Roma Tre University
Ph.D. in Computer Science and Engineering

Complex Technological Networks: Structure, Dynamics and Algorithms.

Sandro Meloni

Complex Technological Networks: Structure, Dynamics and Algorithms.

A thesis presented by
Sandro Meloni
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Computer Science and Engineering

Roma Tre University
Dept. of Informatics and Automation

March 2011

Committee:
Prof. Stefano Panzieri
Prof. Yamir Moreno

Reviewers:
Prof. Albert Díaz Guilera
Prof. Luis Mario Floría Peralta

To
The Inhabitance of Space in General
And V. R. in Particular
This Work is Dedicated
By a Humble Native of Flatland
In the Hope that
Even as he was Initiated into the Mysteries
Of Three Dimensions
Having been previously conversant
With Only Two
So the Citizens of that Celestial Region
May aspire yet higher and higher
To the Secrets of Four Five or even Six Dimensions
Thereby contributing
To the Enlargement of The Imagination
And the possible Development
Of that most and excellent Gift of Modesty
Among the Superior Races
Of Solid Humanity

Edwin A. Abbott - Flatland: A Romance of Many Dimensions

ai miei genitori e a mio fratello

Acknowledgments

It's hard after three years try to thank all the persons that helped me in this work. So, let's start with the easy ones and then try to remember all of them. My first and greatest thank goes to my family: my father, my mother and my brother. Without them I would be a worst person. I've to be grateful to them for almost everything I am and especially for supporting and believing in me. A huge thank goes to my father, he taught me that the hard work always gives great satisfactions and honour. My mother also deserves a huge thank because she has always approved all my choices (although sometimes I made stupid ones) and taught me that doing what you like is the first rule. It's hard to admit but also my brother deserves a special place... He is my guide since I was a child and I've to say that, although I'm almost thirty, he's still the person I would like to be.

Now it's the time for all my (un)official advisors. First of all Professor Stefano Panzieri that believed in me without even know me. He gave me the opportunity to interact with some of best scientists in the field and travel across the world to become a better researcher. Professor Yamir Moreno; he has been more that a tutor in all these years and a wonderful person. From the first time I went to Zaragoza he has been a guide and a friend. He always helped me and treated me as a peer. For all these reasons he deserves one of my best thanks. Dr Vittorio Rosato that was my first tutor and introduced me to this strange world of networks, he's still my guide and I've to thank him for all he did and still does for me. A special thank goes to Professor Vito Latora I've to thank him for the wonderful and prolific time I had in Catania, for his hospitality and for his inspiring course on complex networks, one of the best course I ever followed. A special mention is for my unofficial fifth advisor Jesus Gómez Gardeñes for his friendship, tutoring, maybe one of the craziest paper I will ever write and many many other stuffs.

I cannot forget the persons that assessed and improved this work: Mario

Floría and Albert Díaz-Guilera. Thank you for our time and advices, I really appreciate.

I've also to thank all the persons that worked with me these years: Alex Arenas, Sergio Gomez, Javier Borge-Holthoefer, Arturo Buscarino, Luigi Fortuna, Mattia Frasca, Andrea Gasparri, Antonia Azzini and all the persons I forgot.

Now it's time to give a very huge thank to a group of special persons that played and still play a crucial role in my life: the entire gastroenterology department of San Camillo de Lellis Hospital in Rieti. Without them this thesis and my life would be very different. A special thank goes to Doctors Maurizio Giovannone and Alessandro Gigliozzi and Nurse Rita Emiliani. They went much further than their professional duties to give me the opportunity to have a normal life and voyage around the world. Thank you.

Now it's time to repair a mistake I made in my previous thesis. In the last thesis I unforgivably forgot one of my best friends and maybe the smartest person I know: Pidi. I decided to continue my studies only to thank you in this page, this is for you.

