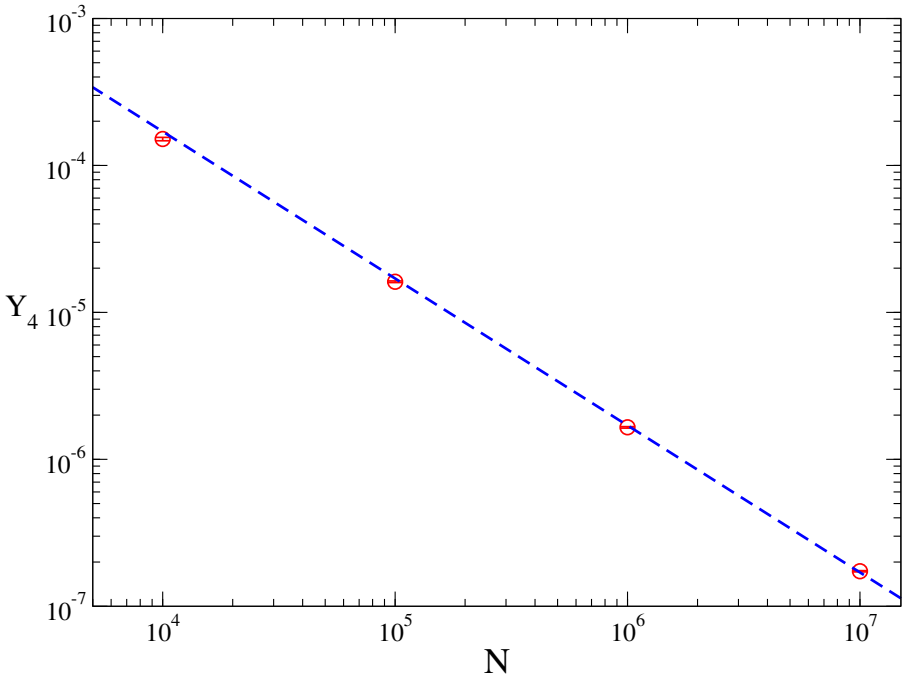


Figure B.6: IPR as a function of the network size for annealed network presenting a power-law degree distribution with $\gamma = 6$. The dashed line indicates the scaling corresponding to a delocalized state where $Y_4 \sim N^{-1}$.

Bibliography

[1] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, UK, 2008.

[2] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833), 2001.

[3] F. Cheng, I. A. Kovács, and A.-L. Barabási. Network-based prediction of drug combinations. *Nat. Commun.*, 10(1197), 2019.

[4] M. Rosas-Casals, S. Valverde, and R. V. Solé. Topological vulnerability of the european power grid under errors and attacks. *Int. J. Bifurcat. Chaos*, 17(07):2465–2475, 2007.

[5] R. Albert, I. Albert, and G. L. Nakarado. Structural vulnerability of the north american power grid. *Phys. Rev. E*, 69:025103, 2004.

[6] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PLoS One*, 5(7):1–9, 2010.

[7] 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(1):35, 2019.

[8] N. Dwi Prasetyo and C. Hauff. Twitter-based election prediction in the developing world. In *Proceedings of the 26th ACM Conference on Hypertext and Social*

Media, page 149158, New York, USA, 2015. Association for Computing Machinery.

[9] A. Bovet and H. A. Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nat. Commun.*, 10:7.

[10] D. Gayo-Avello. A meta-analysis of state-of-the-art electoral prediction from twitter data. *Soc. Sci. Comput. Rev.*, 31(6):649–679, 2013.

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

[12] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87:258701, 2001.

[13] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.

[14] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, 2015.

[15] W. Wang, M. Tang, H. E. Stanley, and L. A Braunstein. Unification of theoretical approaches for epidemic spreading on complex networks. *Rep. Prog. Phys.*, 80(3):036603, 2017.

[16] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[17] F. A. Rodrigues, T. Peron, C. Connaughton, J. Kurths, and Y. Moreno. A machine learning approach to predicting dynamical observables from network structure. *arXiv:1910.00544*, 2019.

[18] C. Castellano and R. Pastor-Satorras. Competing activation mechanisms in epidemics on networks. *Sci. Rep.*, 2:371, 2012.

[19] 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.

