Università degli studi di Padova

Dipartimento di Fisica e Astronomia

Corso di laurea magistrale in Fisica

Decay of localized states in the SIS model

Laureando:
guido zampieri

Relatori:
Sergey dorogovtsev
fulvio baldovin

Co-relatore:
Ronan Silva Ferreira

Anno Accademico 2013/2014

Abstract

The SIS model is a stochastic double-state process, valuable to represent the epidemic spreading within idealized populations of interacting individuals. Its exact mathematical formalization has been one of the goals of recent research on dynamical processes running on networks. In particular, investigations has focused on graphs characterized by highly heterogeneous degree distributions and short average distances between nodes. Such architectures are very common in nature and are known with the name of complex networks. Besides, in the thermodynamic limit the SIS model exhibits an absorbing phase transition from a regime of short-lived active contagion to a stationary state with constant non-null density of infective individuals. Critical features - such as finite size scaling - can be explored thanks to traditional phase transitions tools, inherited from related non-equilibrium processes defined on a lattice.
Novel developments in the study of the SIS model on networks concern structures wherein a few strongly connected hubs join a large multitude of peripheral nodes. Under such circumstances, it has been shown that the adjacency matrix - which is the mathematical representation of connections pattern - undergoes a localization transition, akin to localization in condensed-matter systems. Theoretical investigations suggest that a dynamical consequence is the perpetual outliving of infections within a finite number of individuals, vanishing in the large size limit. Here we inspect these predictions by means of a numerical approach applied on networks with uniform connectivity over all nodes with the exception of a single hub. The configuration model is exploited to randomly build ensembles of graphs, while a continuous time algorithm simulates the SIS dynamics. Findings partially agree with analytical conjectures. Indeed, a double peaked epidemic susceptibility is observed, indicating a modified transition occurring in two stages: by stepwise increasing the infection rate, a contagion outbreak in the hub neighbourhood anticipates the usual global transition. However, the double peak scenario is not guaranteed by the mere presence of the hub, but it turns out to be dependent on nodes degree relationships. This fact reveals a competition between the hub and the rest of the network in the disease propagation. Moreover, the precise expected threshold for the localization transition is not obeyed by simulation data.

As additional computational aspect, we analyze the same typology of network through a second order parameter, recently claimed to be more efficient than susceptibility. The quantity under scrutiny is the life span of non-exploding realizations of the infection process. Results indicate that it is sensitive to the localization transition, but not able to furnish any useful information regarding the competition between hub and network. Nevertheless, we take advantage of this method to verify that the instability brought by the hub is actually localized within a restricted number of nodes.

Contents

Introduction

The complex network theory is nowadays understood as a paradigm for studies on emergent phenomena from the interaction of agents in a system. With the advancement on computer processing and storing of large amounts of data - from technological and genetic, as well as socio-economics databases - this theory has gained a strong interdisciplinary character attracting researchers from numerous fields. In particular, the investigation on dynamical processes occurring on these structures is an issue of central importance in a broad spectrum of knowledge, ranging from physics and biology to computer and economic sciences. In the last decades, a big impulse was given to the development of a theoretical framework capable to describe and make predictions on this interdisciplinary subject. All systems of interest consist of many elements interacting together, hence the proper formalism was found in graph theory: in mathematics and computer science, a *graph* is the formal representation of a *network*, made of abstract entities called nodes (or vertices) connected by pairwise relations called edges. Moreover, several typologies of real networks are made of very large systems, whose macroscopic behaviour cannot be understood from elementary features. This fact has favored the use of techniques inherited from statistical physics, which properly capture the connection between microscopic dynamic evolution and collective phenomena. Such instruments were applied to a variety of real world situations and constitutes the foundation of a conspicuous amount of research. Hot topics are, for instance, epidemic models [1][3][7][23], random walks [3], percolation [1][3][7][23], synchronization [3][7][23], and game theory models [3].
In particular, epidemic models touch a wide variety of applications. Examples are the disease contagion within a population, the virus programs diffusion throughout a computer network, the spreading of ideas and habits within a social group, and more. Epidemic phenomena such as these can be efficiently represented within a graph formalism and modeled according to specific aspects of interest. In the case of a human population, nodes may represent individuals - or groups of individuals - and links are the channels through which an infection can pass from one to another - air transmission, physical contact or others. In alternative, one can think of nodes as computers and edges as cables and telephone lines, to consider the propagation

of virus programs.
A typical property of epidemics is their ability in periodically manifesting: they can have an outbreak, grow, vanish and manifest again. A useful tool is then provided by the SIS model. Within the SIS model, the complex scenario of disease transmission, progression or regression is reduced to a double state framework: the possible states that each individual can assume are only susceptible and infective. An infective individual transmits the infection to all other individuals he interacts with, at a certain infection rate, and recovers spontaneously at a certain recovery rate. Vertices evolve by becoming infective, then turning again susceptible and so forth. This stochastic chain explains the name of the model. However, though the simplicity of its rules, it is still not fully understood and lacks of an exhaustive mathematical description.
The main parameter of the model is the reproductive number, that is the ratio between the infection and the recovery rates. It determines whether an initial seed of contagion leads to en epidemic outbreak or not: below a particular value of the reproductive number the disease quickly dies out. Above, instead, the system may enter in a steady state in which a finite fraction of vertices remains permanently infected. This critical value is called *epidemic threshold*. The passage between the two regimes is associated to an absorbing phase transition and shares many features with other transitions in completely different systems. In particular, SIS model belongs to the same universality class of *directed percolation* model [19][20]. This latter process was conceived to simulate the diffusion of agents through random media, and thus it is traditionally studied on lattices. Later on, it was applied also on networks to study the resistance of particular configurations of graphs to attacks and failures which can compromise some nodes or links. Physics of phase transitions acquires, in this context, a key role in the comprehension of dynamic processes in network science.
Two very common features in real networks are the power-law connectivity distribution of nodes - which translates in a topological absence of scale - and the small-world structure - i.e. high node clustering and short average distance between vertices. They have been verified both for the Internet, cellular networks and for smaller systems as scientific collaborations nets, plus many others [6]. The impressive universality of these features led to an intensive research on scale-free and small-world graphs [5][9][11][12][13]. In particular, recent findings spread light on the role of hubs, namely nodes with strongly large connectivity with respect to the average. Depending on topological properties of the network, it has been shown that hubs may lead to the outliving of infections indefinitely in time, localized within a restricted number of individuals. This fact finds a practical confirmation in the surviving of computer viruses for long periods of time - even years - in a very low density [32]. Further investigations are thus required to comprehend which factors lead to the outbreak or to the extinction of an epidemic.

Objectives

The first aim of this project is to verify analytical predictions on the role of hubs in epidemic diffusion. In the specific, recent studies [15][16] claim the possibility of a localization transition in networks, analogous to localization in disordered condensed-matter physics. Spectral analysis reasonings predict that this fact would spoil the traditional double phase scenario, bringing an intermediate phase of low density stationary contagion. In this context, hubs would act as centers of localization of the disease. However, such predictions have not been accurately verified yet.
As a second aim, we focus on computational strategies devoted to the determination of epidemic thresholds. A new technique was recently proposed, building a parallelism between the duration of contagion process realizations and the size of clusters in percolation [31]. The average life span of non-enduring realizations is thus indicated as more efficient order parameter of the transition. The efficiency of this method is put to the test, both in homogeneous and heterogeneous environments.
At the purpose of fathoming the complexity or real networks, it can be instructive to reduce the problem and explore simple but representative structures. We thus concentrate on regular graphs - i.e. graphs with constant connectivity on all vertices - with a single embedded hub of large connectivity. The phase transition is computationally studied around the epidemic threshold, whose estimation is achieved through the measure of fluctuations in the fraction of infective nodes. Their average value in complex networks epidemiology is known as *susceptibility*. Such technique belongs to traditional phase transition tools, used - for instance - in magnetic systems [19][20][23] and, recently, in the contact process on scale-free networks [23][27]. The measure is repeated for different vertex connectivities and for growing sizes of networks, in order to find a route to the thermodynamic limit. In addiction, average life span measures are performed on fully regular networks and regular networks with an additional hub, in order to make a comparison with the susceptibility method and gain an overall perspective on the localization transition.
All results are discussed embracing both network theory and phase transitions theory, trying to get deeper into the understanding of epidemic thresholds, which is still a pressing challenge in the field of network science.

Chapter 1

Fundamentals of complex networks

Complex networks benefit of a self-consistent mathematical formalization, able to represent a broad variety of real systems. Key instruments are macroscopic quantities - taking the form of statistical distributions or matrices - as well as variables referring to single vertices or nodes, like centrality.

A network is a group of interconnected elements. Ordinary experience tells us that we can build a network starting from physical entities such as people, objects, or even both combined together. They are usually joined by relationships of exchange or sharing of resources or information. The world where we live in contains plenty of different typologies of networks - both in nature and in human civilization - and many systems of scientific interest display this structure.
In this highly technological age, the first example one is brought to think about is the World Wide Web. The bulk of web addresses we daily use are interconnected by hyperlinks in such a way to allow the navigation on the web, jumping from a web page to the other. Doing a research on Google or clicking on a page link means to move from a node of the net to another. Under this untouchable web, a physical network lies - the Internet - made of computers and calculation centers scattered on the globe and put into contact by cables and telephone lines. Data packages travel all through this network to make people access to information. It is easy to understand how optimization of data transport and counter-strategies to attacks or damages at parts of the global structure are important issues.
Considering the natural world, a very important type of nets are biochemical networks. They are chains of intracellular interactions intervening in cellular metabolism, in the regulation of genetic expression and in signal

transduction. The whole set of pathways constitutes an intricate web of interactions between macromolecules and other chemical compounds, whose reconstruction and deciphering are fundamental to comprehend the complexity of cellular systems. Changing the scale, one can find nets also in ecology. An ecological network is the set of all species of an ecosystem and interactions between them. Studying that, it becomes possible to understand the origin of some habits in animals or the dynamics leading to the extinction of a species, or again to quantify phenomena concerning energy and carbon flow.
In all cases, it is possible to study the nature of single elements of the network - how a computer works or how an organism behaves - or the nature of the interactions - the communication protocols used on the Internet or the dynamics of human friendship. In alternative, it is possible to concentrate on the pattern of the interconnections. Data on the Internet, for instance, follow routes dictated by the specific architecture of the web, from which depends the delivery efficiency and speed. In the same way, connections in a social network influence how people gather information and form opinions. This third aspect can then have an essential role in the operating principles of a system.
A systematic investigation on networks - in mathematical jargon called *graphs* - requires proper tools. From an abstract point of view, networks are studied as sets of *nodes* - or *vertices* - connected by links called *edges*. This work starts with a survey of the basic mathematical instruments which will be useful in the following.

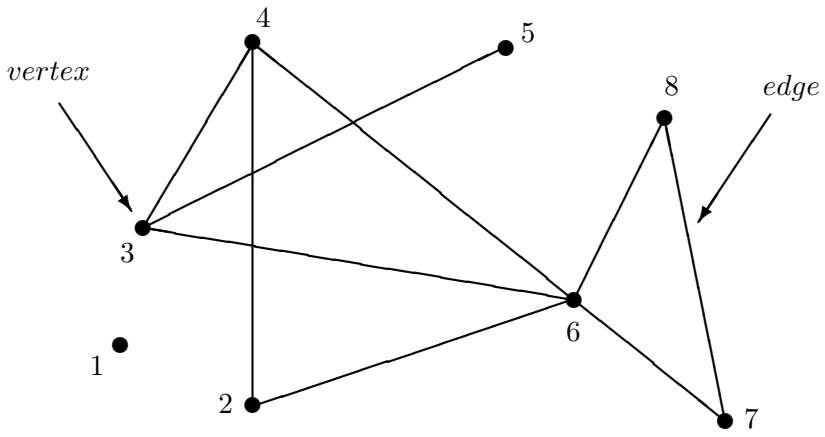


Figure 1.1: A small network composed by eight nodes and nine edges.

1.1 The adjacency matrix

The first useful thing when dealing with networks is to find a comfortable manner to handle and work with the multitude of nodes and edges. In this sense, one needs an appropriate representation. Trivially, the simplest way to do it is with points and lines, as in Figure 1.1. In the early days of graph theory, this was also the only way. However, it is likely to have a very high number of elements to draw - for example, the WWW reaches millions of nodes (see Picture 1.4). The picture becomes then confused and any kind of analysis unfeasible.
A more efficient tool is the matrix representation: given a graph with N nodes, the **adjacency matrix** A is defined as the $N \times N$ matrix whose non diagonal elements A_{ij} correspond to the number of links between two nodes i and j. In the majority of situations, only an edge is present between two vertices, so A contains only 0 and 1. For example, the adjacency matrix of the graph in Figure 1.1 is

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

In case that more than one edge is present between a couple of nodes, they are collectively called *multiedge*. Moreover, it can also happen to find a link connecting a node to itself. In this case this link is called *self-edge* or *self-loop*. The diagonal elements A_{ii} are twice the number of self-loops attached to the node i. For example, for the graph on the right in Figure 1.2 is represented by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \end{pmatrix}$$

It is possible to find multiple self-edges as well. When a network contains no multiedges and no self-loops is said *simple*. In the following, we will assume to consider exclusively simple graphs, unless specified.
Obviously, when the size of the graph is very large it becomes less convenient

also to work with a $N \times N$ matrix. In some cases, it can be avoided, as it will be explained later on.

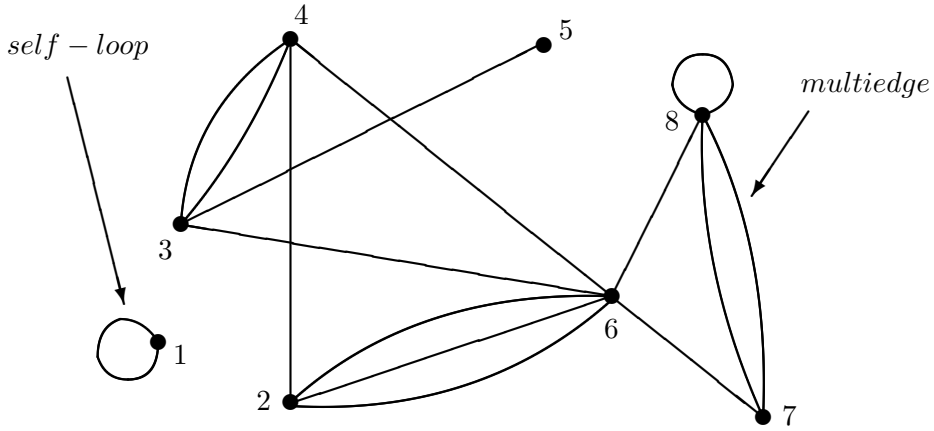


Figure 1.2: A small network with two self-loops and three multiedges.

1.2 Centrality

An advantage of the graphical representation of graphs is the minimal effort necessary to spot the crucial vertices. Looking at Figure 1.3, in only a glance one can tell that the most "important" vertex is the central one: all the branches of the tree need pass through the trunk to communicate within each other. Nevertheless, the definition of centrality is not unique in network theory. According to what is sought, different types of centrality can play their role.
The most common is the **degree centrality**, or simply the **degree**[1], which is the number of links departing from (or ending in) a node. Vertex number 5 in the simple graph in Figure 1.1 has degree one, while vertex number 6 has degree five. Intuitively, in the adjacency matrix representation, the degree of a node i is simply the sum of the elements on the i-th row $k_i = \sum_j A_{ij}$. This type of centrality awards a node uniquely on the amount of contacts. In a social network, it can give the idea of how much influence or access to information an individual has, supposed that all acquaintances are equivalent. But in real world, often this is not the case.

[1]In the following, the degree will be always labeled by letter k.

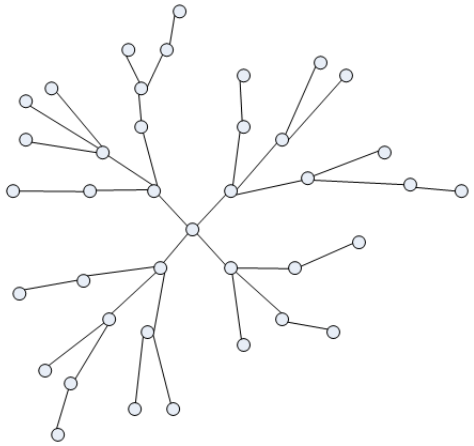


Figure 1.3: A tree-like graph.

To more effectively approach some kind of real situations, the **eigenvector centrality** can be more valuable. In social networks, for instance, being in the circle of popular people or knowing persons with anti-social tendencies can make the difference. But also popular people are more influential if their acquaintances are in turn popular. What is to be taken into account is then the popularity not only of nearest neighbours, but also of second nearest neighbours, third and so on. Mathematically, this translates in the following expression of eigenvector centrality v_i of node i:

$$v_i = \frac{1}{\Lambda} \sum_j A_{ij} v_j \tag{1.1}$$

The meaning of Λ becomes clear when grouping the centralities v_i as components of a single vector v. One obtains then that $Av = \Lambda v$, so that Λ is an eigenvalue of the adjacency matrix.
Besides as being mere topological features, centralities are fundamental quantities in dynamical processes. In particular, degree heterogeneities are the most important factor in how an epidemic spread or how well a network resists to attacks. Intuitively, a node with a large number of connections will spread an infection more efficiently than an isolated node. Recently also the eigenvector centrality has been invoked to describe the role of hubs in epidemic spreading. Moreover, global behaviour of the degree determine the critical properties in phase transitions on networks, as will be shown.
Other typologies of centrality exist, but they are not relevant to the topic of the present work.

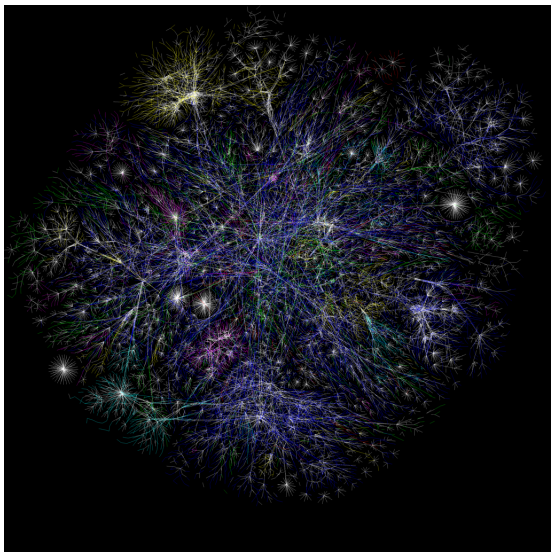


Figure 1.4: Map of the Internet, updated in 2005 (source: www.opte.org). This is how a scale-free topology looks like.

1.3 Degree distribution

In the preceding paragraph we saw that each node in a graph is characterized by a degree, which is an integer number expressing its number of connections. Expanding this idea to the whole network, one can talk about *degree distribution* $P(k)$, defined as the relative frequency of occurrence of nodes with degree k. Equivalently, it is the probability of choosing a node of degree k with random extraction. Indicating with $N(k)$ the number of nodes of degree k and with N the total number of nodes in a graph, one has therefore that $P(k) = N(k)/N$.
Many real-world networks present slowly decaying degree distributions, typically of a power-law form $P(k) \sim k^{-\gamma}$ for large k. This fact has been verified for an astonishing variety of systems, such as the Internet and WWW, geographical networks, protein and genetic networks, energy-grid networks, but also for smaller systems, such as the collaboration and citation networks in scientific publications [6].
However, the degree distribution does not establish univocally the topology of a graph. Given a set of nodes of specified degree, there may be multiple combinations for drawing links, even if the number of nodes is small. For instance, take the two graphs of Figure 1.5. They have the same degree distribution

$$P(1) = \frac{2}{5} \qquad\qquad P(2) = \frac{3}{5}$$

yet their framework is different. This fact leads to the concept of *random graph*. Here the word "graph" does not denote a single network but a statistical ensemble, whose members are all the possible choices for the assignment of edges. Every configuration has its own statistical weight and represents a realization of the ensemble. Depending on how random graphs are built, the degree sequence can be constant for all the ensemble or vary. That is, the exact degree of each single vertex can be fixed or not. In the latter case, the degree distribution $P(k)$ is the probability that a randomly chosen node has degree k, with the following definition

$$P(k) = \frac{\langle N(k) \rangle}{N} \tag{1.2}$$

Here $\langle N(k) \rangle$ is the average number of nodes of degree k, where the averaging is made over the whole statistical ensemble. We assume the total number of nodes N in each element of the ensemble to be the same, that is $N = \sum_k \langle N(k) \rangle$. Starting from these concepts, one can define the n-th moment of the degree k as

$$\langle k^n \rangle = \sum_{k'} k'^n P(k') \tag{1.3}$$

As will turn out, in mean field approaches for collective processes the first and the second moment are central quantities in the position of the critical point.