This time I cannot forget all the friends that followed me in this crazy voyage that is a Ph.D. in Italy: Matteo De Felice, Guido Drovandi, Francesco Corman and Maurizio Di Rocco and all the friends I found at the department of informatics and automation: the entire Robotics lab that put up my strange research and supported me, the entire Networks lab that always offered protection and funny time and all the technical and administrative staff (especially Alessandro, Ivana and Maria). To all these persons I've to add the friend I encountered in the past years: Stefano, Iuri, Angela, Chiara, Giordana and the entire vacanze piemonte group. The years I spent at the university were one of the finniest in my life and I've to thank all of you for this.

As I spent almost half of my Ph.D. around europe it's time to thank all the friends I've encountered in my visits. Let's start from the spanish crew: Sergio "Torone" Perez, Joaquin, Luca, Isabel, Julia, Raquel, Mauro, Carlos, Emanuele, Luce, Jorge and many more. The Sicilian brigade: Alessio, Giuseppe Angilella, Salvo Scellato, Salvo Assenza, Roberta, Enzo, Mattia, Arturo, Violetta, Glorietta, Agata, Antonio. The Enea squad: Elena, Enzo, Alberto, Silvia, Elisa, Antonio.

I'm sure that I forgot at least half of the persons I've to thank. I took the chance to thank them all also for not to get angry for this.

Surely I cannot forget her. She supported me all the time, she suffered when I was away and she always waited my return. I don't know what to do, you are so special... Thank you.

I would like to end this long pages with a phrase from my grandmother Ida: "Macché Sandro lavora?" that sounds like: "is Sandro really working?". When you are so happy to do your job that people thinks you are not working, you are infinitely lucky.
Thank you,
Sandro

Contents

Part I

Introduction

Chapter 1

Introduction

Networks are all around us; from the simplest forms of life to the complexity of our brain. Also ourselves are part of many networks from the social interaction that we engage or as a result of biochemical interactions inside a single cell. Networks have demonstrated to be present not only in social or biological contexts but also in technological systems. The major example is the Internet, maybe the biggest technological network. Other examples include highways, transportation systems, power generation and distribution networks. Thus, the study of the characteristics of the networks is of primary importance for the advance of sciences.

The study of networks historically started as a branch of discrete mathematics known as *graph theory*. The birth of graph theory dates back to in 1736 when the Swiss mathematician Leonhard Euler published the solution to the Königsberg bridge problem (the problem of finding a round trip that traversed each of the bridges of the prussian city of Königsberg exactly once). Since 1736 graph theory has provided a prolific framework for the solution of many practical and theoretical problems such as the optimal flow in a network of pipes or how to fill n jobs by n people with maximum total utility.

Started as a branch of mathematics in the 20th century other fields discover the potentialities of the graph theory applied to real problems. In the early 20's social scientists initiate analysing relationships between social entities exploiting the tools offered by graph theory. Also, in physics interactions modelled as graphs start to emerge.

In the last decade a new way of thinking about networks arose. Thanks to the availability of huge amounts of digital data, computational power and the