[20] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[21] M. F. C. Gomes, A. Pastore y Piontti, L. Rossi, D. Chao, I. Longini, M. E. Halloran, and A. Vespignani. Assessing the international spreading risk associated with the 2014 west african ebola outbreak. *PLOS Currents Outbreaks*, 6, 2014.

[22] Á. G. Muñoz, M. C. Thomson, A. M. Stewart-Ibarra, G. A. Vecchi, X. Chourio, P. Nájera, Z. Moran, and X. Yang. Could the recent zika epidemic have been predicted? *Front. Microbiol.*, 8:1291, 2017.

[23] 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(1):1–8, 2011.

[24] D. J.D. Earn, J. Dushoff, and S. A. Levin. Ecology and evolution of the flu. *Trends Ecol. Evol.*, 17(7):334 – 340, 2002.

[25] M. R. Golden, W. L. H. Whittington, H. H. Handsfield, J. P. Hughes, W. E. Stamm, M. Hogben, A. Clark, C. Malinski, J. R. L. Helmers, K. K. Thomas, and K. K. Holmes. Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infection. *N. Engl. J. Med.*, 352(7):676–685, 2005.

[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] 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.

[28] M. M. de Oliveira and R. Dickman. How to simulate the quasistationary state. *Phys. Rev. E*, 71:016129, 2005.

[29] 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.

[30] 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.

[31] S. Chatterjee and R. Durrett. Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.*, 37(6):2332–2356, 2009.

[32] T. Mountford, D. Valesin, and Q. Yao. Metastable densities for the contact process on power law random graphs. *Electron. J. Probab.*, 18:36 pp., 2013.

[33] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.

[34] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, 63:066117, 2001.

[35] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E*, 65:035108, 2002.

[36] Y. Wang, D. Chakrabarti, W. Chenxi, and C. Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.*, pages 25–34, 2003.

[37] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 10(4), 2008.

[38] M. Boguñá and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Phys. Rev. E*, 66:047104, 2002.

[39] 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.

[40] J. P. Gleeson. High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.*, 107:068701, 2011.

[41] J. P. Gleeson. Binary-state dynamics on complex networks: pair approximation and beyond. *Phys. Rev. X*, 3:021004, 2013.

[42] E. Cator and P. Van Mieghem. Second-order mean-field susceptible-infected-susceptible epidemic threshold. *Phys. Rev. E*, 85:056111, 2012.

[43] G. F. de Arruda, F. A. Rodrigues, and Y. Moreno. Fundamentals of spreading processes in single and multilayer complex networks. *Phys. Rep.*, 756:1–59, 2018.

[44] 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.

[45] P. Van Mieghem. The viral conductance of a network. *Comput. Commun.*, 35(12):1494–1506, 2012.

[46] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.*, 90:028701, 2003.

[47] 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.

[48] D. H. Silva, F. A. Rodrigues, and S. C. Ferreira. High prevalence regimes in the pair-quenched mean-field theory for the susceptible-infected-susceptible model on networks. *Phys. Rev. E*, 102:012313, 2020.

[49] M. Shrestha, S. V. Scarpino, and C. Moore. Message-passing approach for recurrent-state epidemic models on networks. *Phys. Rev. E*, 92:022821, 2015.

[50] C. Castellano and R. Pastor-Satorras. Relevance of backtracking paths in recurrent-state epidemic spreading on networks. *Phys. Rev. E*, 98:052313, 2018.

[51] V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reactiondiffusion processes and metapopulation models in heterogeneousănetworks. *Nat. Phys.*, 3:276–282, 2007.

[52] D. H. Silva and S. C. Ferreira. Activation thresholds in epidemic spreading with motile infectious agents on scale-free networks. *Chaos*, 28(12):123112, 2018.

[53] D. H. Silva. Efeito de mobilidade no limiar epidêmico da dinâmica sis em redes livres de escala. Master's thesis, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil, march 2016.

[54] M. Newman. *Networks: an introduction*. Oxford University Press, Oxford, USA.

[55] F. A. Rodrigues. *Network centrality: an introduction*.

[56] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, Nan Zhang, G. Li, and R. Chen. Topological structure analysis of the proteinprotein interaction network in budding yeast. *Nucleic Acids Res.*, 31(9):2443–2450, 2003.

[57] W. W. Zachary. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.*, 33(4):452–473, 1977.