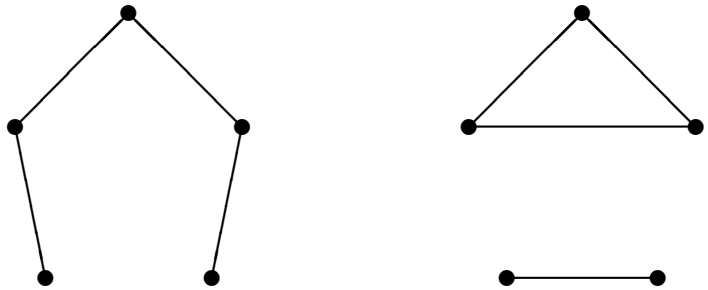


Figure 1.5: Two graphs with the same degree distribution and different edge configuration.

To conclude this quick review on complex networks, we remark that, apart from the degree distribution, a series of other features characterize graphs. A wide zoology is generated by different construction algorithms or properties of nodes and edges [1][3][2]. For instance, a weight or a direction can be assigned to each edge. In these cases, one speaks about weighted or directed graphs, respectively. In the following, we will assume to treat only unweighted undirected graphs.

Chapter 2

Epidemic spreading on networks

Epidemic dynamics is successfully modeled within the complex networks paradigm, but mathematically described only by a series of mean field approaches. Latest improvements concern the inclusion of dynamical correlations in master equations and the role of highly connected vertices.

Epidemics concern plenty of aspects in nature. First of all, infectious diseases spreading, which propagate from individual to individual mainly through physical contact or air transmission. Any sick person has a chance to infect every other person he enters in contact with, so that contagion is based on interactions among people. In this sense, a population can be seen as a network in which each individual or community corresponds to a node and edges correspond to physical interactions[1]. In the case of virus programs propagation the dynamics is similar, as well as in the case of ideas diffusion. Ideally, each system made of several components exchanging some content can be brought back to the epidemic paradigm. However, in order to study the real world it is reasonable to reduce the problem to its backbone and start from a few fundamental assumptions. Simplified populations of individuals, organized in a network configuration, are the basis for epidemic spreading.
An extremely simple model is the so called **SIS model**. The complex scenario of disease transmission, progression or regression is reduced to a double state framework. The basic rule is, so, that every node can assume two possible states: susceptible or infective. An infective node transmits the infection at a certain rate to all its nearest neighbours, i.e. the vertices with which it shares an edge. At the same time, it can spontaneously recover at

[1]To take into account the variable frequency and effectiveness of interactions, weighted edges can be used. This case is not discussed here though.

another rate. Infection and recovery rates are usually indicated with greek letters β and μ, respectively. The ratio between these two parameters is the **reproductive number** λ, a fundamental indicator of the behavior of the system. Vertices pass then from a state to the other following a stochastic chain of the form: susceptible $\rightarrow$ infective $\rightarrow$ susceptible.
A second important model is the so called *SIR model*, wherein the possible states of nodes are susceptible, infective or recovered. Within this model the recovery rate μ is the rate for an infective node to get recovered. Recovered vertices become immune to the infection and cannot return susceptible or infective again, so this class may actually include also dead individuals.
A fundamental notion in epidemiology is the **prevalence**, that is the fraction $\rho(t)$ of infected vertices at a given time. The evolution of this quantity with time is determined by the epidemic model used. Of particular interest in the SIS context is $\rho(t \rightarrow \infty)$. Below a particular value of the reproductive number the disease quickly dies out. This critical value is called **epidemic threshold**. Above, instead, the system approaches a steady state in which a finite fraction of vertices is infected. In the case of the SIR model $\rho(t \rightarrow \infty)$ is trivially null, since all vertices reach the absorbing recovered state.
The existence and the position of this critical epidemic threshold has been object of study for many years and is still a crucial topic in the comprehension of epidemics diffusion. Main results in the mathematical description of the SIS model are reviewed in this chapter.

2.1 Mean field approaches

Models like SIS and SIR are examples of compartmental models, in the sense that vertices are subdivided in classes (or compartments) depending on the stage of the disease. Classes are in these cases the set of vertices in the "S" susceptible state and set of vertices in the "I" infective state, plus the set of those in "R" recovered state within the SIR context. Additional compartments can be introduced to improve the predictive power and to better resemble real situations, where there is a higher number of possibilities. People can be immune to a disease, and yet be a mean of propagation of the virus. Moreover, after the contamination a period of latency can follow, before being contagious. Compartments can also be added according to the age of individuals, to take into account weaknesses or propensities to some infections [8]. Or again, geographic movement of individuals can be added into the picture.
In general this kind of models can be mathematically represented by a set of simple deterministic dynamical equations, which can be exactly solved at the price of some approximations. As in the most investigations of this kind, the base lies in a mean field approach, bypassing the stochastic nature of the process. In this chapter the main progresses are reviewed, from the

early times up to the last advances. Though we will focus uniquely on the SIS model, the same considerations can be extended to the SIR.

2.2 Homogeneous mean field

The simplest assumption is to consider the power of infection at a given instant of time proportional to the average number of contacts with infected individuals. Each infective vertex spreads the infection with a rate β, so the probability of a healthy vertex to become infective during a time interval dt is $1-(1-\beta dt)^n$, where the exponent n is the number of its infective neighbours. This number is estimated by noticing that, on average and neglecting fluctuations, a vertex with k connections has $k\rho$ infective neighbours, where ρ is the prevalence. Consequently, for $\beta dt \ll 1$ one obtains $1-(1-\beta dt)^{k\rho} \simeq \beta k\rho\, dt$. This reasoning is carried on a single vertex scale and is in general true for any network topology. If the vertex degree has only small fluctuations, it is possible to make the approximation $k \simeq \langle k\rangle$ for each vertex. Thus, the prevalence evolution equation is given by

$$\frac{d\rho(t)}{dt} = -\mu\rho(t) + \beta\langle k\rangle\rho(t)(1-\rho(t)) \tag{2.1}$$

The right side of the equation is made by a first spontaneous recovery term and a second contagion term.
This equation can be solved exactly in the early stage of the spreading process, when the prevalence is still very small, that is $\rho(t) \ll 1$. Thus, the term with a ρ^2 factor can be neglected, yielding the linear differential equation

$$\frac{d\rho(t)}{dt} = -\mu\rho(t) + \beta\langle k\rangle\rho(t) \tag{2.2}$$

The resolution is straightforward and leads to

$$\rho(t) = \rho_0\, e^{t/\tau} \tag{2.3}$$

where ρ_0 is the prevalence at time $t=0$ and τ is the characteristic outbreak time

$$\tau = \beta\langle k\rangle - \mu \tag{2.4}$$

This last equation introduces the concept of **epidemic threshold**: if $\tau > 0$ - i.e. $\beta\langle k\rangle > \mu$ - the number of infected individuals explodes and $\rho(t)$ increases exponentially. On the other hand, if $\tau < 0$, then the disease does not spread far and fades away. The value of the reproductive number λ marking the passage between the two regimes is [7]

$$\lambda_c^{HOM} = \frac{1}{\langle k\rangle} \tag{2.5}$$

However, this is still a simplified calculation, determined by the approximation $k \simeq \langle k \rangle$. The possibly complex architecture of the network is reduced to an homogeneous structure of identical vertices. Most of networks in nature are strongly inhomogeneous, with a few central hubs connected to the majority of the other peripheral vertices. In order to deal with such systems it is necessary then to rise the level of complexity of the equations.

2.3 Heterogeneous mean field

To obtain a more accurate analysis it is required to take into account topological heterogeneities, which means introducing an explicit dependence of variables on the degree of vertices. The first step in this direction is a degree block approximation approach: nodes are subdivided into classes labeled by their degree and elements of the same class are assumed statistically equivalent. In other words, all vertices of coordination number k are grouped into the same class and respond to the variable $\rho_k(t)$, reproducing the prevalence in the class. At each of these class variables corresponds an evolution equation, so that the dynamics of the system is described by a set of differential equations of the type

$$\frac{d\rho_k(t)}{dt} = -\mu\rho_k(t) + \beta(1-\rho_k(t))k\,\Theta_k(t) \tag{2.6}$$

Here $\Theta_k(t)$ represents the density of infected neighbours of nodes belonging to the class k:

$$\Theta_k(t) = \sum_{k'} \frac{k'-1}{k'} P(k'|k)\rho_{k'}(t) \tag{2.7}$$

The probability $P(k'|k)$ is the conditional probability that a link emanating from a node of degree k arrives to a second node of degree k'. It is usually difficult to estimate, except when doing a further approximation - i.e. neglecting degree correlations. This means assuming that the probability $P(k'|k)$ does not depend on k. It is possible to show that, in this case, $P(k'|k) = k'\,P(k')/\langle k \rangle$. Exploiting this last assumption and Equation (2.6), the evolution equation of $\Theta(t)$ can be written as

$$\frac{d\Theta(t)}{dt} = \beta\left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right)\Theta(t) \tag{2.8}$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ indicate respectively the first and the second moment of the degree distribution $P(k)$, being the n-th moment is defined according to Equation (1.3). Equation (2.8) can be easily solved and allows to earn the solution of Equation (2.6) as well. The expression giving the epidemic threshold λ_c becomes then [7]

$$\lambda_c^{HMF} = \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{2.9}$$

In networks with uniform degree distribution, this outcome coincides with that of Equation (2.5) for the homogeneous assumption. If, instead, there are heterogeneities, they weigh more on the denominator and, in general, the resulting critical λ is smaller. In the extreme cases with diverging degree second moment the epidemic threshold gets even to disappear. For instance, in scale-free networks - i.e. $P(k) \simeq k^{-\gamma}$ - with $\gamma < 3$, it can be shown that the second moment diverges with maximum degree k_{max}. This leads to a threshold scaling as $\lambda_c^{HMF} \sim k_{max}^{\gamma-3}$, which goes to 0 in the thermodynamic limit. Thus, for $\gamma < 3$ the system is always in a super-critical regime. The epidemic threshold and consequently the phase transition appears only for $\gamma > 3$, when the second moment is finite. The relevance of this finding is due to the fact that several real world scale-free networks have a degree distribution with a scaling exponent between 2 and 3 [9].
This heterogeneous mean-field theory finds confirm in numerical simulations only in annealed networks, which mean that their configuration changes more quickly than typical dynamics rate, in order to destroy degree correlations. The degree distribution is assumed fixed.
A more accurate line of reasoning is accomplished in the *quenched mean field theory*. Still neglecting dynamical correlations and fluctuations between infected neighbours, it focuses on the particular topological configuration of the network - in the specific through the spectral analysis of the adjacency matrix A [12]. The term "quenched" underlines that the particular edges configuration is fixed inside the adjacency matrix, in contrast with annealed networks. Most of real networks have actually a quenched structure, so this approach is more valuable to explore real world systems. The evolution equation is written, this time, for each vertex i of the system:

$$\frac{d\rho_i(t)}{dt} = -\mu\,\rho_i(t) + \beta(1-\rho_i(t))\sum_{j=1}^{N} A_{ij}\rho_i(t) \tag{2.10}$$

Mathematical analysis of Equation (2.10) tells us that the solution is a linear combination of exponential functions of the eigenvalues Λ_i of A. The dominant behaviour is dictated by the term of the largest eigenvalue Λ_1, i.e. $\rho(t) \sim e^{\Lambda_1 t}$. The new value of the epidemic threshold is so

$$\lambda_c^{QMF} = \frac{1}{\Lambda_1} \tag{2.11}$$

For networks made of nodes with the same degree k, the main eigenvalue is equal to $\Lambda_1 = k$ and the outcome coincides again with that of homogeneous mean field. In general, it can be demonstrated that $\lambda_c^{QMF} < \lambda_c^{HMF}$. This fact has important repercussions, in particular, on scale-free nets. For $\gamma < 3$, both approaches agree on the vanishing of the epidemic threshold, although with a different scaling in the interval $2,5 < \gamma < 3$. But when $\gamma > 3$, the quenched mean field approach still predicts a vanishing threshold in the

thermodynamic limit, even if the network maintains a finite $\langle k^2 \rangle$. Furthermore, the exact value of Λ_1 has been calculated for several real networks, obtaining strongly different results in the corresponding epidemic threshold with respect to the heterogeneous mean field [16].
An alternative strategy to locate the epidemic threshold is to study what happens for $t \to \infty$. Indeed, one expects to find a zero prevalence below the epidemic threshold and a positive prevalence above. Setting to zero the derivative in Equation (2.10), one obtains the following expression for $\rho_i = \rho_i(\infty)$:

$$\rho_i = \frac{\beta \sum_j A_{ij}\, \rho_j}{\mu + \beta \sum_j A_{ij}\, \rho_j} \tag{2.12}$$

This equation has a trivial solution $\rho = 0$ and a non-null solution $\rho_i > 0$ over a certain value of λ, which marks the epidemic threshold. The outcome so obtained coincides with the result in Equation (2.11).

2.3.1 Pair-Quenched Heterogeneous mean field

Previous analysis still do not take into account dynamical correlations between neighbouring nodes. Only recently, this difficulty was partially overcome by means of a perturbative technique, yielding new powerful results [14].
Let us then introduce a further notation, wherein $[A_i]$ is the probability that the vertex i is the state A. In the same way, $[A_i, B_j]$ is the probability for vertices i and j to be respectively in the states A and B. So on, this idea extends to groups of indefinitely many nodes. Consistently with the previous notation, we have that $\rho_i = [1_i]$ and, therefore, $[0_i] = 1 - \rho_i$. Additional variables can now be presented:

$$\psi_{ij} = [1_i, 1_j], \qquad \omega_{ij} = [0_i, 0_j] \tag{2.13}$$

$$\phi_{ij} = [0_i, 1_j] \qquad \bar{\phi_{ij}} = [1_i, 0_j] \tag{2.14}$$

The single-vertex evolution equation takes now the form

$$\frac{d\rho_i}{dt} = -\mu\, \rho_i + \beta \sum_j \phi_{ij} A_{ij} \tag{2.15}$$

where the dependence on time was dropped. Within this notation it is possible to add an equation to the problem, which is the dynamical equation for a pair of connected vertices (i, j) in the state (0,1)

$$\frac{d\phi_{ij}}{dt} = -\mu\, \phi_{ij} - \beta\, \phi_{ij} + \mu\, \psi_{ij} + \beta \sum_{\substack{l \in \mathcal{N}(l), \\ l \neq i}} [0_i 0_j 1_l] - \beta \sum_{\substack{l \in \mathcal{N}(l), \\ l \neq j}} [1_l 0_i 1_j] \tag{2.16}$$

where $\mathcal{N}(i)$ is the neighbourhood of vertex i. First three terms represent the possible events within the pair that permit to leave or get to the state

(0,1) - i.e. vertex j spontaneously recovers, j infects i, or i recovers. The two summations consider the interaction of i and j with their other neighbours. Equations (2.15) and (2.16) were solved thanks to the one vertex approximation $\phi_{ij} \approx \rho_i(1-\rho_i)$ and to the following pair approximation:

$$[A_i, B_j, C_l] \approx \frac{[A_i, B_j][B_j, C_l]}{[B_j]} \tag{2.17}$$

Performing a linear stability analysis around $\rho_i = \phi_{ij} = \psi_{ij} = 0$ an analytical solution was found in some simple cases, namely the random regular graph, the star graph and the wheel graph. The most relevant result for the scopes of this project is for the random regular network, that is

$$\lambda_c^{PQMF} = \frac{1}{k-1} \tag{2.18}$$

Numerical simulations confirm this predicted threshold.

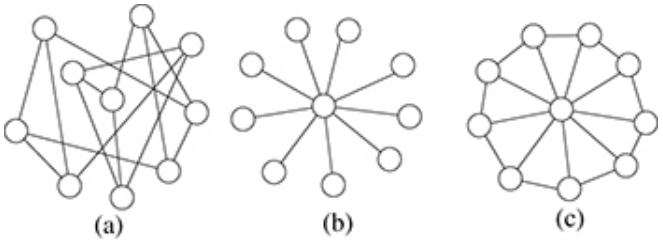


Figure 2.1: Three types of network: (a) Random regular graph. (b) Star graph. (c) Wheel graph.

2.4 Localization transition in networks

Recent efforts in SIS model research concern the role of eigenvector centrality in epidemic spreading[2]. In particular, it has been shown that important consequences might arise depending on the eigenvector centrality of the most connected vertices. Indeed, the leading eigenvector of the adjacency matrix can undergo a *localization transition* [15]. This implies that most of the weight of the vector concentrate around a few components, leaving a negligible weight on the others. If this happens the eigenstate is said to be **localized**. The weight of eigenvector components is tied to the entity of entries in the adjacency matrix - namely to the number of edges of corresponding nodes - as one can see from Equation (1.1). Hubs - i.e. nodes with high connectivity - are thus the responsible for the localization transition. If, instead, the network has weak heterogeneities, all nodes have similar eigenvector centrality and all entries of the eigenvector are of the same order of

[2]See Paragraph 1.2.

magnitude. In this case the eigenstate is said to be **delocalized**.
It has been demonstrated that this fact might break the traditional double-regime picture of the SIS dynamics [16]. In fact, if Λ_1 corresponds to a localized eigenstate then, in the upper neighbourhood $\lambda_c^{QMF} = 1/\Lambda_1$, the system can reach a stationary state wherein the contagion is restricted within a finite number of vertices. In the infinite graph limit, this number constitutes a negligible fraction of the total number of vertices and vanishes. With further increase of λ more and more nodes get permanently infected until they become a finite fraction of the total. On the other hand, if Λ_1 corresponds to a delocalized eigenstate, right above λ_c^{QMF} the disease infects a finite fraction of vertices and the usual transition to the active state is preserved.
The demonstration exploits again the spectral analysis of the adjacency matrix A. Probabilities $\rho_i = \rho_i(\infty)$, taken from the steady state Equation (2.12), can be written as a linear superposition of components of the eigenstate $v_i(\Lambda)$

$$\rho_i = \sum_{\Lambda} c(\Lambda)\, v_i(\Lambda) \tag{2.19}$$

where coefficients $c(\Lambda)$ are the projections of the vector ρ on eigenstate $v(\Lambda)$. Setting $\mu = 1$, so that without loss of generality $\beta = \lambda$, and substituting the last equation into Equation (2.12), one obtains

$$c(\Lambda) = \lambda \sum_{\Lambda'} \Lambda'\, c(\Lambda') \sum_{i=1}^{N} \frac{v_i(\Lambda)\, v_i(\Lambda')}{1 + \lambda \sum_{\Lambda''} \Lambda''\, c(\Lambda'')\, v_i(\Lambda'')} \tag{2.20}$$

Solving this equation with respect to $c(\Lambda_1)$ yields $\lambda_c = 1/\Lambda_1$, returning the result of quenched mean field. Near the epidemic threshold, only the principal eigenvector counts and $\rho_i \approx c(\Lambda_1)\, v_i(\Lambda_1)$. At $\lambda \geq \lambda_c$, expanding $\rho = \sum_i \rho_i$ as a function of $\epsilon = (\lambda - \lambda_c)/\lambda_c$ one obtains

$$\rho \approx \epsilon\, \frac{\sum_{i=1}^{N} v_i(\Lambda_1)}{N \sum_{i=1}^{N} v_i^3(\Lambda_1)} \tag{2.21}$$

This expression gives the prevalence right above the epidemic threshold as a function of the components of the leading eigenvector. The parameter establishing whether it is a localized eigenvector is the inverse participation ratio:

$$IPR(\Lambda) = \sum_{i=1}^{N} v_i^4(\Lambda) \tag{2.22}$$

If, in the limit $N \to \infty$, $IPR(\Lambda)$ is of the order of $O(1)$, then the eigenvector $v(\Lambda)$ is localized. This because for a localized $v(\Lambda)$, only a few components are of the order $v_i(\Lambda) \sim O(1)$, while the rest are negligible. If the main eigenvector is localized, the fractionary factor in Equation (2.21) becomes of the

order of $O(1/N)$, implying that $\rho \sim O(1/N)$. The stationary infected state in the upper neighbourhood of λ_c is then limited to a ρN number of vertices. Conversely, if a generic $v(\Lambda)$ is delocalized usually one has $v_i(\Lambda) \sim O(1/\sqrt{N})$ for all $v_i(\Lambda)$ and therefore $IPR(\Lambda) \to 0$. If this is the case for $v(\Lambda_1)$ one obtains $\rho \sim O(1)$. A finite fraction of vertices are then trapped into the active stationary state and the traditional phase transition is preserved.