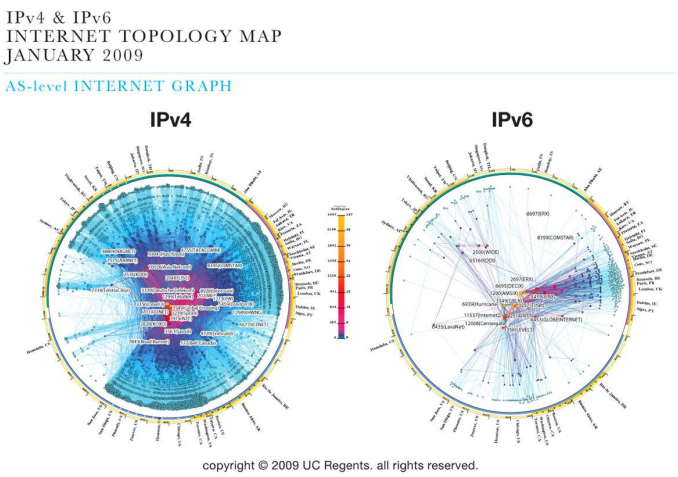

Figure 1.1: This visualization represents macroscopic snapshots of IPv4 and IPv6 Internet topology samples captured in January 2009. The plotting method illustrates both the extensive geographical scope as well as rich interconnectivity of nodes participating in the global Internet routing sytem.For the IPv4 map, CAIDA collected data from 33 monitors located in 30 countries on 5 continents. Coordinated by our active measurement infrastructure, Archipelago (Ark), the monitors probed paths toward 7.4 million /24 networks that cover 95% of the routable prefixes seen in the Route Views Border Gateway Protocol (BGP) routing tables on 1 January 2009.For the IPv6 map, CAIDA collected data from 6 Ark monitors located in 4 countries on 2 continents. This subset of monitors probed paths toward 1,491 prefixes which represent 88.9% of the globally routed IPv6 prefixes seen in Route Views BGP tables on 1 January 2009. Original picture from The Cooperative Association for Internet Data Analysis (CAIDA) $http://www.caida.org/research/topology/as_core_network/$

quickness in communications a different kind of networks has been analyzed. Such networks are defined as *Complex Networks*. As suggested by their name complex networks are composed by a large quantity of units connected in a irregular and sometimes time evolving fashion.

The birth of this new research field was triggered by two papers published at the end of the 90's. The first, by Watts and Strogatz, appeared in Nature in

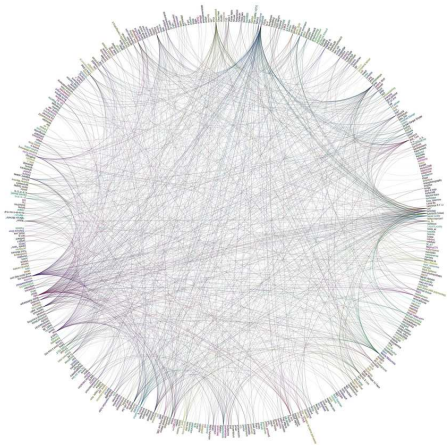

Figure 1.2: Visualization of the connections between users in the flickr *Processing.org* group. Nodes represent users and links friendship between them. Original from the flickr processing.org group *http* : *//www.flickr.com/photos/eskimoblood/*2111672366/

1998 and explained the emergency of a property that has been conjectured in many real world networks, the so-called *small world* phenomenon. Although some real networks can be very large, i.e. the Internet, the distance between two random nodes in most cases remains relatively small, this concept being already known in social sciences from the 60's and expressed by the idea of 6 degrees of separation (the idea that everyone is on average approximately six steps away from any other person on Earth), Watts and Strogatz proposed a model in which starting from a regular network, the random rewiring of a small fraction of the links provokes that the mean distance between nodes rapidly drops to very low values.

In the second seminal paper, appeared in Science in 1999, Barabási and Albert show that in real networks nodes are not equal but few of them have a vast amount of connections and the majority of the nodes are only poorly connected. This peculiar organization has been showed to be present in most

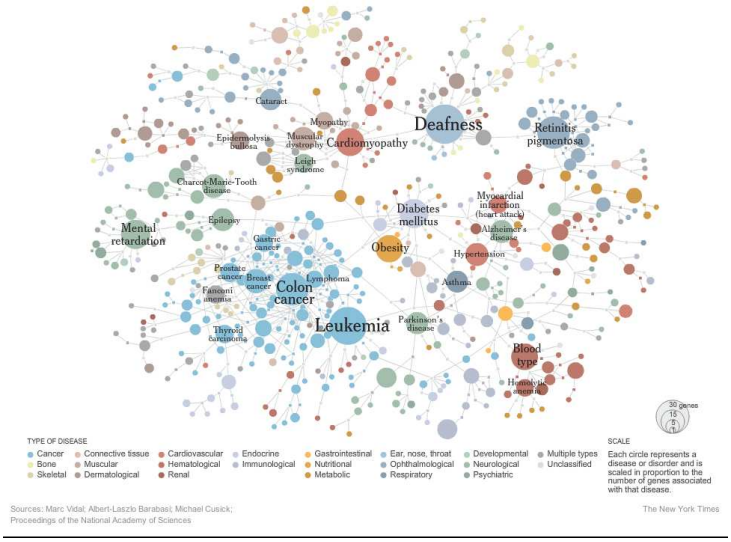

Figure 1.3: The map of the human diseases networks (HDN). Round nodes represent genetic disorders, square nodes are genes and a link between a square and round nodes is drawn if the gene is considered responsible for the disorder. Colors indicate the type of disease. Sizes are proportional to the number of genes associated with the disease.

of the networks we know about and gave the impulse to study networks as a separate field.