[58] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. 2009.

[59] D. Dünker and J. Kunegis. Social networking by proxy: a case study of catster, dogster and hamsterster. *arXiv:1501.04527*, 2015.

[60] Dogster friendships network dataset – KONECT, apr 2017.

[61] Youtube friendship network dataset – KONECT, apr 2017.

[62] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th*

ACM SIGCOMM Conference on Internet Measurement, page 2942, New York, USA, 2007. Association for Computing Machinery.

[63] F. Radicchi. Predicting percolation thresholds in networks. *Phys. Rev. E*, 91:010801, 2015.

[64] Air traffic control network dataset – KONECT, apr 2017.

[65] P. Bonacich. Power and centrality: a family of measures. *Am. J. Sociol.*, 92(5):1170–1182, 1987.

[66] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Adv. Phys.*, 51(4):1079–1187, 2002.

[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] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B″*, 38(2):205–209, 2004.

[69] S. Weber and M. Porto. Generation of arbitrarily two-point-correlated random networks. *Phys. Rev. E*, 76:046111, 2007.

[70] P. Erdős and A. Rényi. On random graphs. *Publ. Math. 6*, page 290297, 1959.

[71] A. Barrat and R. Pastor-Satorras. Rate equation approach for correlations in growing network models. *Phys. Rev. E*, 71:036127, 2005.

[72] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30(4):1141–1144, 1959.

[73] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms*, 6(23):161–180, 1995.

[74] 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.

[75] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs and mathematical tables*. Dover Publications, Inc., New York, USA, 1965.

[76] F. Chung, L. Lu, and V. Vu. Eigenvalues of random power-law graphs. *Ann. Comb.*, 7(1):21–33, 2003.

[77] 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.

[78] R. Pastor-Satorras and C. Castellano. Eigenvector localization in real networks and its implications for epidemic spreading. *J. Stat. Phys.*, 173:1110–1123, 2018.

[79] R. Pastor-Satorras and C. Castellano. Distinct types of eigenvector localization in networks. *Sci. Rep.*, 6:18847, 2016.

[80] P. Pradhan, A. Yadav, S. K. Dwivedi, and S. Jalan. Optimized evolution of networks for principal eigenvector localization. *Phys. Rev. E*, 96:022312, 2017.

[81] A. S. Mata and S. C. Ferreira. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *EPL*, 103(4):48003, 2013.

[82] J. T. Matamalas, A. Arenas, and S. Gómez. Effective approach to epidemic containment using link equations in complex networks. *Sci. Adv.*, 4(12):4212, 2018.

[83] Q. Wu, R. Zhou, and T. Hadzibeganovic. Conditional quenched mean-field approach for recurrent-state epidemic dynamics in complex networks. *Physica A*, 518:71–79, 2019.

[84] D. ben Avraham and J. Köhler. Mean-field (n,m)-cluster approximation for lattice models. *Phys. Rev. A*, 45:8358–8370, 1992.

[85] M. Henkel, H. Hinrichsen, and S. Lübeck. *Non- equilibrium phase transition: Volume1: Absorbing phase transitions*. Spring, Netherlands, 2008.

[86] J. A. Hołłyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki. Universal scaling of distances in complex networks. *Phys. Rev. E*, 72:026108, 2005.

[87] Z.-W. Wei and B.-H. Wang. Susceptible-infected-susceptible model on networks with eigenvector localization. *Phys. Rev. E*, 101:042310, 2020.

[88] B. Karrer and M. E. J. Newman. Message passing approach for general epidemic models. *Phys. Rev. E*, 82:016101, 2010.

[89] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, 2011.

[90] C. Castellano and R. Pastor-Satorras. Relevance of backtracking paths in recurrent-state epidemic spreading on networks. *Phys. Rev. E*, 98:052313, 2018.

[91] L. Ménard and A. Singh. Percolation by cumulative merging and phase transition for the contact process on random graphs. *Ann. Sci. de l'Ecole Norm. Superieure*, 49:1189–1238, 2016.

[92] C. Castellano and R. Pastor-Satorras. Cumulative merging percolation and the epidemic transition of the susceptible-infected-susceptible model in networks. *Phys. Rev. X*, 10:011070, 2020.

[93] 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.