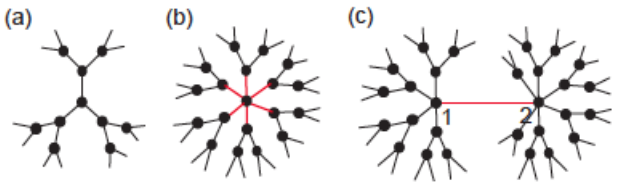


Figure 2.2: (a) Regular Bethe lattice with degree $k = 3$. (b) Bethe lattice with one hub of degree $q > k$. (c) Bethe lattice with two connected hubs.

Localization phenomena has been verified for scale-free networks, wherein the localization of the main eigenvalue appears to be bound to the maximum degree k_{max} and occurs if it exceeds a certain threshold k_{loc}. This happens if the degree distribution is slowly decaying, namely if $\gamma > 5/2$. In real networks known to have a scale-free topology, the estimation of $IPR(\Lambda_1)$ led to the conclusion that the localization transition does not take place for all cases considered and this was imputed to the fact that $k_{max} < k_{loc}$ [16]. Moreover, a simple analytical condition for the localization transition has been found in the case of a *Bethe lattice*, which is an infinite graph with a tree-like structure and vertices of constant degree (see Picture 2.2). The adjacency matrix for such a graph with vertices of degree k has all eigenvalues equal to $\Lambda_1 = k$, corresponding to delocalized eigenvectors with $v_i(\Lambda_1) = N^{-1/2}$. Introducing a hub of degree $q > k$, one instead finds

$$\Lambda_1 = \frac{q}{\sqrt{q - k + 1}} \tag{2.23}$$

and entries of the leading eigenvector have maximum value in correspondence of the hub and exponentially decrease with increasing distance. Moreover, if $q \gg k$ the inverse participation ratio tends to $IPR(\Lambda_1) \to (1 + 1/q)$, so the corresponding eigenstate is localized. With two connected hubs an analogous solution was found. Note that Equation (2.23) returns the same threshold of the homogeneous mean field approach in the completely homogeneous case $q = k$.
However, a satisfying numerical confirmation on this issue is missing and up to now it was not clear what are the implications of the localization transition in the SIS dynamics. This is the central point of this project and will be faced in the final chapter.

Chapter 3

Phase transition in the SIS model

Statistical mechanics is essential to comprehend collective phenomena on complex networks. In particular, current efforts focus on non-equilibrium processes characterized by absorbing phase transitions - like epidemic spreading. Universal relationships bind the SIS model and the percolation process.

In this chapter we embrace a more general perspective to present the properties of the SIS model seen as a stochastic process. We start by recalling that the spectra of the reproductive number splits into two distinct regions - or phases[1]: the first phase has a rapidly collapsing prevalence, while the second phase has an ever-lasting non-null prevalence. The boundary region between these regimes is what in statistical physics is called a **phase transition**. More in general, a phase transition is the passage between different regimes in systems governed by stochastic rules. Such phenomena were historically first observed in thermodynamical fluid systems and so they are usually described in terms of statistical mechanical quantities. Within the modern classification, they are discerned into two main categories. In *first order phase transitions*, the two phases are separated by a coexistence line, on which they live together in a mixed-phase regime. Systems crossing this line accomplish the transition part by part and thermodynamic quantities present a discontinuous behaviour. *Second order phase transitions* are instead characterized by a smooth uniform variation in the behaviour of the system and do not present a coexistence line. For these reasons, they are also called *continuous phase transitions*. This latter typology is induced by long range correlations that span the whole size of the system. In systems

[1] See Chapter 2.

made of many interacting agents, this phenomenon translates in a *collective behaviour* in the microscopic degrees of freedom.
Research on phase transitions started on equilibrium systems, typically many-particle systems in contact with a thermal reservoir that controls the temperature. However, much of the knowledge on equilibrium transitions can be extended to non-equilibrium ones, and particularly to **absorbing phase transitions**. These can be found when the microscopic dynamics is irreversible and manifest as continuous transitions from fluctuating stationary states into absorbing states. The essential difference is that here time plays the role of independent degree of freedom.
In the last few years, an intense effort was put in extensively investigate phase transitions on networks. Examples are the equilibrium processes such as the Ising model [3] [17] [7], but also those of non-equilibrium like epidemic models [3] [17] [7], percolation [3] [17] [7], synchronization [3] [17] [7] and many others [3] [17] [7]. A motivation for investigating phase transitions on networks is the theoretical interest in understanding how the topology of the substrate affects critical phenomena, traditionally studied on regular $\mathbb{Z}^d$ spaces. Indeed, a network with N vertices is equivalent to a space of dimension $N-1$, since in principle a vertex can be connected to all other vertices. Depending on the degree distribution, networks can thus work as strongly irregular spaces, even infinite dimensional in the thermodynamic limit. On the other hand, there are also more practical repercussions: for example, the Ising model framework may represent not only magnetic systems, but also other entities, such as social systems wherein people are reciprocally influenced. Unraveling the complexity of a model can then return new implications on more fronts.
Intuitively, the SIS phase transition belongs to the category of non-equilibrium absorbing phase transitions. As soon as the systems reaches the condition of being fully healthy, it becomes trapped in a completely inactive state forever. In the following, features of this transition are presented, briefly retracing main exemplifying models.

3.1 Ising model

The Ising model is a paradigmatic case of study in the field of phase transitions and critical phenomena. It is the precursor of a variety of models useful in physics as well as in other fields, like epidemiology. Originally, it was created to describe ferromagnetic systems. Indeed, in materials such as iron, nickel or cobalt, the dipole orientation of each atom is temperature sensitive and, at the same time, neighbouring atom dipoles tend to align in the same direction to minimize interaction energy.. When temperature is high, all atoms change orientation very quickly and the average magnetization is null. When, on the other hand, the temperature is lowered under a

certain value, spins in the material freeze and a nonzero net field appears. The passage from an ordered regime with a preferential axis and a disordered one represents a spontaneous symmetry breaking.

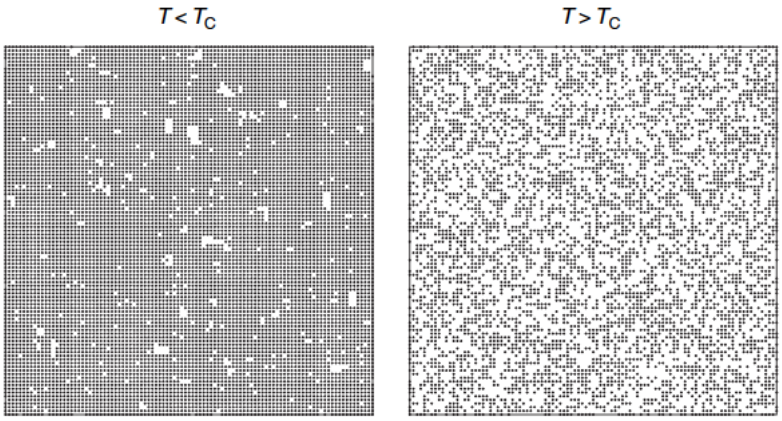


Figure 3.1: On the left, subcritical regime of the Ising model, where spins freeze trying to align with the neighbours. On the right, overcritical regime, where spins continuously change orientation at random.

The set up of the Ising model is the following: imagine to have a regular pattern made of all equal cells[2], filling an N-dimensional space to the infinite. This is what is called a *lattice*. Now suppose every site on the lattice contains a particle, in order to obtain a crystalline structure. Particles are polarizable and have a spin which can freely rotate in space. In principle spins can be oriented in any direction, but we assume that the admissible values of the spin for each particle i are only $\sigma_i = \pm 1$. This corresponds to the simplified idea that polarization exists only in an "up or down" direction[3]. Any spin conveys the orientation of other spins, so that it tends to assume the same orientation of neighbour sites. Plus, spins randomly switch orientation with a frequency proportional to the temperature T. In presence of an external field H, the total energy of the system is given by the Hamiltonian

$$\mathcal{H}_{\text{Ising}} = -\sum_{\{i,j\}} J_{ij}\sigma_i\sigma_j - \sum_i H_i\sigma_i \tag{3.1}$$

The first sum runs over all couples of neighbour spins i and j. The factor J_{ij} expresses the coupling energy of a pair of confining spins and H_i represents the local field on atom i. In the simplest case $J_{ij} = J$ and $H_i = H$. This system can pass from any global configuration $\{\sigma_i\}$ to any other, so

[2]Usually, cells are called sites.

[3]Other models, as the XY model, consider spins assuming continuous values between -1 and $+1$.

the dynamics is reversible and there is equilibrium. When the external field H is turned off a second order phase transition separates the disordered regime at large T and zero overall magnetization $M = \sum_i \sigma_i$ from the ordered regime at small T and non-null M. Consequently, one can study the phase transition on the basis of the behaviour of M, which is then the *order parameter* of the transition.
In general, equilibrium phase transitions are described in terms of order parameters associated to conjugated fields, imposed on the system by external conditions. The response in the order parameter to variations of the external field is given by the susceptibility χ, defined as the derivative of the order parameter with respect to the field. In the considered case, we have that the total magnetization M is associated to the external field H. The **susceptibility** is then defined as

$$\chi(T, H \to 0^+) = \frac{\partial M}{\partial H}\Big|_{H=0^+} \tag{3.2}$$

This variable has the peculiar beahaviour to diverge at the particular combination of parameters which marks the phase transition, called **critical point**. For this reason, it is a useful tool to investigate the specific conditions that trigger the transition, as will be shown later on.
In recent times the Ising model was applied also on networks [3] [17] [7]. Here the pairs of spins in (3.1) are not the spatial confining sites, but nearest neighbours in the graph configuration. The Hamiltonian acquires thus the form

$$\mathcal{H}_{\text{Ising}} = -\sum_{i<j} J_{ij} A_{ij} \sigma_i \sigma_j - \sum_i H_i \sigma_i \tag{3.3}$$

where A_{ij} is the entry of the adjacency matrix corresponding to the pair i and j. As was shown in previous chapters, the topology of graphs can vary considerably and Ising on graphs is solvable in a mean field approximation. Exactly like in the SIS heterogenous mean field[4], we can partition the vertices in classes of the same degree. The criticality condition was calculated for general uncorrelated networks and the critical temperature obtained is given by

$$T_c = 2J / \ln\left(\frac{\langle k^2 \rangle}{\langle k^2 \rangle - 2\langle k \rangle}\right) \tag{3.4}$$

It is important to notice that, even in equilibrium processes like this, the value of the critical point is influenced by connectivity properties of the graph. In particular, in networks with a heavy-tailed degree distribution, the second moment of k diverges in the thermodynamic limit $N \to \infty$, and the critical temperature goes to infinity. Consequently, in scale-free networks the phase transition disappears, just like in the SIS mean field.

[4] see Paragraph 2.3.

3.2 Critical scaling

In the modern era of critical phenomena, the research focused more and more on the determination of sets of numbers called **critical exponents**, which are deeply bound to the nature of the system considered. Let us now show how.
Let us assume the external field to be given by the temperature T, like for the Ising model. In general, around the critical point, a function $f(T)$ behaves as $f(\epsilon)$, where

$$\epsilon \equiv \frac{T - T_c}{T_c} \tag{3.5}$$

is a dimensionless coordinate expressing the distance from the critical temperature. We assume that $f(\epsilon)$ is a positive and continuous function for sufficiently positive and small values of ϵ and that the limit

$$\alpha \equiv \lim_{\epsilon \to 0} \frac{\ln f(\epsilon)}{\ln \epsilon} \tag{3.6}$$

exists. Close to the critical point it is then true that $f(\epsilon) \sim e^{\alpha}$. The number α takes the name of critical exponent of $f(\epsilon)$.
A ferromagnetic system described by the Ising model presents critical scaling for many quantities. For example, the total magnetization scales as $M \sim |\epsilon|^{\beta}$ and the susceptibility scales as $\chi \sim |\epsilon|^{-\gamma}$. However, all critical exponents can be related to the critical exponent of the *correlation length* ξ, which defines the typical size of domains of aligned spins. For $T \to T_c$, the correlation length diverges according to the law $\xi \sim |\epsilon|^{-\nu}$. This means that the systems becomes more and more "aware" of itself approaching the critical point. At criticality, domains of all sizes appear on the lattice with a frequency that decreases as a power-law with growing size and the system is invariant under suitable scaling transformations. This phenomenon is known under the name of *scale invariance*.
However, the power-law form is just an approximation and contains less information than the complete form of the function. So why study critical exponents? There are two main reasons for which this is important. First, there is the experimental fact that, close enough to the critical point, the power-law behaviour dominates. This means that critical exponents are always measurable, while the entire functions may not be. Linear fitting of experimental data in log-log plots are usually used at this purpose.
A second reason is that critical exponents are bound with each other by relationships arising from general statistical reasonings, which transcend any peculiar system. The discovery that different systems might share the same critical scaling lead to the concept of **universality class**. Universality classes are the groups of systems with an equal set of critical exponents. SIS model belongs to the directed percolation universality class, that is presented in Paragraph 3.4.

3.3 Finite-size scaling

So far, we tacitly treated phase transitions only in the thermodynamic limit. However, boundaries have relevant effects on the critical point, emerging more and more as the system size diminishes. For instance, consider the susceptibility χ_L of the 2D Ising model on an infinitely long slab of finite width L. The susceptibility does not diverge as $\chi_L \sim |\epsilon|^{-\gamma}$, but grows up to a maximum at some pseudo-critical temperature $T_c(L)$. If the system is sufficiently large it is expected that the distance from the real critical temperature scales with a certain exponent λ',

$$\frac{T_c(L) - T_c(\infty)}{T_c(\infty)} \sim L^{-\lambda'} \qquad (\text{if } L \to \infty) \tag{3.7}$$

which is called *shift exponent*[5] [17]. A second outcome is the broadening of the curves for finite L. It is described in terms of a *rounding temperature* $T^*(L)$, defined such that if $(T - T_c)/T_c \geq (T^* - T_c)/T_c$, then $\xi_L(T) \simeq \xi_\infty(T)$. In analogous way as before,

$$\frac{T^*(L) - T_c(\infty)}{T_c(\infty)} \sim L^{-\theta} \qquad (\text{if } L \to \infty) \tag{3.8}$$

where θ is the *rounding exponent*. These considerations influence the scaling of all variables. The fundamental hypothesis of the finite-size scaling asserts that these two exponents are equal.
The concept of phase transition is then defined only in the thermodynamic limit. However, real systems have finite size, as well as systems modeled in numerical simulations, hence it is fundamental to comprehend the role of finite size effects. Moreover, in finite size systems the order parameter is a fluctuating quantity and fluctuations have maximum intensity at the critical point. It is then convenient to define the variance of the magnetization per unit volume [18]

$$\chi' = \lim_{L \to \infty} L^d (\langle M^2 \rangle - \langle M \rangle^2) \tag{3.9}$$

where the brackets $\langle \dots \rangle$ denote the temporal average, L is the lateral size of the system and d its dimensionality. In finite size systems - such as those of computational use - and in the absence of a symmetry-breaking field, this quantity scales with an exponents equal to γ for $T < T_c$, so that it can be measured in the place of the real susceptibility. This will turn very important in the final chapter.

3.4 Percolation

The Ising model phase transition belongs to the more general category of equilibrium phase transitions, in the sense that the microscopic evolution of

[5] The $'$ superscript aims to prevent confusion with the reproductive number λ.

the system is reversible. Processes such as the SIS model manifest instead a non-equilibrium phase transition, wherein an absorbing state compromises the microscopic reversibility. These latter processes inherit all critical features from the formers, yet with some differences. A paradigmatic group of models is in this case that of percolation models [19][20][25].
The term percolation derives from Latin and means "filter gradually through a porous surface or a substance". In science, it is traditionally adopted as a geometrical model for the study of the permeability through random media. It can be applied first of all on the filtration of solutions in solid materials, but also on the spreading of wildfires in forests. However, percolation has also a dynamical interpretation. Indeed, it has been widely studied through simple models which try to imitate the microscopic penetration dynamics in a porous filter, just like the Ising model tries to imitate the microscopic dynamics in magnetic materials. The filter is conceptually represented by a lattice, whose sites are "open" and "closed" if they correspond, respectively, to an empty interstice or to an impenetrable cell. The medium is made irregular by asserting that every site is open with probability p - the *percolation probability* - and closed with probability $1 - p$. The dynamics can be formulated as a stochastic process driven by several choices of rules. Usually it starts from a single site, the seed, and proceeds along an axis representing a spatial or temporal dimension, as in Picture 3.2. For ease of visualization it is drawn as a waterfall subjected to the gravity force field. At each step the water can flow on open sites of the underlying layer. The phase transition occurs at a precise threshold of the percolation probability p_c. At low p, only open paths of finite size exist, but over p_c the lattice can become macroscopically permeable.
Moreover, percolation exists in two main forms: in the undirected or isotropic percolation, the flowing is allowed in all directions, while in the directed percolation the agent is forced to flow in a preferred direction of space. Both versions of the model display a phase transition, yet they have diverse critical parameters and universal properties. *Directed percolation* constitutes the landmark universality class for non-equilibrium phase transitions into absorbing states. Its importance is comparable to that of the Ising model for equilibrium phase transitions. In this kind of systems, the steady state is characterized by a constant density of active sites $\rho_s > 0$ in the thermodynamic limit. For finite size systems, however, ρ_s is a fluctuating quantity. Analogously to the equilibrium case, one can measure the magnitude of ρ fluctuations - which is so adopted as order parameter - and obtain information on the critical point. Here one therefore uses the following definition:

$$\chi' = \lim_{N\to\infty} L^d(\langle\rho^2\rangle - \langle\rho\rangle^2) \tag{3.10}$$

In the non-equilibrium case, this quantity scales with an exponent $\gamma' \neq \gamma$, though.

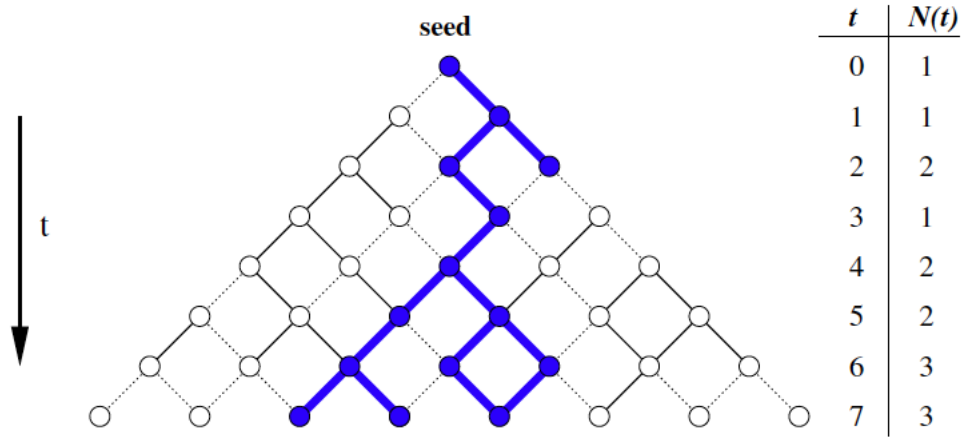


Figure 3.2: Visualization of directed percolation on a square lattice. In the table on the right, $N(t)$ represents the number of open sites found at layer t by the percolated water.