The first steps in research on complex networks began with the definition of new tools to characterize the topology of real systems. The most important discovery was the existence of a series of unifying principles common to many networks in very different fields. In this way systems as diverse as phone call networks, acquaintances graphs, the Internet, genetic and metabolic networks, and the actors collaboration networks have been shown to share the same structure and in many cases the same laws governing their functioning. The most striking of these principles is the almost omnipresence of a specific degree distribution function named *power law* that sensibly deviates from the expected poissoninan function. As this particular structure is the result of the evolutionary forces that act over it, one can suppose that the structure must play a key role in the dynamical processes that can take place on it. This supposition

drove the research into a second stage in which the focus changed from the analysis of already formed networks to the understanding of the evolutionary principles that shape real networks structure. At this stage studies highlighted the huge effects of the topology over the dynamics on a network and a series of unexpected results showed that many counter-intuitive behaviors, as the emergence of cooperation in a selfish world, can be explained by the complex interconnections between nodes. This new approach permitted to create a set of theoretical tools to study the behavior of large ensembles of interconnected dynamical units that before have only been observed in an empirical way. In the last years, as usual in every research field, the focus changes from a simply descriptive point of view to a predictive one, in which the knowledge starts to be used to make predictions about the future behavior of the system and to exploit system characteristics to control the overall dynamics. This is the case i.e. of the epidemic spreading in acquaintance networks or congestion control in communication ones, in which effective immunization policies or routing algorithms have been proposed. In figures 1.1, 1.2 and 1.3 some examples of complex networks are presented. Figure 1.1 depicts the map of the Internet at the Autonomous Systems (AS) level for the IP v. 4 and v. 6 protocols. Nodes positions in the graph are related to AS geographical positions and number of connections. Figure 1.2 presents an example of virtual social networks with the connections between users in the Flickr photo-sharing service. In figure 1.3 a biological network is represented in which genetic correlations between human genetic disorders are highlighted [24].

The aim of this work is to analyze, model and control dynamical systems through complex networks theory. Specifically, we will propose a series of dynamical models on large graphs to represent complex and non linear dynamics. We mainly focus on theoretical models that, at a high level of abstraction, are representations of real world technological and social systems. In the cases where it's possible, we will propose an analytical treatment of these systems and we corroborate our predictions via numerical simulations and real data analysis. In particular we will try to predict the dynamical evolution of the systems under study and express the conditions for the emergence of a desired collective behavior. During the work we will present several types of models to illustrate some of the dynamics that are studied in complex networks literature. We will cover four different aspects of networks dynamics and we consider three distinct field of application.

The first is the field of sensor and monitoring networks that nowadays are becoming more and more important in our automated world. In this case we consider two different types of applications. In one we study the robustness of

a static sensor network. Specifically we will present a distributed algorithm for the generation of robust topologies in a two dimensional environment. In the second we look at a spatial environment in which agents can move randomly to simulate mobile sensors and robots. In the second field of application we focus our attention on maybe the most important type of technological networks: communication networks such as the internet and the complex traffic interchange that takes place on its nodes. We will focus on two aspects of traffic dynamics. On one side we will investigate the attributes of traffic flows in such networks. On the other side we will propose an empathetic traffic control strategy that allows for a greater traffic volume to be forwarded to the destination before congestion arises. Finally, in the third applied field, we consider both technological and transportation networks to study how a disease (biological or a computer virus) can spread on a complex environment.

To guide the reader in the logical organization between different applications the work has been separated in three main parts each one covering one type of dynamics on complex networks and two other introductory and concluding parts.