[94] R. S. Sander, G. S. Costa, and S. C. Ferreira. Sampling methods for the quasi-stationary regime of epidemic processes on regular and complex networks. *Phys. Rev. E*, 94:042308, 2016.

[95] P. Van Mieghem, H. Wang, X. Ge, S. Tang, and F. A. Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *Eur. Phys. J. B*, 76:643–652, 2010.

[96] X. Huang and R. Durrett. The contact process on random graphs and galton-watson trees. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 17:159–182, 2020.

[97] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno. Discrete-time markov chain approach to contact-based disease spreading in complex networks. *EPL*, 89(3):38009, 2010.

[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] R. S. Ferreira and S. C. Ferreira. Critical behavior of the contact process on small-world networks. *Eur. Phys. J. B*, 86:462, 2013.

[100] R. Juhász, G. Ódor, C. Castellano, and M. A. Muñoz. Rare-region effects in the contact process on networks. *Phys. Rev. E*, 85:066125, 2012.

[101] M. M. de Oliveira, S. G. Alves, and S. C. Ferreira. Dynamical correlations and pairwise theory for the symbiotic contact process on networks. *Phys. Rev. E*, 100:052302, 2019.

[102] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3):75 – 174, 2010.

[103] D. H. Silva and S. C. Ferreira. Dissecting localization phenomena of dynamical processes on networks. *Journal of Physics: Complexity*, 2021.

[104] H. E. Stanley. *Introduction to phase transitions and critical phenomena*. Oxford University Press, New York, United States, 1987.

[105] R. S. Ferreira, R. A. da Costa, S. N. Dorogovtsev, and J. F. F. Mendes. Metastable localization of diseases in complex networks. *Phys. Rev. E*, 94:062305, 2016.

[106] C. Castellano and R. Pastor-Satorras. Non-mean-field behavior of the contact process on scale-free networks. *Phys. Rev. Lett.*, 96:038701, 2006.

[107] C. Castellano and R. Pastor-Satorras. Routes to thermodynamic limit on scale-free networks. *Phys. Rev. Lett.*, 100:148701, 2008.

[108] S. C. Ferreira, R. S. Ferreira, and R. Pastor-Satorras. Quasistationary analysis of the contact process on annealed scale-free networks. *Phys. Rev. E*, 83:066113, 2011.

[109] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80:1275–1335, 2008.

[110] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, 71:027103, 2005.

[111] G. Ódor. Localization transition, lifschitz tails, and rare-region effects in network models. *Phys. Rev. E*, 90:032110, 2014.

[112] 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.

[113] S. C. Ferreira, R. S. Ferreira, C. Castellano, and R. Pastor-Satorras. Quasistationary simulations of the contact process on quenched networks. *Phys. Rev. E*, 84:066102, 2011.

[114] A. S. Mata, R. S. Ferreira, and S. C. Ferreira. Heterogeneous pair-approximation for the contact process on complex networks. *New J. Phys.*, 16(5):053006, 2014.

[115] H. Hong, M. Ha, and H. Park. Finite-size scaling in complex networks. *Phys. Rev. Lett.*, 98:258701, 2007.

[116] M. Ha, H. Hong, and H. Park. Comment on "non-mean-field behavior of the contact process on scale-free networks". *Phys. Rev. Lett.*, 98:029801, 2007.

[117] C. Castellano and R. Pastor-Satorras. Castellano and pastor-satorras reply:. *Phys. Rev. Lett.*, 98:029802, 2007.

[118] G. St-Onge, V. Thibeault, A. Allard, L. J. Dubé, and L. Hébert-Dufresne. Master equation analysis of mesoscopic localization in contagion dynamics on higher-order networks. *Phys. Rev. E*, 2021.

[119] K. Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties*, pages 211–280, Tokyo, Japan, 1989. Mathematical Society of Japan.

[120] 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.

[121] P. Holme. Modern temporal network theory: a colloquium. *Eur. Phys. J. B*, 88:234, 2015.

[122] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.

[123] F. A. Rodrigues, T. K. DM. Peron, P. Ji, and J. Kurths. The kuramoto model in complex networks. *Phys. Rep.*, 610:1–98, 2016.

[124] F. Radicchi and C. Castellano. Breaking of the site-bond percolation universality in networks. *Nat. Commun.*, 6:10196, 2015.