After the advent of network science, percolation was also exploited to deal with resistance and resilience of network structures, traffic and congestion in a roadway tissue and even epidemic spreading. The most considered case is, however, the resistance to removal or failure of nodes or links[6]. Indeed, in some contexts the solidity of the structure of a network is of crucial importance. Take, for example, the case of power grids or transport networks. In case an element of the net damages or fails to do its job it is necessary to have an efficient reaction strategy. This requires the knowledge of which are more important at their specific function.
In this case, the phase transition lies between a regime with only finite disconnected sub-graphs and a regime wherein a giant component of connected vertices appears. The order parameter is the probability P_G for a node to belong to a giant cluster. The complementary probability $1 - P_G$ is the probability, for a node of degree k, that none of its edges lead to the giant component. This definition translates into the following expression [7] [17]

$$P_G = 1 - \sum_k P(k) q^k \tag{3.11}$$

where we define q as the probability that a randomly chosen edge does not lead to a vertex of the giant cluster. It is possible to write the following self-consistent equation for q:

$$q = \sum_k \frac{kP(k)}{\langle k \rangle} q^{k-1} \tag{3.12}$$

[6] As in the lattice version, it is possible to distinguish between node or edge percolation.

The condition for the existence of a giant cluster is determined by searching a non-trivial solution for Equation 3.11. In uncorrelated networks, the critical value for the probability p is found to be

$$p_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{3.13}$$

In particular, for regular graphs this expression is equivalent to $p_c = 1/(k-1)$.
In network percolation, the role of the susceptibility is played by the average size of non-giant clusters $\langle s \rangle$. Critical scaling affects order parameter and relative response response function as

$$P_G \sim (p - p_c)^{\beta} \tag{3.14}$$

$$\langle s \rangle \sim |p - p_c|^{-\gamma} \tag{3.15}$$

The use of γ as exponent of $\langle s \rangle$ stresses its function, equivalent to susceptibility in the lattice case.
The main difference between equilibrium and non-equilibrium critical phenomena - such as for percolation - is in the correlation length. In the first category, continuous phase transitions are characterized by a single correlation length ξ that diverges at the critical point as $\xi \sim |\epsilon|^{-\nu}$. Besides, in the second case time acts as additional degree of freedom, implying that there are two different correlation lengths, $\xi_\perp$ and $\xi_\parallel$. Subscripts $\perp$ and $\parallel$ denote, respectively, spatial and temporal properties. Approaching the critical point, these quantities diverge as

$$\xi_\perp \sim |\epsilon|^{-\nu_\perp} \qquad\qquad \xi_\parallel \sim |\epsilon|^{-\nu_\parallel} \tag{3.16}$$

with exponents that are generally different. Critical phenomena of non-equilibrium can all be related to these two correlation lengths, as happens for the contact process and the SIS model.

3.5 Contact process

The process that is most strictly bound to the SIS model is the **contact process**. This latter is a stochastic model for interacting particle systems, traditionally applied on reaction kinetics. It is the simplest version of a category of models aimed to mathematically describe how the concentration of one or more substances changes because of the effect of local chemical reaction and spread in space. Such models are called *reaction-diffusion processes*. Later on, the contact process was exported also on networks environment to investigate epidemic spreading. Within its framework, vertices are sites that can be occupied by a particle or be empty. Particles can self-annihilate with a rate μ and generate new offspring particles with a rate λ/k in each

neighbour site - with k the degree of the generating particle. Usually, the annihilation rate is set at $\mu = 1$ - choice that fixes the time scale without loss of generality. The similarity with the SIS dynamics is evident, though the terminology is different. The only gap is that in contact process the strength of spreading is equally subdivided within the neighbours of occupied vertices, and so depends on their degree. In the SIS, sick vertices infect with a constant rate, independently on connectivity heterogeneities. The two models coincide only when applied on lattices or networks of constant degree distribution. To give a proof of this fact, we briefly present a theoretical approach for the determination of the critical point in contact process. Just like for the SIS model, the mathematical survey has been attempted by means of a heterogeneous mean field approach. Similarly to what was presented in Paragraph 2.3, the master equation for the class of nodes of degree k reads as

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda(1-\rho_k(t))k\sum_{k'}\frac{P(k'|k)\,\rho_{k'}}{k'} \tag{3.17}$$

where we set $\mu = 1$. Despite its formulation, this approach fails in taking into account degree fluctuations and predicts without distinctions $\lambda_c = 1$. In the attempt of fixing this issue, a heterogeneous pair-approximation approach was used [27]. The symbolism is analogous to that of Paragraph 2.3.1, with the only exception that subscripts refer to the degree of node compartments and not to the degree of single nodes. The set of single-vertex density evolution equation becomes

$$\frac{d\rho_k}{dt} = -\rho_k + \lambda\, k\sum_{k'}\frac{\phi_{kk'}\,P(k'|k)}{k'} \tag{3.18}$$

and the evolution equation for $\phi_{kk'} = [0_k, 1_{k'}]$ is

$$\begin{aligned}\frac{d\phi_{kk'}}{dt} = -\phi_{kk'} - \lambda\frac{\phi_{kk'}}{k'} + \psi_{kk'} + \lambda(k'-1)\sum_{k''}\frac{[0_k 0_{k'} 1_{k''}]\,P(k''|k')}{k''} \\ - \lambda(k-1)\sum_{k''}\frac{[1_{k''} 0_k 1_{k'}]\,P(k''|k)}{k''}\end{aligned} \tag{3.19}$$

Performing a quasi-static approximation - $d\rho_k/dt \approx 0$, $d\chi_{kk'}/dt \approx 0$, in the limit $t \to \infty$ - it is possible to get to a general condition equation for criticality in uncorrelated networks. It was analytically solved for random regular networks, i.e. with degree distribution $P(k) = \delta_{km}$. The solution is

$$\lambda_c = \frac{m}{m-1} \tag{3.20}$$

and achieves to get very close to numerical simulation results. The result is then the same as for the SIS model, since at criticality the effective creation

rate through each link is rescaled of a factor m.
Contact process undergoes a non-equilibrium phase transition whose order parameter is the particle density ρ_∞ in the stationary state. The subcritical phase has rapidly decaying $\rho(t)$ into an absorbing state, with null particle density. The overcritical phase arises at a certain value of λ and is an active state, with constant average particle density. In this context, the susceptibility is defined as the answer in the order parameter to a variation of an external field h, which can be interpreted as a source of spontaneous creation of particles. However, more often the alternative definition is used for computational purposes, that is

$$\chi = N(\langle \rho^2 \rangle - \langle \rho \rangle^2) \tag{3.21}$$

This is variance of the particle density multiplied by the size N of the graph, namely a measure of stochastic fluctuations in the density of particles.
In $d-$dimensional lattices, a full set of critical exponents was calculated. Defining $\Delta = \lambda - \lambda_c$, the system enters in the active phase according to $\rho \sim \Delta^\beta$. Close to the critical point, the correlation length and time diverge scaling as $\xi_\perp \sim |\Delta|^{-\nu_\perp}$ and $\xi_\parallel \sim |\Delta|^{-\nu_\parallel}$. Other critical exponents exist, but they can all be expressed in terms of β, $\nu_\perp$ and $\nu_\parallel$ by scaling and hyperscaling relations [20][22].
Analogous considerations are valid for the SIS model. In fact, both the contact process and the SIS model belong to the directed percolation universality class in homogeneous lattices. In complex networks, however, the topologic heterogeneities dramatically affect both models. At a heterogeneous mean field level, they show different critical exponents.

Chapter 4

SIS model on networks

In networks science, computer simulation assumes a key role in the confirmation of theoretical predictions. Probabilistic operative methods and algorithms are required to fulfill this task. To reproduce networks with precise connectivity properties, the configuration method is the simplest technique.

In this chapter are presented methods and algorithms used to face the issues introduced in the prelude. The SIS model is a stochastic process and consequently has to be investigated by means of a probabilistic approach. For dynamical processes acting on complex networks, this does not only mean to perform a large number of repetitions of the process and extract average values. To get some insight of general validity, it is required also to equip of all possible graphs with the requested degree distribution. In other words, we need a random graph[1].
In continuity with the the study of Dorogovtsev et al. [16], the type of network here considered is the **random regular graph**, a network with equal degree on all the randomly linked vertices - i.e. having degree distribution $P(k) = \delta_{km}$, with m constant. Taking any node as central, the structure of such a graph can be seen as a tree-like with long random loops added. In the large size limit, loops in short distances are so rare that, in any point, the graph is locally tree-like and can be seen as a Bethe lattice [23]. The trivial degree distribution makes this kind of network easy to study and, if a hub is introduced, its effects can be clearly observed. Plus, it is important for graphs to be simple, that is they must not contain self connected nodes and multiple links. Operative methods are presented in the following.

[1] See Paragraph 1.3

4.1 The configuration method

A series of models for the generation of random graphs have been developed. The majority of them are able to work only for specific degree distributions. Above all, the *configuration method* - or *pairing method* - is able to produce graphs of whatsoever degree distribution [1]. Actually, this model allows to create a network starting from any arbitrary degree sequence. That is, the starting point is the exact set of degrees of every node. This, in turn, fixes the number of edges L by means of the constraint $\sum_i k_i = 2L$.
Consider a set of vertices of specified degree k_i of arbitrary choice. To each of them is assigned a number k_i of edge stubs. In total, there are initially $2L$ stubs. The procedure consists in picking a pair of stubs uniformly at random and joining them together, in order to create a complete edge. After that, $2L - 2$ free stubs are present. The operation is iterated until no free stubs are present and all the L links are built. This final matching corresponds to a network in which each vertex has exactly the desired degree.
Clearly, the graph so obtained is just one possible configuration. Choosing each single stub pairing with equal probability implies that each final matching has equal probability to be produced. One can then see how the effective overall result is the random graph in which every configuration of specified degree sequence appears with the same probability.

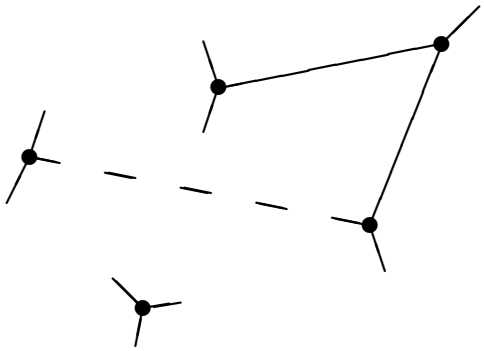


Figure 4.1: Configuration method for random regular graph of degree distribution $P(k) = \delta_{k3}$. The dashed line represents the union process between two stubs.

However, there are some issues that make this method imperfect. In principle, the pairing can take place also between stubs of the same vertex, or between stubs of vertices already connected. It means that a fraction of the ensemble generated is made of non simple networks, containing self-loops and multiedges. Indeed, it was calculated that, for random regular graphs,

simple configurations appear with probability proportional to $e^{-k^2/4}$, at least for relatively small k [29].
A second fact is that different final matchings can be mapped in the same network, simply by permuting the stubs in every possible way. The number of permutations for the k_i stubs in a node i is $k_i!$ and consequently the number of final matchings corresponding to a given graph is $M(\{k_i\}) = \Pi_i\, k_i!$. This number is the same for all graphs, since every degree k_i is fixed. Indicating with $\Omega(\{k_i\})$ the number of all possible matchings, each network appears with constant probability M/Ω. Yet, this property is valid only for simple networks. Admitting self-loops and multiedges, it is no longer true that any permutation of the stubs leads to a new matching. This fact becomes clear looking at Figure 4.2. Furthermore, in the context of dynamical processes running on networks - such as the SIS model - impurities like multiedges and self-loops may alter the dynamics.
Though the density of impurities goes to zero in the thermodynamic limit, we will work mostly at small network sizes. For this reason, we need to ensure to have only simple graphs.

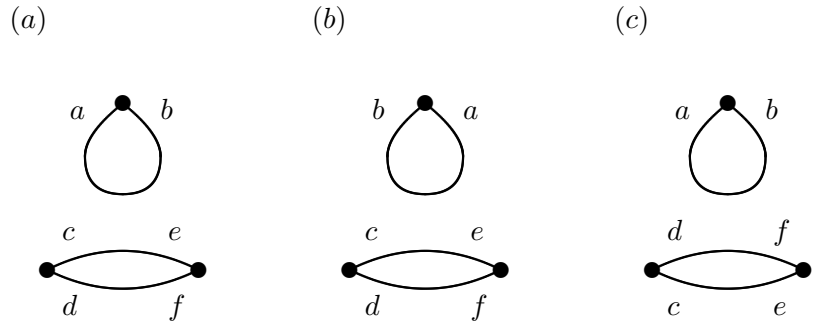


Figure 4.2: Permutation of stubs in self-loops and multiedges. Exchanging stubs a and b the corresponding edge remains the same. Switching members both in the couple $c - d$ and in the couple $e - f$ returns the same two edges.

4.2 The Steger-Wormald method

An improvement of the configuration method was given by A. Steger and N.C. Wormald [29]. The new algorithm guarantees of generating only simple graphs and, at the same time, maintains the runtime low. Though their formulation is specifically for random regular graphs, it is here presented in the general case of arbitrary degree sequence.
The procedure works in the same way as in the configuration method, but additionally checking at each stubs pairing that the operation does not create

a multiedge or a self-loop. In an operative way, one proceeds as follows. The starting condition is a set U of $N(\sum_i k_i)$ unpaired points, divided in N groups of cardinalities k_i. Then, one chooses uniformly at random two points i and j in U. If they are suitable, they are deleted from U and the pair $i - j$ is created. The term suitable indicates two points that lie in different groups and such that no currently existing pair contains points in the same two groups. The pairing is repeated until no suitable couple can be found. In the end, a graph is created, with a link between vertices a and b if and only if there is a pair containing points in the a-th and in the b-th groups. The drawback of this procedure is that it might not return a regular graph and it might be necessary to start all over from the beginning. For instance, one possible situation is to remain only with two unpaired points within the same group. In this case there is no suitable matching, and at the same time two stubs are left free. However, even considering this inconvenience, the algorithm is still fast. In calculators, the overall runtime for random regular graphs was estimated to go as kN.
The Steger-Wormald method is the algorithm adopted in this project to generate random regular graphs, both fully regular and with one hub added. In the former case, the initial condition sees N groups of k points. In the latter case, being q the degree of the hub, the initial set of unpaired points U consists of one group of q points and $N - 1$ groups of k points, for a total of $q + k\,(N - 1)$ points.

4.3 The adjacency list

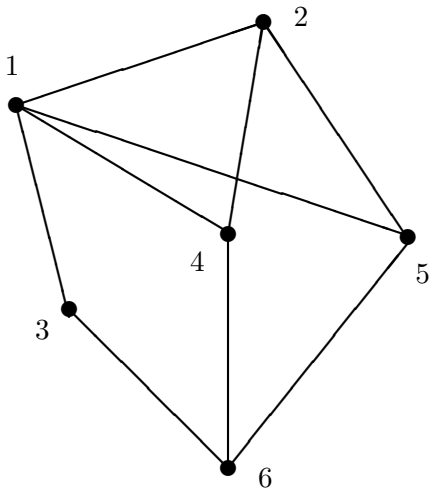
The easiest way to represent a graph is usually through its adjacency matrix, but for large N it requires a huge amount of memory resources. In the case of random regular networks with small coordination number $k \ll N$, however, using a so large amount of memory would be pointless, since most of the elements are null. Instead of the adjacency matrix, it is convenient to use the adjacency list. It is a set of N vectors where the components of the vector i are the labels of the nearest neighbors of node i, with $i = 1, 2, ..., N$. This is in general convenient for all sparse networks, namely networks with small average degree $\langle k \rangle$. The total number of required entries is close to $\langle k \rangle N$, versus the $N^2 \gg \langle k \rangle N$ elements of the adjacency matrix.
Furthermore, the finite size of the networks generated requires to take into account some restrictions in the degree of the nodes. As pointed out in Paragraph 4.1, setting the degree sequence implies in turn to fix the total number of edges L through the constraint $\sum_i k_i = 2L$. Since each edge occupies two entries in the adjacency list, which for random regular networks has kN components, the following condition must be satisfied:

$$kN = 2L \tag{4.1}$$

Then, to have a physically reasonable L, the number kN must be even.
In the case a hub is introduced the condition becomes

$$k(N-1)+q=2L \tag{4.2}$$

Then, $k(N-1)+q$ must be even.



1 →	2	3	4	5
2 →	1	4	5	
3 →	1	6		
4 →	1	2	6	
5 →	1	2	6	
6 →	3	4	5	

Figure 4.3: Example of adjacency list of a small graph.

4.4 Epidemic dynamics

Once the network is built it is possible to assign to each vertex its initial state at time $t_0 = 0$, that can be susceptible or infective. The state of the entire system is given by the particular configuration of susceptible and infective vertices. The temporal evolution of the SIS model is a Markov jump process, that is performed by means of a continuous time algorithm [30][31].
In the contagion process two quantities determine the successive state of the system and when the state transition occurs. These quantities are the number of infected vertices $N_I(t)$ and the number of active links $L_A(t)$, defined as the total number of links emanating from any infected vertex. In this definition, links between two infected vertices are counted twice. Two different lists keep track of vertices and edges belonging to these two groups.

The event determining the state transition is selected at random: recovery with probability

$$P_R(t) = \frac{\mu N_I(t)}{\mu N_I(t) + \beta L_A(t)} \tag{4.3}$$

and infection with probability

$$P_I(t) = 1 - P_R(t) = \frac{\beta L_A(t)}{\mu N_I(t) + \beta L_A(t)} \tag{4.4}$$

As in Chapter 1, here μ indicates the recovery rate and β the infection rate. If the recovery is chosen, a vertex is picked randomly in the list of infected vertices and removed. Otherwise, if the infection is chose, a link is chosen randomly in the list of active links and, if it ends to a susceptible vertex, this latter is turned into infective and added to the respective list. In both the circumstances, time is increased according to

$$t \longrightarrow t + \Delta t = t + \frac{1}{\mu N_I(t) + \beta L_A(t)} \tag{4.5}$$

This procedure is iterated until all vertices are in the healthy state or the process reaches the maximum time.
It is possible to simply rearrange expressions 4.3 and 4.4 as follows[2]:

$$P_R(t) = \frac{\mu N_I(t)}{\mu(N_I(t) + \lambda L_A(t))} = \frac{N_I(t)}{N_I(t) + \lambda L_A(t)} \tag{4.6}$$

$$P_I(t) = \frac{\mu \lambda L_A(t)}{\mu(N_I(t) + \lambda L_A(t))} = \frac{\lambda L_A(t)}{N_I(t) + \lambda L_A(t)} \tag{4.7}$$

This way it visible that the evolution of the system is determined by a single parameter, the reproductive number. It is thus convenient to set $\mu = 1$ and vary β. In such a way one sets the time scale as

$$\Delta t = \frac{1}{N_I(t) + \beta L_A(t)} = \frac{1}{N_I(t) + \lambda L_A(t)} \tag{4.8}$$

In this project, this algorithm was implemented and used to study the SIS phase transition. Results and relative discussions are the subject of the next chapter.

[2] $\lambda = \beta/\mu$, see Chapter 2

Chapter 5

Investigation on the epidemic threshold

A computational investigation on the SIS dynamics is accomplished by means of two different techniques. The first derives from traditional phase transition tools, while the second exploits a parallelism with percolation process. The focus is the critical point position and features in homogeneous structures in presence of isolated hubs of large connectivity.

In this chapter we present the analysis and the results object of this work. The SIS phase transition is investigated on networks generated by using the configuration method, described in Chapter 4. The dynamics is performed through the continuous time algorithm outlined in the same chapter.
It is widely accepted that the measure of the susceptibility - meant as magnitude of fluctuations in the order parameter[1] - represents a valuable method to determine the critical point [14] [27] [30]. Conversely, a new method was recently proposed, exploiting the duration of the contagion process [31]. A first comparison between the two techniques is made on random regular networks, while later on we focus on the analysis of localized states introducing a hub in the network architecture. By using both methods we get a complete comprehension of the dynamical implications of localization transition. As final discussion, we present their advantages and disadvantages.