In the first part an introduction to complex networks theory and its tools is given. The concepts introduced are the ones needed to clearly understand the rest of work without a profound knowledge of the field. Specifically, in chapter 2 we present the tools acquired from graph theory and social sciences to analyze the topological structure of complex networks. We will start showing the building blocks of a graph as nodes and links and different graph classes. Then, we will focus on single nodes and links properties like the degree and the strength. We will follow introducing group of nodes properties like the degree distribution of a network and the clustering coefficient. Chapter 2 ends with a presentation of some of the most important models available in the literature for the creation and evolution of complex networks.

Chapter 3 is devoted to introduce the recent advances in modeling dynamics on networks. In this chapter different dynamics will be taken into account covering most of the concepts that will be used in the rest of work and, specifically, the basic models that will be taken as foundations in parts II, III and IV. To study the robustness of a network some basic definitions of percolation theory will be proposed. We also present the state of the art of evolutionary game theory on networks, traffic and congestion dynamics and the most recent advances in disease spreading on large graphs.

Please, note that, although we try to cover the most important concepts in the field, a systematic review of the recent literature is outside the scope of this work. We will just resume the basic informations needed to fully understand

the entire work. We will also provide at the beginning of each part a short intro including the key concepts needed to understand the part alone. For a complete review of the field please refer to [1, 2, 3, 4].

In the second part we start the description of the research activity that represents the center of this thesis work. In this part we concentrate on models for sensor and mobile networks. Specifically, we will cover two aspects: the generation of robust topologies for sensor networks on a 2D space and the study of the conditions that allow the emergence of cooperation in a mobile agents environment.

In chapter 4 we cope with the problem of creating a fault tolerant topology for a sensor network in a distributed fashion. To do so we consider a two dimensional space in which sensor nodes are distributed randomly (i.e. dropped from a plane). In our model nodes have an interaction radius r representing the range of the communication antenna. We propose a distributed algorithm that draws nodes desired degree from a fixed degree distribution and tries to establish the desired number of connections. The degree distribution is chosen to assure high robustness both for random failures and intentional attacks.

In chapter 5 we move a little further considering a 2D environment in which agents can move randomly to study the conditions under which cooperation becomes a convenient strategy. To this end, we examine the behavior of random moving agents that play a typical social game as the prisoner's dilemma and we explicit the insurgency of an all cooperators output as a function of nodes density, velocity and game's parameters.

The third part is fully devoted to traffic models in communication networks. Chapter 6 presents the study of traffic flows characteristics. Through random walk theory we analyze fluctuations in mean flows and predict their value. We also test our theory with numerical simulations and real data analysis. In chapter 7 another fundamental problem for communication networks is addressed: how to optimally drive traffic to cut down the point at which congestion emerges. We present a minimal traffic model to study the emergence of congestion and propose a self adaptive strategy to forward higher traffic values. We also permit neighboring nodes to collaborate on the routing optimization and finally a proof that this local empathetic behavior produces the same results as global optimization is given.

Part IV offers a presentation of the work done in epidemics dynamics. In this part we present three different topics ranging from a fully analytical study of simple disease spreading dynamics to a complex model that incorporates human responses to the disease. The first study is presented in chapter 8. We propose an alternative formulation to the classic Heterogeneous Mean-Field

analysis that can analytically recover the whole disease phase diagram and also predict the single node's infection probability. The proposed approach also allows to study a complete class of infectious dynamics ranging from a contact process to a fully reactive interaction.

In chapter 9 we deepen the study of the effects of non-reactive interactions on epidemic spreading. Two coupled models are presented. To describe nodes interaction a traffic model is introduced, in which packets represents quanta of interaction and infection can be transmitted only if two nodes exchange traffic. We also consider different traffic scenarios and propose an analytical treatment for the study of the epidemic threshold.

Chapter 10 presents a more realistic model in which human responses to the rise of an epidemic are considered. We extend the model proposed in chapter 9 in a metapopulation scenario in which individuals move according to traffic flows and adapt their conduct to the development of the disease. We introduce two types of responses based respectively on cancel a journey or avoid infected nodes and, through an heterogeneous mean-field approach, we derive an expression for the critical mobility threshold.

Finally, in the last part of this thesis the conclusions of the work are drawn and a list of the publications and the other research activities carried out are presented.