5.1 General picture

In the SIS framework, the ensemble of states of the system is the set of the possible infective-susceptible configurations and includes an absorbing state

[1] See Chapter 3

- i.e. the configuration with all vertices in the S state. A contagion process is meant to start from an initial configuration with $n(t=0)=1$ seed nodes and stop until the absorbing state is reached. Depending on the value of the reproductive number λ, the infection can remain active inside the system for a certain lapse of time. At low λ the number of infected nodes falls rapidly to zero. At growing λ, the time needed to reach a null prevalence gradually increases and at a point it dramatically diverges. After this point, the epidemic may enter in a stationary state of non-null prevalence in the thermodynamic limit. Otherwise, in a finite network it survives until an extreme fluctuation kills it completely. Picture 5.1 displays the behaviour of $n(t)$ for three different values of λ, evolving from an initial configuration with a single contagious node in a network of size $N = 5 \cdot 10^4$.

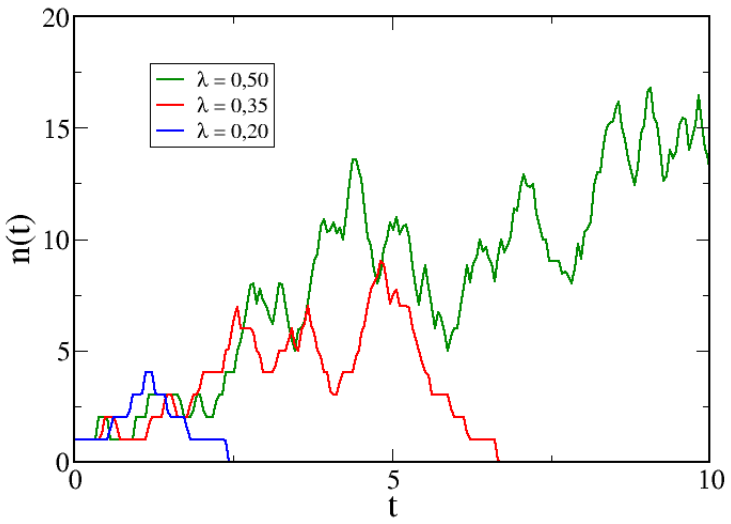


Figure 5.1: Absolute number of infected vertices in function of time, for different values of the reproductive number λ. The initial condition is $n(0) = 1$ and the graph size $N = 5 \cdot 10^4$.

However, we are analyzing a stochastic process, so we need a sample of realizations to build a statistics. Averaging over a large number of repetitions, the prevalence shows an initial outbreak terminating in a peak or in a plateau depending whether λ is set below or above the epidemic threshold. In the former case, after the peak the prevalence exponentially relaxes to zero with a law of the form

$$\rho \sim e^{-t/\tau} \tag{5.1}$$

as it is visible in Picture 5.2. The quantity τ is called relaxation time and is the characteristic time of the prevalence decay. Its behaviour as a function

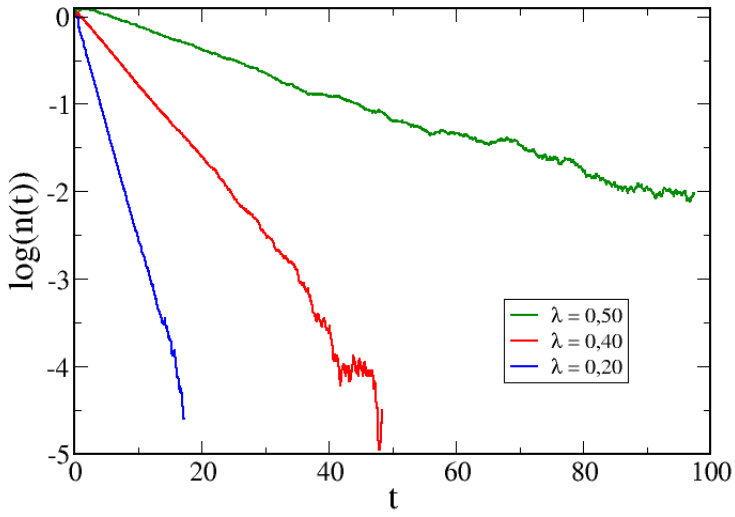


Figure 5.2: Number of infected vertices versus time on the log-normal scale, for different values of λ. Below λ_c, after a brief transient interval the decay is exponential.

of the reproductive number was calculated and is shown in Picture 5.3. Calculations were performed with graphs of size $N = 5{\cdot}10^4$ and coordination number $k = 3$. The relaxation time τ was calculated through a linear fitting of $-log(\rho(t))$ and averaging over 100-200 realizations in 10 different graphs to reduce fluctuations. The neat divergence of τ around $\lambda \simeq 0,54$ evidences the proximity of the phase transition and translates into a sudden increase in the average lifetime $\langle T \rangle$ of the spreading process. We measured this latter quantity in function of λ, together with the maximum duration T_{max}. Results are shown in Pictures 5.4 and 5.5.

The analysis proceeds with a profile of the prevalence in the intermediate region between the sub-critical and the over-critical regimes. The heigh of the peak preceding the exponential decay grows with increasing reproductive number, meaning that the prevalence on average grows. At greater λ the initial boost in the number of infected nodes ends in a plateau, lasting indefinitely in time.

Given a set of a high number of realizations of the process, it is possible to build the prevalence distribution $\bar{P}_n$, defined as the probability distribution for the system to be in a state with n infected vertices. It is computed by taking the absolute frequency of the state n over all simulation runs. We can now define the moment of order m of the number of infective nodes

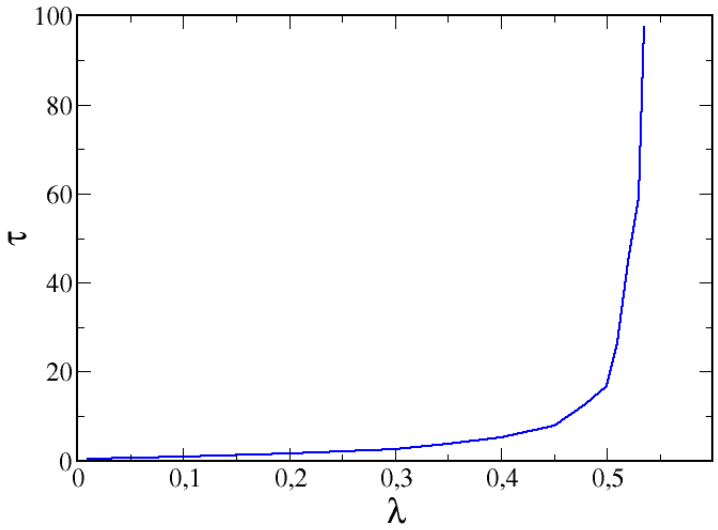


Figure 5.3: Relaxation time as a function of λ in networks of size $N = 5 \cdot 10^4$ and connectivity $k = 3$.

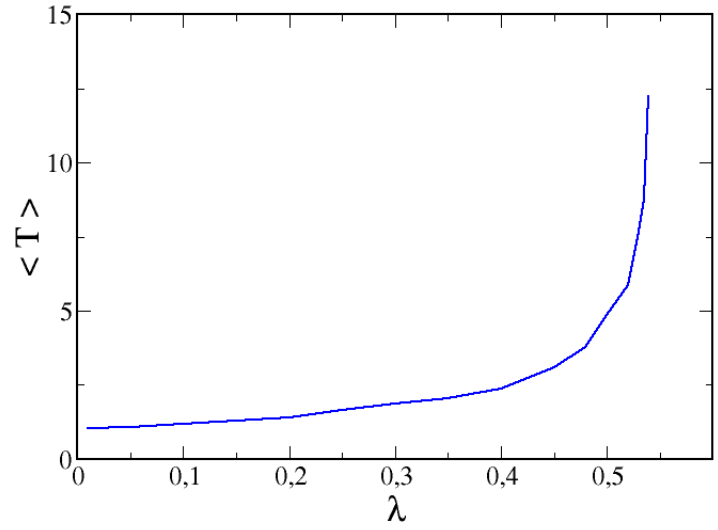


Figure 5.4: Average life span of the whole set of infection processes $\langle t \rangle$ as a function of λ. Networks size is $N = 5 \cdot 10^4$ and connectivity is $k = 3$.

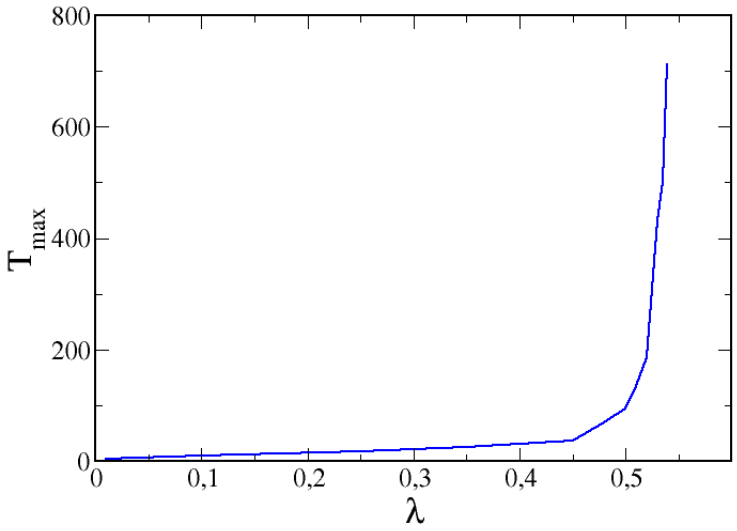


Figure 5.5: Maximum lifetime of the infection process T_{max} as a function of λ. Parameters used are $N = 5 \cdot 10^4$ and $k = 3$.

distribution as

$$\langle n^m \rangle = \sum_{n=1}^{N} n^m \bar{P}_n \tag{5.2}$$

where the sum extends up to the network size N. The first moment is the average number of infective vertices $\langle n \rangle$.
Calculations focus on $\langle n \rangle$ and on the maximum number of infected nodes ever reached n_{max}. As previously, they are performed over 100-200 realizations for 10 different graphs of size $N = 5 \cdot 10^4$. Clearly, at high λ, $\langle n \rangle$ will be close to the value in the plateau, and what is observed is again a continuous divergence in the proximity of $\lambda \simeq 0,54$, as can be seen in Picture 5.6.

Variables τ and $\langle n \rangle$ help us to draw a general picture of what goes on in a typical contagion process. We can see that the position of the critical point is generously out of range with respect to the predictions of both homogeneous, heterogeneous and quenched mean field theories. In the case of random regular networks, in fact, they all state $\lambda_c = \lambda_c^{HOM} = \lambda_c^{HMF} = \lambda_c^{QMF} = 1/3 \simeq 0,33$. This confirms the fact that dynamical correlations, that are not taken into account in these approaches - have a primary role in the value of the epidemic threshold. Indeed, the pair-quenched mean field approach succeeds in getting closer, predicting $\lambda_c^{PQMF} = 1/2 = 0,5$. However, these are just qualitative reasonings. To get deeper into the investigation it is needed to know the exact position of the critical point.

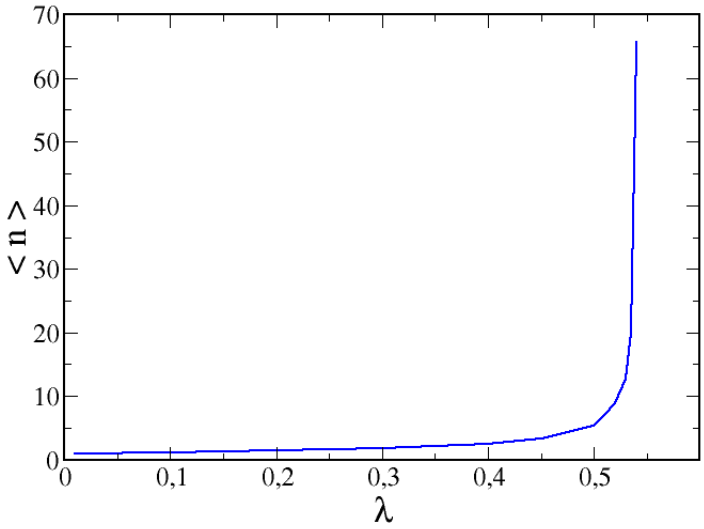


Figure 5.6: Average number of infective nodes over the whole infection process $\langle n \rangle = \langle \rho \rangle N$ as a function of λ. Networks size and connectivity are, respectively, $N = 5 \cdot 10^4$ and $k = 3$.

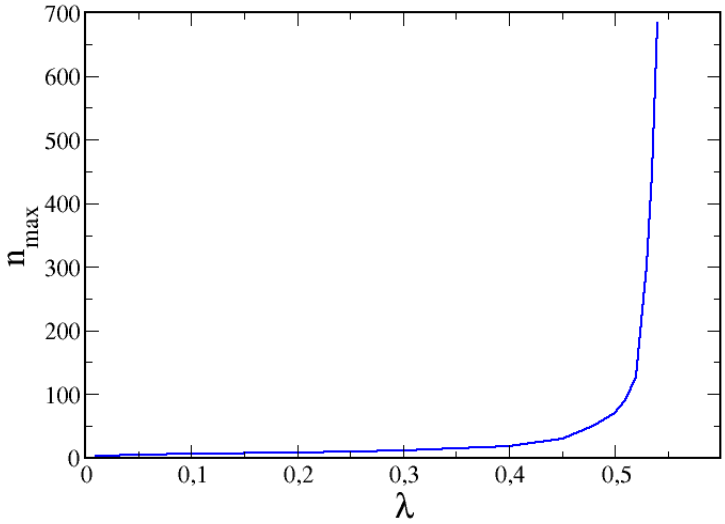


Figure 5.7: Maximum number of infected nodes reached in the ensemble of realizations of the infection process $n_{max} = \rho_{max} N$, as a function of λ. Networks size is set equal to $N = 5 \cdot 10^4$, while the connectivity is set equal to $k = 3$.

5.2 Susceptibility method

As described in Chapter 2, the exact theoretical prediction of the epidemic threshold in a generic network is an open task. On the side of the numerical prediction, one needs to equip with an effective technique in order to compare and verify analytical results. It is then fundamental to choose the right order parameter, able to spot the real critical point in an arbitrary graph architecture. A widely adopted technique is the study of the **susceptibility**, inherited from traditional absorbing phase transitions research[2]. Here it represents the magnitude of stochastic fluctuations in the density of infective vertices and, in the thermodynamic limit, it diverges at the epidemic threshold. In fact, in the sub-critical and over-critical phase $\langle\rho\rangle$ approaches the null value of the absorbing healthy state and the finite value of the stationary fluctuating state, respectively. In the intermediate regime, any realization of the contagion process can undergo a different evolution and ρ presents extreme fluctuations. In the limit $N \to \infty$, the value of $\langle\rho\rangle$ in the steady state is infinite, which entails fluctuations of infinite amplitude at the transition.

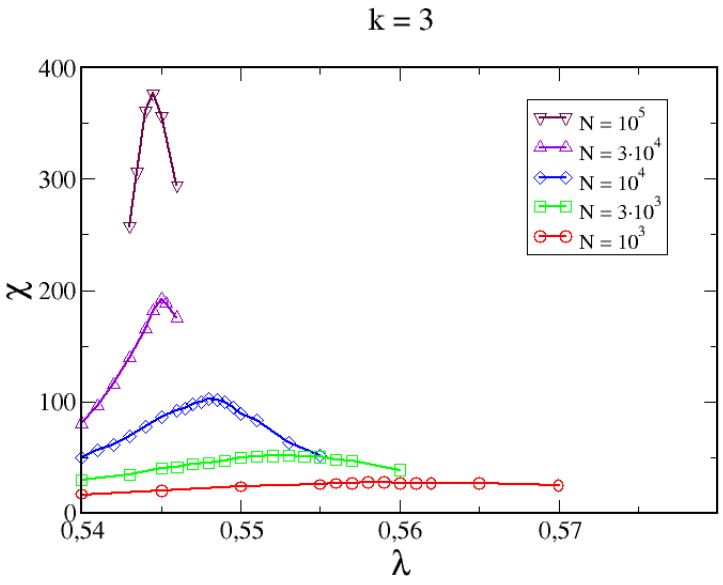


Figure 5.8: Susceptibility peaks for networks of connectivity $k = 3$ and growing sizes.

Moreover, it has been shown that, rescaling χ by a factor $\langle\rho\rangle$, it maintains its scaling properties, while returning clearer results [30]. Thus, the definition

[2]See Chapter 3.

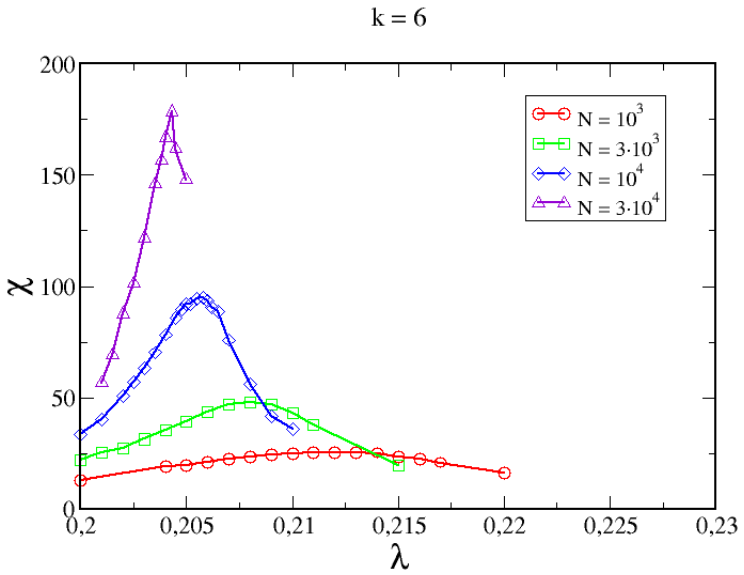


Figure 5.9: Susceptibility peaks for networks of connectivity $k = 6$ and growing sizes.

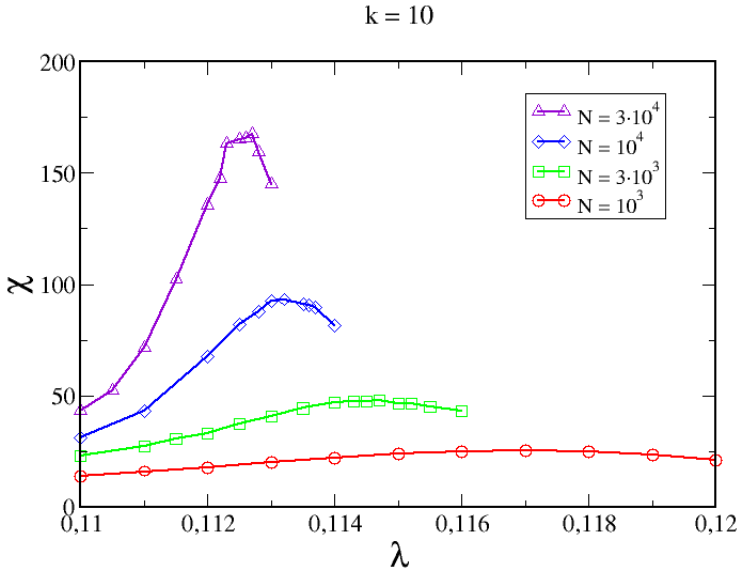


Figure 5.10: Susceptibility peaks for networks of connectivity $k = 10$ and growing sizes.

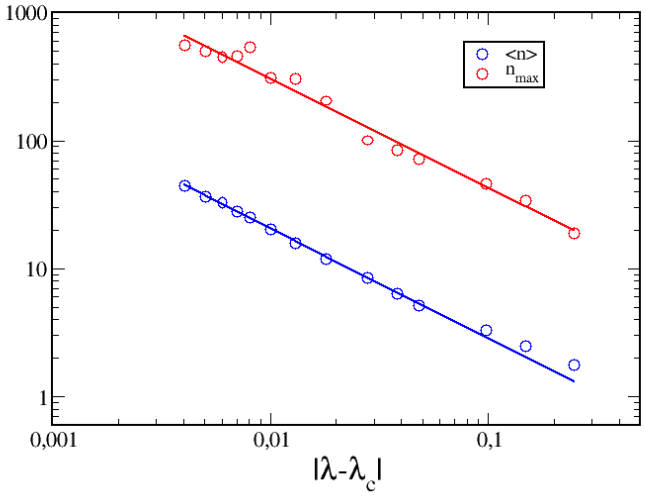


Figure 5.11: Average and maximum number of infective vertices as a function of the distance from the critical point. Parameters are $k = 3$ and $N = 10^4$. Blue line represents a function proportional to $|\lambda - \lambda_c|^{\alpha_1}$ with $\alpha_1 \simeq -0,86$ (correlation index $r \simeq 1$) and red line a function proportional to $|\lambda - \lambda_c|^{\alpha_2}$ with $\alpha_2 \simeq -0,85$ (correlation index $r \simeq 0,96$).

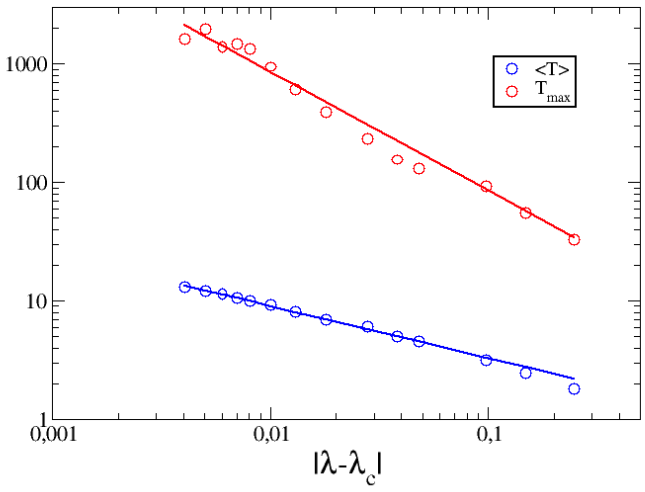


Figure 5.12: Average and maximum life span of the infection process as a function of the distance from the critical point. Parameters are $k = 3$ and $N = 10^4$. Blue line represents a function proportional to $|\lambda - \lambda_c|^{\alpha_1}$ with $\alpha_1 \simeq -0,44$ (correlation index $r \simeq 1$) and red line a function proportional to $|\lambda - \lambda_c|^{\alpha_2}$ with $\alpha_2 \simeq -1$ (correlation index $r \simeq 0,96$).

that will be adopted in the following is

$$\chi = N\frac{\langle\rho^2\rangle - \langle\rho\rangle^2}{\langle\rho\rangle} = \frac{\langle n^2\rangle - \langle n\rangle^2}{\langle n\rangle} \tag{5.3}$$

Simulations were performed to spot the critical point in random regular graphs of degree $k = 3, 6, 10$, whose outcome is illustrated in Picture 5.8, 5.9 and 5.10. Curves are obtained by averaging over 1000-5000 realizations for each of 10-100 different graphs. To overcome finite-size limitations, networks of growing sizes are used. For technical reasons the top value of N is not very large, but still sufficient to build a route to the thermodynamic limit. In all charts, peaks move toward the value depicted by pair-quenched mean field theory as N increases. Indeed, for $k = 3$ one has $\lambda_c^{PQMF} = 1/2 = 0,2$, while for $k = 6$ one has $\lambda_c^{PQMF} = 1/5 = 0,2$ and for $k = 10$ one has $\lambda_c^{PQMF} = 1/9 \simeq 1,11$. These values agree with some published results on random regular networks [30].
Moreover, we performed an analysis on the scaling of quantities presented in the first paragraph of the chapter. Data show a good power-law scaling for all of them.

5.3 Life span method

A second technique, of recent conception, proposes to seek a threshold behaviour in the time variable and concentrate on the duration of epidemics [31]. This method is based, again, on the analysis of the SIS diffusion process starting from a single infected vertex. In the sub-critical regime, the closer the critical point and the longer realizations on average survive. Right over the critical point, instead, if $N \to \infty$ there is a non null probability that the infection lasts indefinitely in time. Realizations for which it is the case are said to be endemic. The appearance of at least an endemic realization signals that the epidemic threshold has been passed over. The situation is thus similar to that of percolation model[3], wherein clusters size increases up until the formation of a giant component. As in that case one measures the average size of finite clusters, here the strategy consists in measuring the average lifetime T_{finite} of those realizations that die out in a finite time. Varying the order parameter, the outcome is a curve with a peak at the value of λ at which realizations are more long-lasting before they start to become endemic.
To discriminate whether a realization is finite or endemic, a new quantity is defined: the **epidemic coverage** (here indicated with C). It represents the fraction of distinct vertices ever infected during the whole realization. In the thermodynamic limit, it is a secure indicator of the belonging to one of the two classes of realizations: if the infection dies out in a finite time it

[3]See Paragraph 3.4

has also a finite coverage, which disappears with N tending to infinity. Vice versa, finite C means finite life span. Moreover, to obtain a coverage equal to 1 it is required an infinite time and if the disease infects the network for an infinite time then C is equal to 1. In definitive, $C = 1$ if and only if the system finds the active stationary state.
Difficulties arise when one wants to measure the coverage, having systems of finite size. How to establish in this case when a realization is endemic? Using the same criteria of the infinite size case can be too expensive in terms of computational time. The proposal is to exploit the fact that, if $N \to \infty$ and C reaches a finite value different from 0, then the realization must be endemic. It would, therefore, be possible to set a threshold of C different from 1 to distinguish between endemic and finite realizations. An immediate advantage is that, after a realization has reached this value of coverage, it is possible to truncate it - saving precious time.
This technique has been successful in confirming epidemic thresholds for scale-free networks, obtained by a new theoretical approach which takes into account the characteristic reinfection time of distant vertices [31]. However, the validity of both the theoretical approach and the life span method is still under debate. Data presented in the following aim to verify the reliability of the life span method in the simplest possible configuration, namely pure homogeneous random graphs.

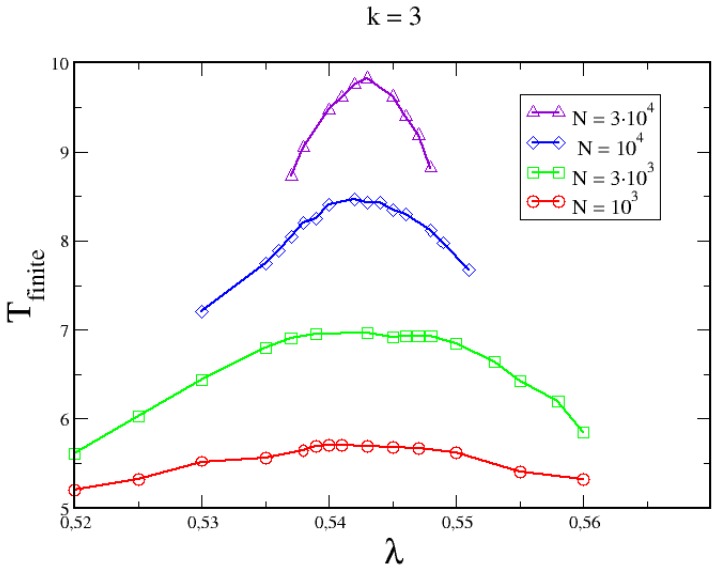


Figure 5.13: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 3$ and growing sizes.

According to previous tests [31], the value of C would be non influential

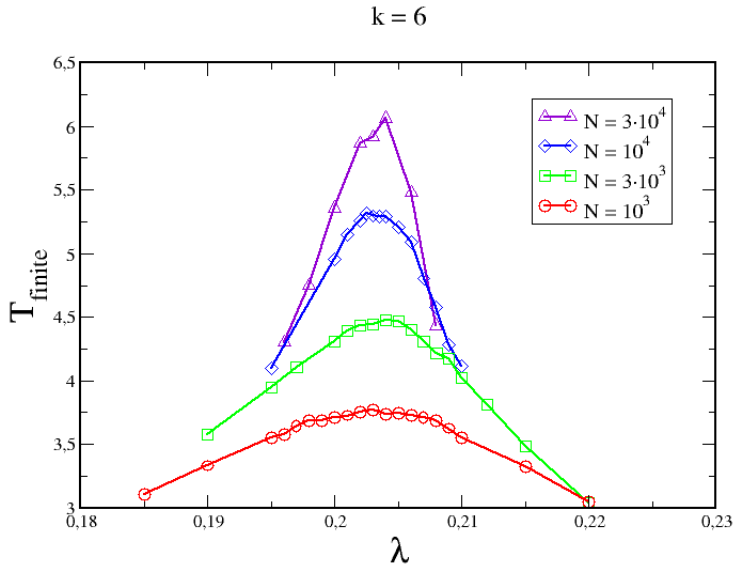


Figure 5.14: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 6$ and growing sizes.

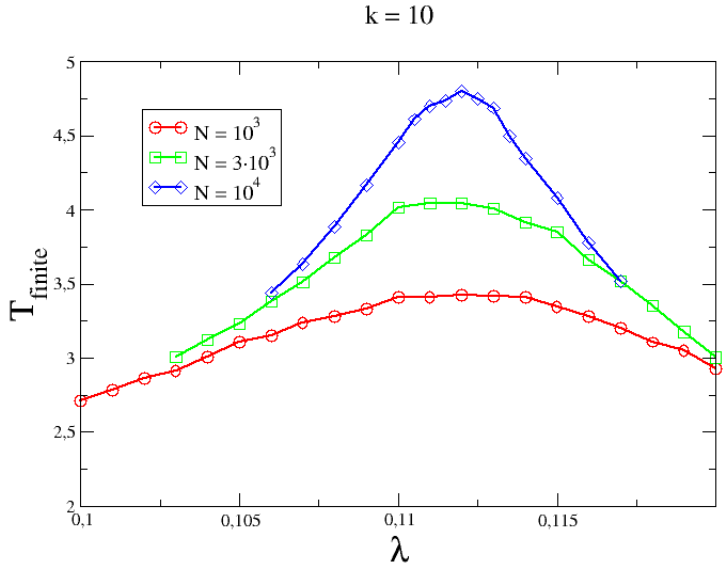


Figure 5.15: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 10$ and growing sizes.

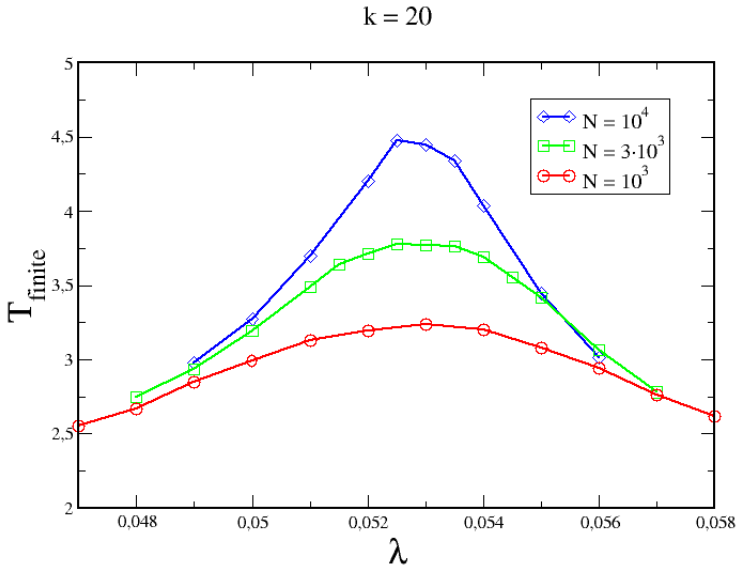


Figure 5.16: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 20$ and growing sizes.

in the position of the critical point. We thus set the coverage threshold at $C = 0,5$ and performed a T_{finite} measure again with $k = 3, 6, 10$ and varying N. Pictures 5.13, 5.14, 5.15 and 5.16 show the curves obtained.
At a first glance one can notice that peaks are stable around a precise value of the reproductive number and the shift exponent is close to 0. At small scales, results strongly differ from those provided by the susceptibility. The critical point, however, is located again in the proximity of λ_c^{PQMF} for all vertex degrees. This technique seems then to catch the thermodynamic limit value of the epidemic threshold much better than susceptibility. Finite size seems to have slight effects on the position of the peak. However, there is an evident failure in finding the critical point at the effective system size. To support this conclusion, we performed a study on critical scaling of the relaxation time. The epidemic thresholds given by both methods were estimated through a quadratic fit of the top of peaks with $N = 10^4$. Taking as λ_c the value given by the susceptibility, data are well fitted by a power-law function. On the other hand, taking the value obtained with the life span method, data do not align on a double logarithmic plot (Figure 5.17).The life span method seems therefore to be imprecise at small size scales. In the large size limit, however, the method seems to successfully yield results in agreement with the susceptibility method, but further tests at larger scales are required.

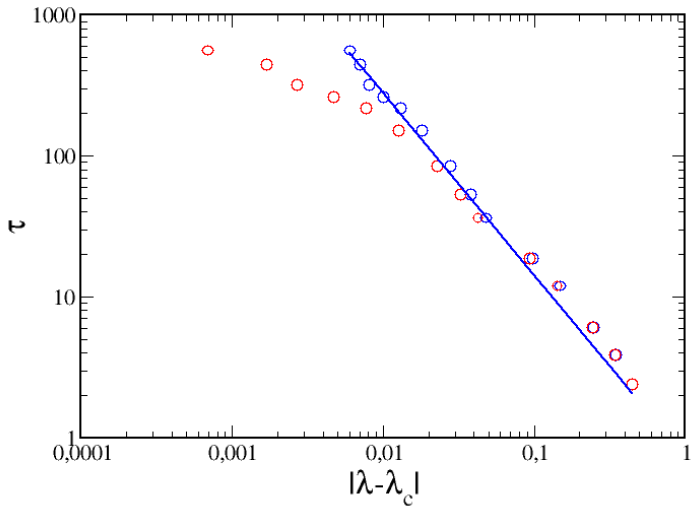


Figure 5.17: Relaxation time of the infection process as a function of the distance from the critical point, obtained both by the χ (blue series) and T_{finite} (red series) measures. Parameters are $k = 3$ and $N = 10^4$. Blue line represents a function proportional to $|\lambda - \lambda_c|^\alpha$ with $\alpha \simeq -1,29$ (correlation index $r \simeq 0,995$).

5.4 Transition with localized states

The subsequent analysis concerns the effect of hubs on the epidemic phase transition, aiming to investigate the predictions on networks with a localized principal eigenstate[4]. In particular, we look for numerical evidences of the outliving of the infection in a restricted number of vertices, most likely around the hub.
A single node of degree $q > k$ was so introduced in the regular architecture and its degree was set at $q = 10 \cdot k$, with $k = 3, 6, 10, 20$. All parameters values give an inverse participation value $IPR(\Lambda_1) \simeq 0,20$, indicating that the principal eigenstate is localized. To efficiently capture the dynamical effects of the hub, this node was also imposed to be the seed of the infection. Analysis on the susceptibility was repeated, revealing a different scenario. The hub has multiple consequences on the susceptibility peaks.

First of all, peaks are shifted toward the left on the order parameter axis, as one can see in Pictures 5.18, 5.19, 5.20 and 5.21. This reflects the intuitive fact that the higher connectivity of the hub allows it to spread the infection more efficiently and so it works as a sort of "amplifier" of contagion. At the same time, even if the hub turns susceptible, it is sufficient that one of its several nearest neighbours infects it again to return an active spreader.

[4]See Paragraph 2.4

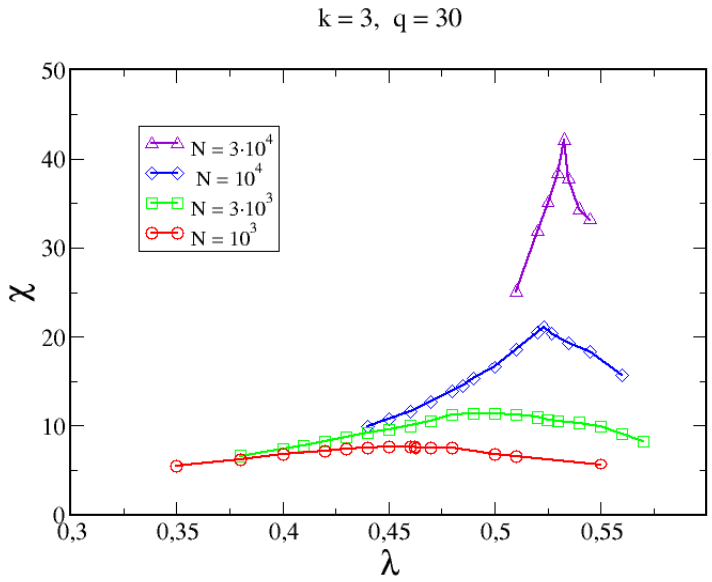


Figure 5.18: Susceptibility peaks for networks of connectivity $k = 3$, hub degree $q = 30$ and growing sizes. Sizes are those indicated in the legend box plus one, to fulfill the condition 4.2.

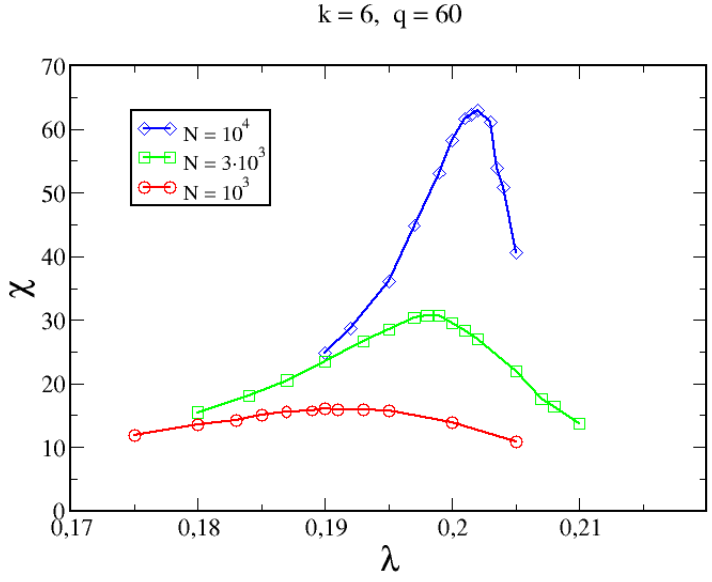


Figure 5.19: Susceptibility peaks for networks of connectivity $k = 6$, hub degree $q = 60$ and growing sizes.

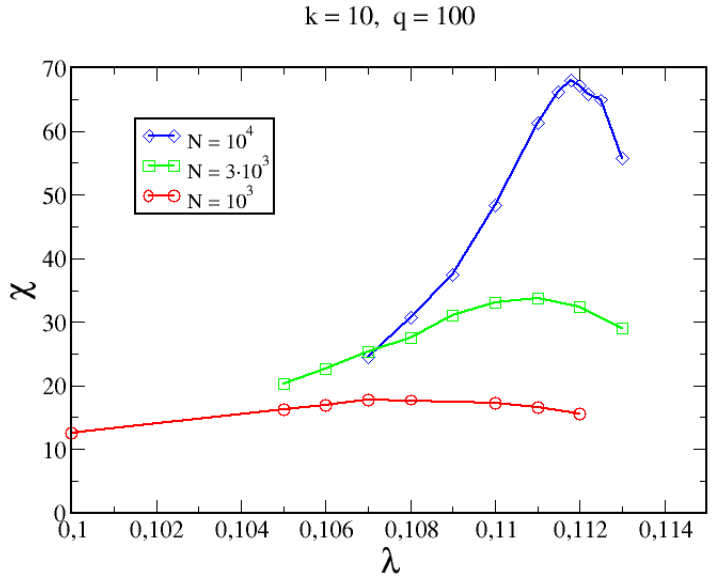


Figure 5.20: Susceptibility peaks for networks of connectivity $k = 10$, hub degree $q = 100$ and growing sizes.