Chapter 2

Topology: Complex networks and networked systems

Since this work is fully devoted to the study of dynamical models on top of complex networks in this first part we consider useful to introduce some preliminary notions regarding complex networks' structure and graph theory. Therefore, the aim of this chapter is to present some basic concepts of networks' topology that will be helpful to understand the rest of the thesis. We first introduce the main definitions and the notation used in the work. Then, we discuss some structural properties observed in real networks and the most important models proposed in the literature to represent them.

2.1 Basics definitions and notation

A complex network can be easily represented by a graph and thus the tools offered by graph theory [5, 6, 7] can be used for the mathematical treatment of networks. A graph $G(N, K)$ consists of two sets namely N and K. N is defined as a set of *nodes* (or *vertices*) and K as a set of pair of nodes defined as *links* (or *edges*). The cardinality of the two sets is referred respectively as the number of nodes and links of the graph.

A node is usually named by its index i in the set N and a link by a couple of nodes i and j denoted by (i, j) or l_{ij}. If a link (i, j) between nodes i and j is present in K the link is said to be *incident* on i and j and they are defined as *adjacent* or *neighboring* nodes.

A particular category of graphs is represented by the graphs in which the K

set is composed by ordered pairs of nodes. Such graphs are defined as *directed* graphs in opposition with *undirected* graphs in which nodes order in links is unimportant. In directed graphs links are directed from the first node to the second one. On the other side, in undirected graphs links are bidirectional and the presence of link (i, j) implies the presence of link (j, i). In graph theory a weight or value can be associated with nodes or links, in this case the graph is defined as *weighted*. If multiple links or selfloop are present the graph is defined as a *multigraph*.

The size of a graph is determined by the cardinality of the set N and the number of links may vary from 0 to $N(N-1)/2$. A graph is *sparse* if the number of links is much smaller than the square of the number of nodes ($K \ll N^2$), otherwise the graph is called *dense*. A *subgraph* $G'(N', K')$ of $G(N, K)$ is a graph in which N' and K' are a subset of N and K respectively and all the links in K' only contain nodes in N'.

As this work mainly cope with diffusion dynamics on graphs a key concept is represented by the *reachability* of two nodes. If nodes i and j are not directly connected by a link it may be possible to reach j from i traversing a list of links from i to j. A *walk* can be defined as a list of links that starts in i and ends in j. A walk in which there aren't repeated nodes is defined as a *path*. The length of a path is represented by the number of links that must be crossed to reach the destination and the smallest path from node i to j is the so-called *shortest path*. A *cycle* is a walk starting and ending on the same node.

If for each pair of nodes in a graph exists a path connecting them the graph is said to be *connected*, otherwise the graph is defined as *disconnected*. A *component* of a graph G is a connected subgraph of G. We can also the define *giant component* as a component whose size scales as N.

To represent a graph on a computer system it could be useful to consider a matricial form of the graph G. The *adjacency matrix* $\mathcal{A}$ is a square $N \times N$ matrix which elements a_{ij} are 1 if there is a link between nodes i and j and 0 otherwise. For undirected graphs the adjacency matrix is symmetric and as no self loops are allowed in normal graphs the principal diagonal always contains zeros. This form of representation offers some advantages (i.e. direct access to each element), but some drawbacks are also present, first of all large memory occupancy as the size of memory needed is $\mathcal{O}(N^2)$. The problem is amplified for sparse graphs in which the matrix is mostly composed by zeros.

Other forms of representation include the so called *list of adjacency*: a linked list in which each element represents a link between two nodes. In this case the memory occupancy is $\mathcal{O}(K)$ but some drawbacks are the high time needed to access an element as it is needed to cross part of the list to reach the

desired link.

2.1.1 Nodes degree and degree distribution

The *degree* of a node i is the number of links in which i is involved and can be defined in terms of the adjacency matrix $\mathcal{A}$ as:

$$k_i = \sum_{j \in N} a_{ij} \ . \tag{2.1}$$

In case of directed graphs it is possible to evaluate two components of the degree of a node, namely the number of incoming links (links that end at the node) $k_i^{in} = \sum_j a_{ji}$ defined as *in-degree* and the number of outgoing links (links that start at the node) $k_i^{out} = \sum_j a_{ij}$ defined as *out-degree*; the total degree of a node is referred as $k_i^{tot} = k_i^{in} + k_i^{out}$. The sequence of the nodes degree is named *degree sequence* of a graph.