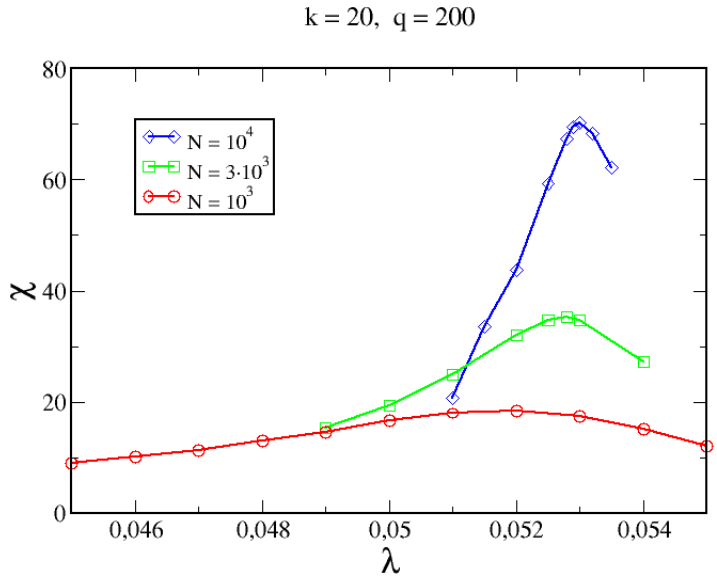


Figure 5.21: Susceptibility peaks for networks of connectivity $k = 20$, hub degree $q = 200$ and growing sizes.

Consequently, a smaller value of λ is necessary to the disease to stay alive. Furthermore, the shift of the peaks with respect to the fully regular case is more evident as network size decreases. The reason is that q is fixed and N varies, so the heterogeneity of the hub is perceived more when the dynamics is restricted to a limited number of vertices, while tends to vanish in larger and larger systems. An evidence of this fact is that the scaling toward infinite size seems to tend again to the value close to prediction of the pair-quenched mean field approach for random regular graphs. Poor agreement is found instead with the epidemic threshold given by quenched mean field $\lambda_c^{QMF}(k,q) = 1/\Lambda_1$. Indeed, by using Equation (2.23), one obtains $\lambda_c^{QMF}(3,30) \simeq 0,176$, $\lambda_c^{QMF}(6,60) \simeq 0,124$, $\lambda_c^{QMF}(10,100) \simeq 0,095$ and $\lambda_c^{QMF}(20,200) \simeq 0,067$. Clearly none of these is the limit value of the scaling for any susceptibility peak. This discrepancy is not a surprise though, considered that these results are extracted within a quenched mean-field approach, not taking into account dynamical correlations.
A second aspect is the height of the peaks, which means the magnitude of prevalence fluctuations. It is clear from the charts that the hub has a dampening effect on susceptibility, since all peaks are lower than in the corresponding regular graphs. This can be explained by conjecturing that the hub creates an island of infective vertices around itself which, on average, determines the mean prevalence in the network and leaves to the dynamics in the homogeneous rest of the graph a little influence.
Later on, simulations were repeated setting $k = 3, 4, 6$ and fixing the hub degree at $q = 100$. This time, results of susceptibility measures are dramatically different from the previous ones. In fact, in charts 5.22 and 5.23 a second peak appears on the left of the peak associated to the phase transition. The new peak is in the same position for all k at small scales - i.e. $N = 10^3$ - and scales with N according to k. In the large size limit the two peaks tend to overlap and add up to a single peak.
In chart 5.24 relative to $k = 6$ only one peak is present, though. This fact led us to the idea that in previous analysis with smaller hubs the same thing could have happened. Consequently, we performed a further analysis using as parameters $k = 3$, $q = 60$ and $k = 2$, $q = 30$.
From charts 5.25 and 5.26 it is possible to see that the position of peaks relative to the localization transition depends on the degree of the hub. With $q = 100$, its top value is around $\lambda \simeq 0,20$, with $q = 60$ is around $\lambda \simeq 0,28$ and with $q = 30$ around $\lambda \simeq 0,40$. The highest is q and the lowest is the reproductive number required to generate an explosive dynamical instability. The new peak reflects thus the strong fluctuations caused by the high number of connections hub. This instability explodes at a precise value of λ determined by the degree q at small scales and shifts with growing size according to the value of k. A threshold value of the reproductive number associated to the hub emerges, which will be indicated as λ_h in the following. Even without precise quantitative results we can extract some observations.

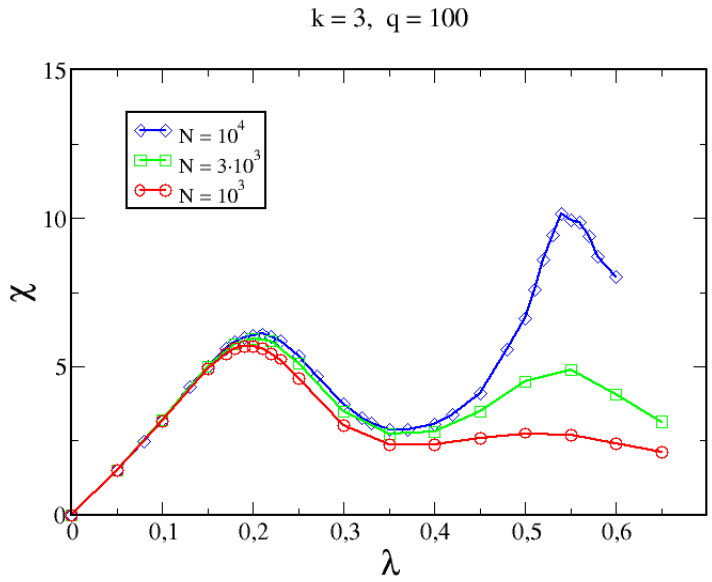


Figure 5.22: Susceptibility for networks of connectivity $k = 3$, hub degree $q = 100$ and growing sizes. Sizes are those indicated in the legend box plus one, to fulfill the condition 4.2.

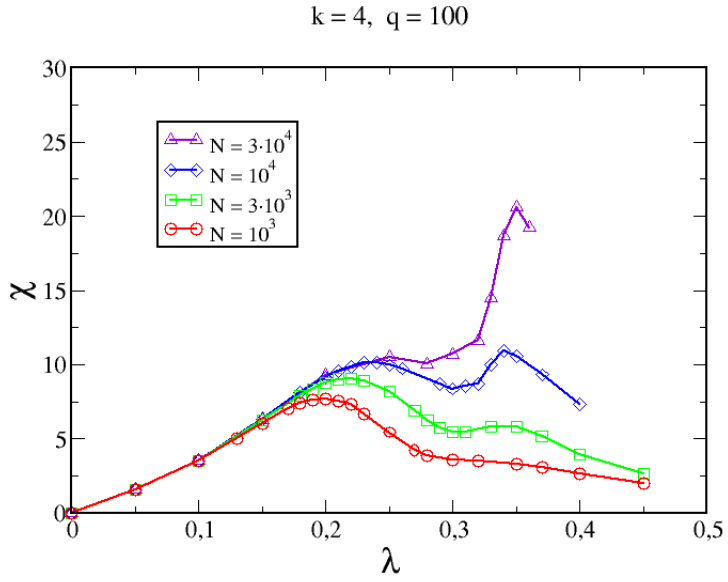


Figure 5.23: Susceptibility for networks of connectivity $k = 4$, hub degree $q = 100$ and growing sizes.

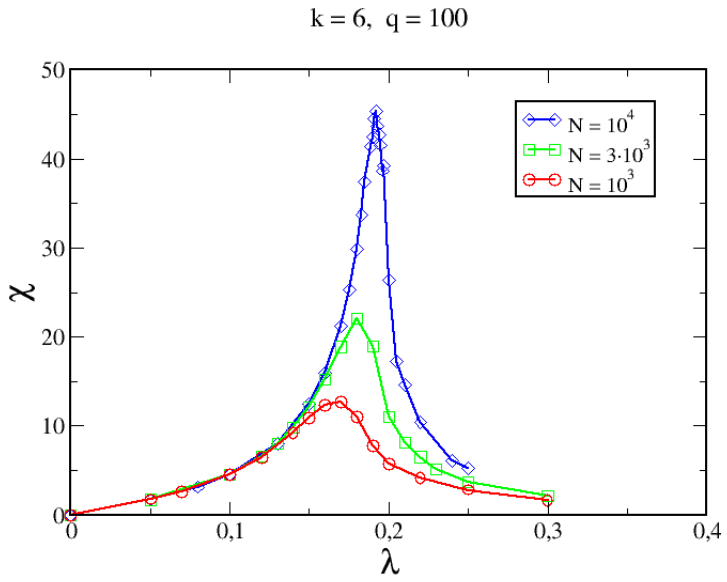


Figure 5.24: Susceptibility for networks of connectivity $k = 6$, hub degree $q = 100$ and growing sizes.

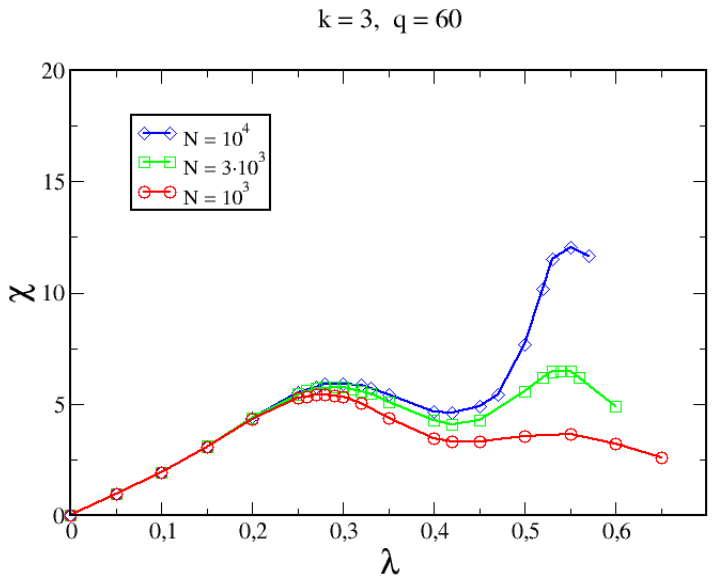


Figure 5.25: Susceptibility for networks of connectivity $k = 3$, hub degree $q = 60$ and growing sizes.

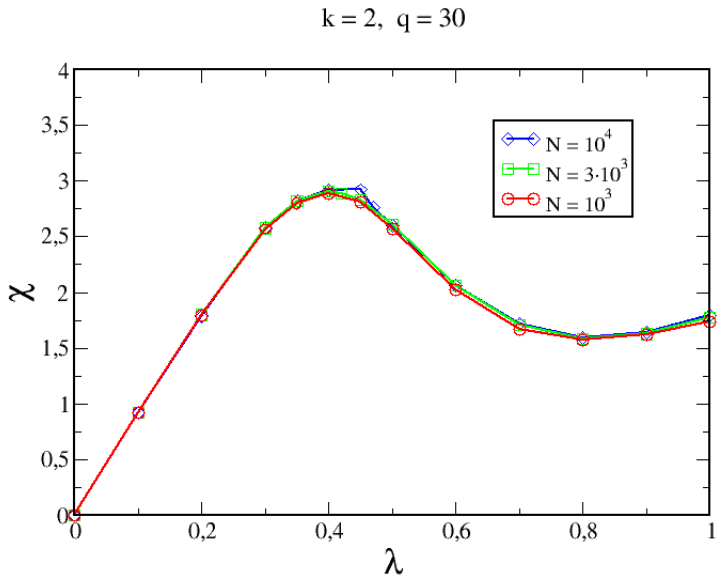


Figure 5.26: Susceptibility for networks of connectivity $k = 2$, hub degree $q = 30$ and growing sizes.

If $\lambda_h < \lambda_c$, then the double peak scenario turns out. At small N, the leftmost peak dominates, suggesting that hub is able to bring a stronger instability then the network itself. Yet, increasing N the rightmost peak grows faster than the other and rapidly absorbs it. This fact indicates that the hub instability is actually restricted within a limited number of vertices and, in the limit $N \to \infty$ vanishes. On the other hand, if $\lambda_h > \lambda_c$ then the result is the usual one peak scenario. Even so, the top value of the peak is lowered by the presence of the hub, which dampens density fluctuations. These two alternative pictures suggest a competition between the hub and the rest of the network in a infection outbreak.

5.4.1 Life span of localized states

The analysis continues with a comparison between the life span and the susceptibility methods in the case a hub is added to the regular configuration. We performed again simulations with the same parameters as before, namely $k = 3, 6, 10, 20$ and $q = 10 \cdot k$.
Also in this case, we find that all peaks are shifted toward low λ values with respect to the corresponding peaks of equal connectivity and network size in the fully regular case. The peak shift here is more consistent at smallest scales, in the same way as in the susceptibility analysis. Again, it reflects the fact that a smaller value of λ is required for the phase transition in pres-

ence of a strong spreader vertex. The limit value of the epidemic threshold seems to be again close to that of pair-quenched mean field. Moreover, it is worth to be noted that the height of the peaks is sensitively larger than in the regular configuration, leading to the observation that the hub actually expands the lifetime of the spreading activity.

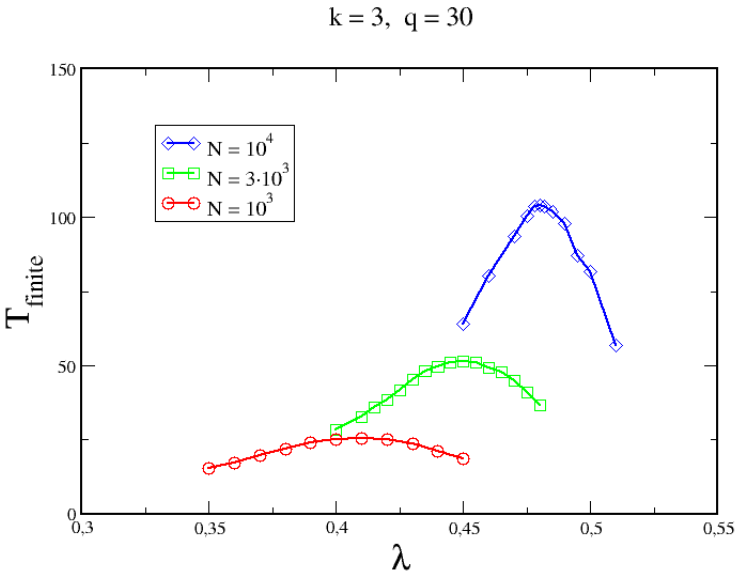


Figure 5.27: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 3$, hub degree $q = 30$ and growing sizes.

Nevertheless, by setting $k = 3, 4, 6$ and $q = 100$ the scenario remains basically the same. Figure 5.31, 5.32 and 5.33 illustrate the outcome. Also in this case only one peak is visible, while with the same parameters the susceptibility presents two peaks. The life span method appears to be sensitive to the localization transition, but is not able to give any information on the competition between the hub and the network. In the specific, it is not useful to say whether the contagion dynamics is local or global. Instead, with parameters corresponding to the double-peaked susceptibility, this method returns a peak that is located in between the top values of χ. Particularly for $k = 3$ and $q = 100$, it is visible that T_{finite} yields a peak that is midway between λ_h and λ_c. In this case the present method clearly fails in its principal task, determining the epidemic threshold.
In any case, combining the susceptibility and the life span method we can verify that, as expected, a hub in a homogeneous network configuration provokes a long-lasting, localized active dynamics, starting below the critical point.

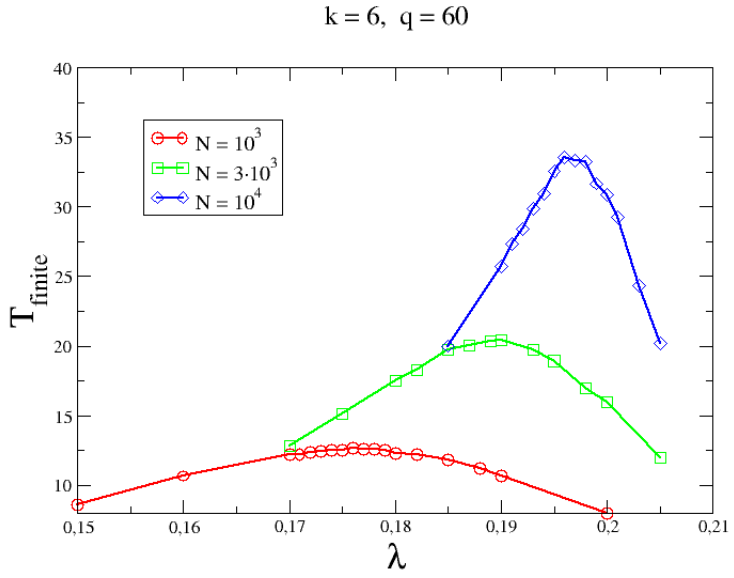


Figure 5.28: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 3$, hub degree $q = 60$ and growing sizes.

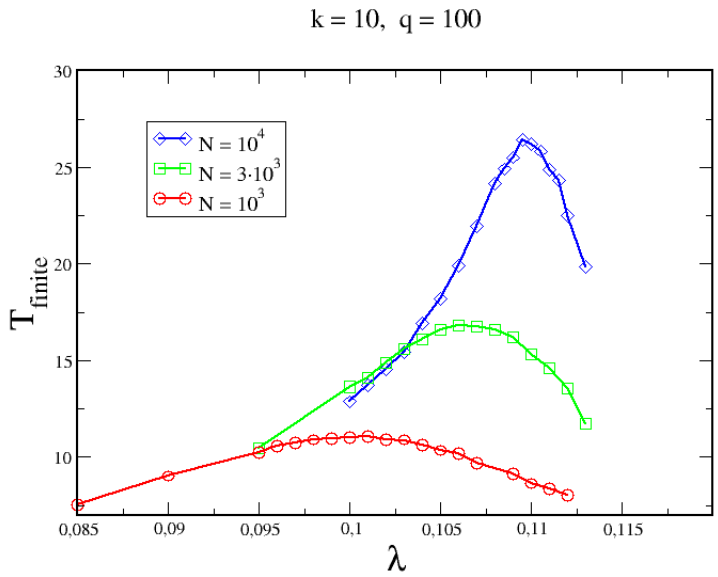


Figure 5.29: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 10$, hub degree $q = 100$ and growing sizes.

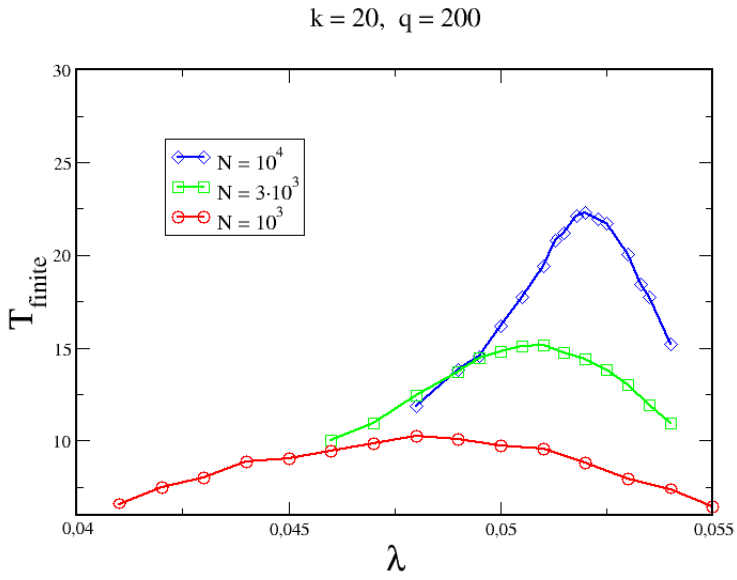


Figure 5.30: Peaks of lifetime of non-endemic realizations for networks of connectivity $k = 20$, hub degree $q = 200$ and growing sizes.

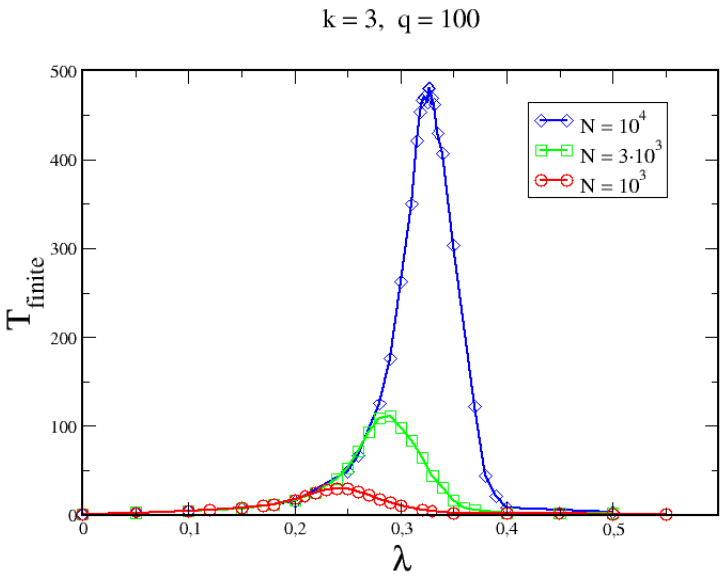


Figure 5.31: Measure of T_{finite} for networks of connectivity $k = 3$, hub degree $q = 100$ and growing sizes.

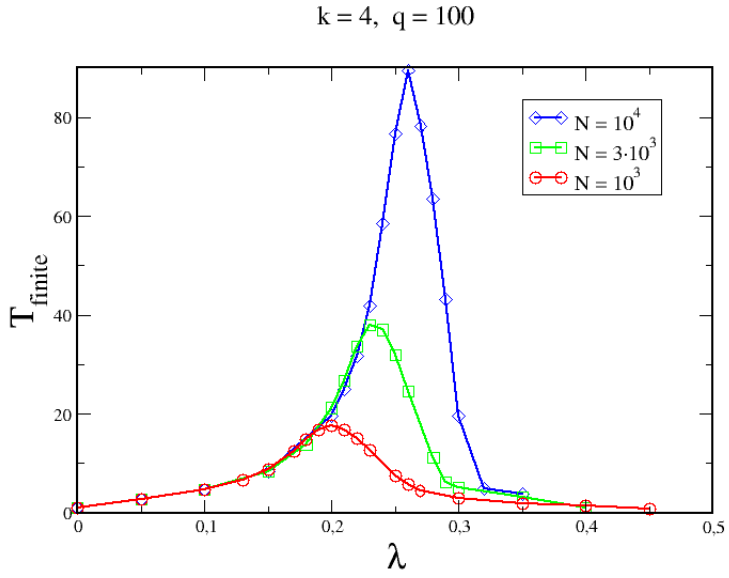


Figure 5.32: Measure of T_{finite} for networks of connectivity $k = 4$, hub degree $q = 100$ and growing sizes.

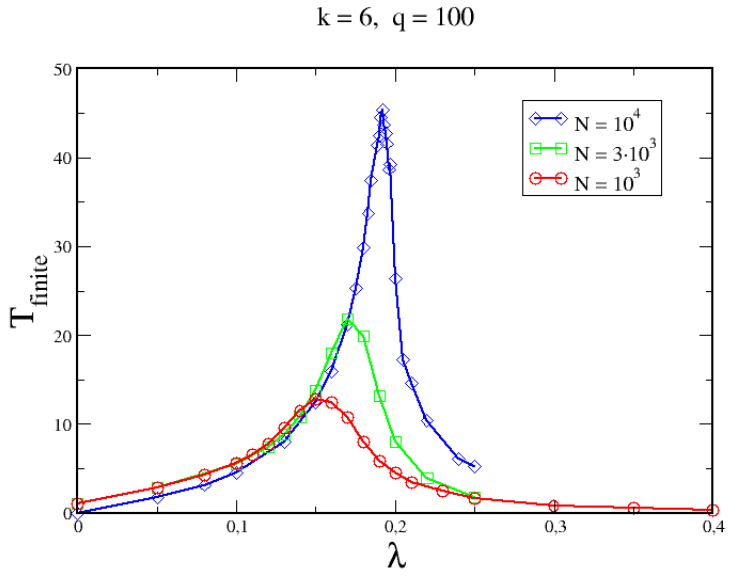


Figure 5.33: Measure of T_{finite} for networks of connectivity $k = 6$, hub degree $q = 100$ and growing sizes.

As additional helpful consideration, we have to cite a very recent work on this subject [33]. Here the life span method is tested by performing simulations on double random regular graphs, namely two random regular graphs of degrees k_1 and k_2 connected by a single link. Results show a double-peaked susceptibility in correspondence of $1/(k_1 - 1)$ and $1/(k_2 - 1)$, in agreement with pair-quenched mean field. Differently, the life span method yields only one peak in an intermediate position with respect to the other two peaks. The scenario is thus similar to that with a random regular network and a hub.
These results suggest also that, using random regular networks with two distinct hubs strong enough, three peaks will emerge. For now, however, this remains an hypothesis.

5.5 Critical scaling of localized states

To gain a wider view of the role of hubs in the dynamics, we studied as final thing the critical scaling of characteristic quantities described in the first paragraph of the chapter. The scaling is considered below the transition point λ_c. In order to get a better comparison with the scaling in the completely delocalized networks, parameters were set, as previously, $k = 3$ and $N = 10^4$. The hub size was taken $q = 100$. We resorted to the susceptibility method and to a quadratic interpolation on the top of the right peak to determine precise values for λ_c.
As regards the density of infected vertices, the power-law scaling is lost and both $\langle n \rangle$ and n_{max} increase exponentially approaching λ_c. Temporal variables appear instead approaching a typical power-law scaling. The duration of the active spreading is enhanced by the hub but follows ordinary scaling rules. This fact underlines two different scaling behaviours for temporal and spatial variables dictated by two different correlation lengths $\xi_\perp$ and $\xi_\parallel$.

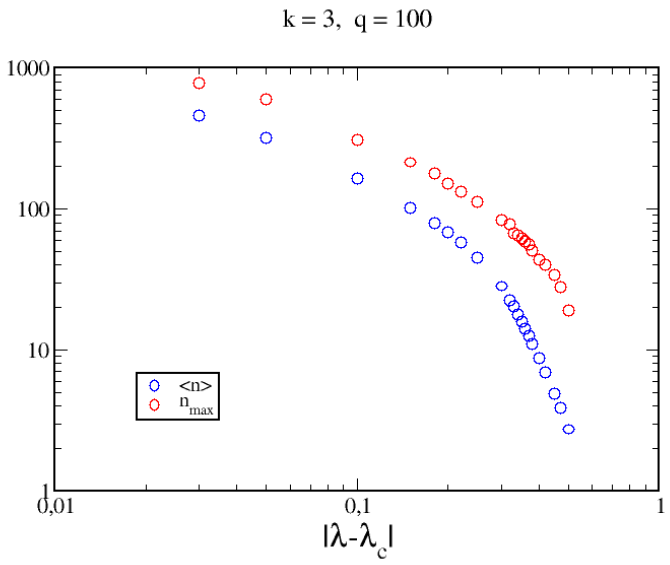


Figure 5.34: Average and maximum number of infected vertices as a function of the distance from the critical point. Networks size is set $N = 10^4$.

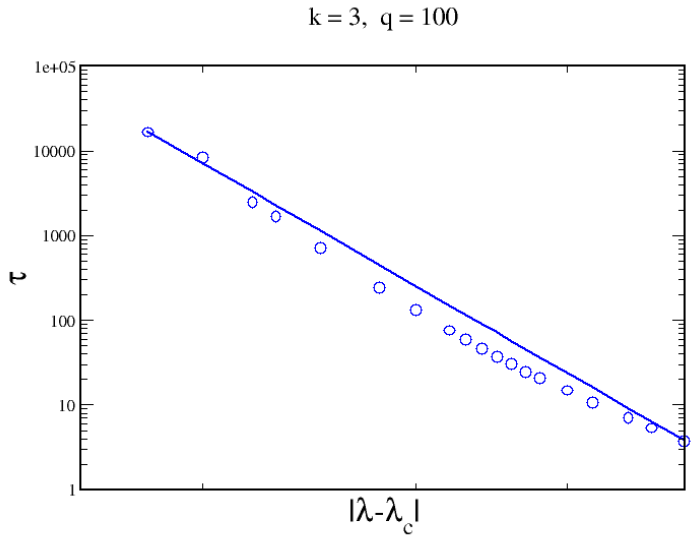


Figure 5.35: Relaxation time of infection process as a function of the distance from the critical point. Networks size is set $N = 10^4$.

Conclusions

In this project, dynamical and critical properties of the SIS phase transition were explored by means of a numerical approach. The analysis focused on the role of hubs inserted in random regular networks, i.e. sets of graphs in which edges extremes are casually assigned in equal number to each vertex. With the addiction of the hub, such structures build a bridge between regular spaces such as a Bethe lattice and complex networks, characterized by highly heterogeneous degree distributions. Moreover, the choice of a homogeneous structure with a single heterogeneity allows to isolate and better capture its effect on the SIS dynamics.
The investigation aimed to verify analytical predictions deriving from a spectral analysis in networks of this typology. The presence of the hub determines a localization transition in the leading eigenvector of the adjacency matrix, which has been predicted to affect the traditional SIS phase transition. In particular, analysis of the contagion process stationary states leads to find an intermediate regime - right over the epidemic threshold - with long-lasting epidemics surviving within a restricted number of vertices. An analytical expression for the critical point was also calculated in the case of a Bethe lattice with a hub.
The transition from the healthy to the endemic phase was inspected through epidemic susceptibility measures, which is the most widely accepted technique to detect and study phase transitions within the contact process and the SIS model. Simulations first concerned fully random regular configurations and yielded outcomes in agreement with pair-quenched mean field theory, in the large size limit, and markedly far from previous mean field approaches. As expected, the pair-approximation approach turns out to be the most effective tool in the theoretical prediction of the epidemic threshold position in homogeneous environments. Indeed, it is the only approach taking into account dynamical correlations of the state space.
Introducing a single hub is sufficient to produce visible repercussions. While the limit value of the critical reproductive number remains the same, a different finite size scaling appears, reflecting the major ease of reaching a powerful contagion in the presence of a strong spreader vertex. The epidemic threshold, thus, shifts from low toward larger values of the reproductive number as the system size grows. At larger and larger size, hub effects tend

to disappear, so that susceptibility peaks approach again the pair-quenched value. In addiction, all peaks display a smaller top magnitude than in the corresponding fully regular networks with same connectivity, suggesting that the contagion dynamics is controlled and stabilized by the hub.
Moreover, by using a stronger hub with very high connectivity - in low degree networks - enables us to testify the emerging of an additional susceptibility peak, associated to a highly fluctuating activity within the hub neighbourhood. The new peak separates the absorbing phase and a regime with long-term - but not endemic - dynamics, wherein the activity is concentrated around the hub and is due to its greater spreading capacity. The endemic phase is reached only at further increase of the reproductive number, when it reaches the typical critical point of random regular graphs. The position of the leftmost peak appears to depend on the hub degree at small graph scales - i.e. 10^3 vertices - and to be affected by finite size scaling according to the degree of the homogeneous component of the graph. Plus, if the epidemic threshold is at a smaller value of the reproductive parameter than the hub threshold, only one peak arises. There seem to be then a sort of competition between the hub and the rest of the graph. Which of the two causes an outbreak in order parameter fluctuations at lower reproductive number depends on the degree of both the former and other vertices. If hub connectivity is large enough, an outbreak in its neighbourhood requires a lower reproductive parameter than the global transition. On the other hand, if a lower reproductive parameter is needed to trigger the phase transition, then relative fluctuations hide the fluctuation outbreak given by the hub. However, even in the double peak case, with increasing network size fluctuations corresponding to the global transition overcome fluctuations of the hub. A wide region of high infective density instability therefore originates. It seems plausible that, in the thermodynamic limit, the leftmost part of this region, relative to a finite number of vertices, tends to vanish under the infinite entity of the phase transition.
To complete the survey, we analyzed the critical scaling of characterizing average quantities in the double peak framework. Findings confirm the breaking in the typical SIS phase transition. Approaching the epidemic threshold from below, in fact, average lifetime of the spreading process scales as a power-law, while average number of infective vertices follows an exponential behaviour. The hub brings then an enhanced dynamics thanks to its high degree.
These outcomes partially agree with the predictions under survey, that claim the appearance of an intermediate regime right above the epidemic threshold. Nevertheless, none of the peaks is located at the expected epidemic threshold for a Bethe lattice with a hub, which results underestimated. This can be understood by remembering that this result was obtained within a quenched mean field context, which was proved to be strongly imprecise. Moreover, the intermediate regime is not guaranteed by simply introducing

a hub, but seems to depend on relationships between degrees of both the hub and the rest of the nodes.
Our second aim was to verify the efficiency of the life span of finite realizations as order parameter of the SIS phase transition. The measure of this variable was proposed as new technique for the detection of endemic transitions in the SIS model on complex networks. Tests were performed again in random regular networks, both in presence and in absence of a hub. In the former case, this method yields a peak in correspondence of values predicted by pair-quenched mean field theory at all network sizes, with shift exponent close to zero. On one hand, it is a positive finding: it allows to obtain a rough estimate of the epidemic threshold by inspecting systems of small size, which require a feeble computational effort. Nevertheless, at small scales the life span measure leads to dramatically discording results with respect to the susceptibility measure. This means that it misses completely the effective epidemic threshold determined by critical finite size effects and turns out to be a bad order parameter for small systems.
In networks with a localized eigenstate, the critical shift appears more evident, one more time reflecting the major ease in activating the phase transition. Even though the life span of finite realizations approaches the pair-quenched value increasing the system size, the critical point is again in systematic defect with respect to the corresponding susceptibility peak at small sizes. Moreover, the life span method totally fails in detecting the twofold behaviour of a network composed by a homogeneous component and a localized heterogeneity. In fact, even at parameters yielding a double-peaked susceptibility, the lifetime of finite realizations displays only one peak. Temporal variables are able to identify only the passage to a long term activity, without discriminating its origin: a localized infection persistence is not distinguished from a global contagion. This is not the greater problem, however. Indeed, when the susceptibility shows a double peak, the limit position of the life span peak is not in correspondence of the right one, nor in correspondence of the left one, but midway between the two. The life span peak seems then trapped between two competing spreading activities due to the hub and the whole network, without being faithful to none of the two. These results lead to the conclusion that attention is needed when dealing with heterogeneous topologies. Certainly, the present investigation requires further developments, but it is not meaningless to presume that the same issues might appear also in other network topologies with strong heterogeneities, even in scale-free and real networks. Also in the homogeneous case, however, the susceptibility method is more reliable, as previously discussed.
To conclude, we take advantage of all results and vulnerabilities of this investigation to suggest possible developments and future directions. Certainly, as the first thing it is required to perform numerical analysis on the SIS dynamics in presence of a single isolated hub at larger network sizes. This

would allow to improve the present analysis and acquire a broader perspective on a wider scale, a step closer to the thermodynamic limit. It would also be useful to quantitatively study the position of the peak connected to the hub activity in relation to its degree and see if a vanishing epidemic threshold is reachable, like in scale-free connectivity configurations. Since hub impact depends on the size of the system, it seems likely that, if there is a threshold for a vanishing critical point, then it will be size dependent. The critical scaling deserves more accurate analysis as well.
Furthermore, the relationship between the localization transition and modifications of the ordinary SIS phase transition remains incomplete. This aspect deserves to be further explored by seeking more precise analytical results, including dynamical correlations and thus most likely combine the adjacency matrix spectral analysis and the pair-approximation approach. Plus, the position of the peak relative to the hub remains to be determined. Another situation to consider is what happens when more than one hub is present on a homogeneous multitude of vertices. Intuitively, an equal number of peaks will appear if at lower reproductive parameters than the global transition. Numerical surveys are required to clarify this occurrence.
In addiction, the emerging of anomalous scaling behaviour gives the role of topology heterogeneities a special role in a wider critical phenomena perspective. Ulterior developments might also focus on absorbing phase transitions in different models, such as the contact process. Definitely, the role of hubs in complex networks is worth to be further investigated, in order to get deeper in the comprehension of complex networks.

Bibliography

[1] Newman M., *Networks: An Introduction*, Oxford University Press, Inc., New York, NY, USA (2010)

[2] Dorogovtsev S. N., Mendes J. F. F., *Evolution of Networks*, Oxford University Press, Inc., New York, NY, USA (2003)

[3] Dorogovtsev S. N., *Lectures on Complex Networks*, Clarendon Press (2010)

[4] Barabási A. L., Albert R., *Emergence of scaling in random networks*, Science 286, 509 (1999)

[5] Barabási A. L., *Scale-free networks: a decade and beyond*, Science 325, 5939 (2009)

[6] Caldarelli G., *Scale-free networks*, Oxford University Press, Oxford (2007)

[7] Barrat A., Barthélemy M., Vespignani A., *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, UK (2008)

[8] Hethcote H. W., *The mathematics of infectious diseases*, SIAM Rev., 42(4) (2000)

[9] Pastor-Satorras R., Vespignani A., *Epidemic dynamics and endemic states in complex networks*, Phys. Rev. Lett. 86, 3200 (2001); Phys. Rev. E 63 066117 (2001)

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

[11] Pastor-Satorras R., Vespignani A., *Epidemic spreading in scale-free networks*, Phys. Rev. Lett. 86, 3200 (2001)

[12] Chakrabarti D., Wang Y., Wang C., Leskovec J., Faloutsos C., *Epidemic thresholds in real networks*, ACM Trans. Inf. Syst. Secur. 10, 1 (2008)

[13] Castellano C., Pastor-Satorras R., *Thresholds for epidemic spreading in networks*, Phys. Rev. Lett. 105, 218701 (2010)

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

[15] Martin T., Zhang X., Newman M. E. J., *Localization and centrality in networks*, arXiv: 1401.5093v1 (not published)

[16] Goltsev A. V., Dorogovtsev S. N., Oliveira J. G., Mendes J.F.F., *Localization and spreading on diseases in complex networks*, Phys. Rev. Lett. 109, 128702 (2012)

[17] Stanley H. E., *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, London, Limited (1971)

[18] Binder K., Heermann D. W., *Monte Carlo Simulation in Statistical Physics*, Springer, Heidelberg (2010)

[19] Henkel M., Hinrichsen H., Lèubeck S., *Non-Equilibrium Phase Transitions no. v. 1*, Springer, London, Limited (2008)

[20] Hinrichsen H., *Non-equilibrium critical phenomena and phase transitions into absorbing states*, Adv. Phys. 49, 815 (2000)

[21] Bak P., *How Nature Works*, Springer-Verlag New York, Inc., New York, NY, USA (1996)

[22] Marro J., Dickman R., *Non-Equilibrium Phase Transitions in Lattice Models* Cambridge University Press, Cambridge, UK (1999)

[23] Dorogovtsev S. N., Goltsev A. V., Mendes J. F. F., *Critical phenomena in complex networks*, Rev. Mod. Phys. 80, 1275 (2008)

[24] Barzel B., Barabási A. L., *Universality in network dynamics*, Nature Physics 9, 2741 (2013)

[25] Stauffer D., Aharony A., *Introduction to Percolation Theory, 2nd Revised Edition*, Taylor & Francis, Philadelphia, USA (2003)

[26] Albert R., Jeong H., Barabási A. L., *Error and attack tolerance of complex networks*, Nature 406, 6794 (2000)

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

[28] Hong H., Ha M., Park H., *Finite-size scaling in complex networks*, Phys. Rev. Lett. 98, 258701 (2007)

[29] Steger A., Wormald N.C., *Generating random regular graphs quickly*, Journal of Combinatorics, Probability and Computing Vol. 8 Issue 4 (1999)

[30] Ferreira S. C., Castellano C., Pastor-Satorras R., *Epidemic thresholds of the susceptible-infected model on networks: a comparison of numerical and theoretical results*, Phys. Rev. E 86, 041125 (2012)

[31] Boguña M., Castellano C., Pastor-Satorras R., *Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks*, Phys. Rev. Lett. 111, 068701 (2013)

[32] Pastor-Satorras R., Vespignani A., *Evolution and structure of the Internet: A statistical physics approach*, Cambridge University Press, Cambridge (2004)

[33] Mata A. S., Ferreira S. C., *Multiple phase transitions of the susceptible-infected-susceptible epidemic model on complex networks*, arXiv: 1403.6670 (not published)