One of the most important characteristics of a graph is its *degree distribution*. Usually denoted as $P(k)$ the degree distribution represents the probability that a randomly chosen node it has degree k, or in other terms the fraction of nodes with degree k. For directed graphs it could be useful to define the in-degree distribution $P^{in}(k)$ and the out-degree distribution $P^{out}(k)$.

The degree distribution gives important information regarding the structure of the graph and can be analysed graphically (plotting $P(k)$ as function of k) or evaluating the moments of the distribution. The generic $n-$moment of $P(k)$ reads as:

$$\langle k^n \rangle = \sum_k k^n P(k). \tag{2.2}$$

The first moment of the distribution $\langle k \rangle$ denotes the mean degree of the graph and the second moment $\langle k^2 \rangle$ the fluctuations of the distribution. This latter factor plays a crucial role on dynamics that take place on a network.

2.1.2 Shortest path and average path length

In the context of diffusion processes on networks the distance between nodes is a key aspect and more importantly the shortest distance. As expressed previously 2.1, the shortest path is the minimal length of a path connecting two nodes in the graph. To have a compact representation of all distances in a graph it could be useful to use a matrix $\mathcal{D}$ whose entries d_{ij} contain the length of the shortest path connecting i and j. It's important to notice that

if the graph is undirected $\mathcal{D}$ is symmetric and if the graph is disconneted $\mathcal{D}$ can contain elements equal to ∞. The maximum value of $\mathcal{D}$ is defined as the *diameter* of the graph and represent the distance to reach the two extreme nodes in the graph. The mean distance between two nodes is defined as the *average path length* or *characteristic path length* [8] and can be calculated as:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N}, i \neq j} d_{ij} \ . \tag{2.3}$$

The main disadvantage in the calculation of L is represented by its divergence in case of disconnected graphs. To overcome this problem a useful solution is to consider the harmonic mean of the shortest path lengths [9] and derive the *efficiency* of a graph:

$$E = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N}, i \neq j} \frac{1}{d_{ij}} \ . \tag{2.4}$$

This latter definition avoids the divergence of L as every couple of nodes belonging to disconnected components gives a zero contribution to the formula.

2.1.3 Centrality measures

To study the dynamical behavior of networks it's convenient to evaluate the importance of a single node in the graph, as the relevance of a node can be related to different factors i.e. the distance to other nodes or how many paths pass though it, multiple measures have been proposed in the literature and are defined as *centrality measures* [10]. A first measure of the importance of a node is related to the number of connections it has, since a very connected node can reach a huge number of neighbors at the same time, this quantity takes the name of *degree centrality*. In diffusion processes the distance between nodes plays a key role in the dynamics and it could be useful to define the importance of a node as how close it is to the other nodes. Thus, we can define the *closeness centrality* as the mean distance of a node to all the others. Maybe the most important centrality measure is represented by the so-called *betweenness centrality* that measures how important a node is with respect to how many traffic flows pass through it. Mathematically the betweenness centrality is determined by the number of shortest paths that pass through a node over the total number of shortest paths in the graph:

$$b_i = \sum_{j,k \in \mathcal{N},\ j \neq k} \frac{n_{jk}(i)}{n_{jk}} \ , \tag{2.5}$$

where n_{jk} is the number of shortest paths between j and k and $n_{jk}(i)$ the number of shortest paths connecting j and k that pass through i. In the same way the concept of betweenness centrality can be extended to the links and is referred as *edge betweenness centrality* [11].

2.1.4 Clustering

Another typical property of complex networks is represented by the *clustering* (or *transitivity*). In social networks analysis the transitivity represents the fraction of connected triangles that can be found in a graph over the total number of connected triples of nodes. This quantity is defined as T and can be calculated as [1]:

$$T = \frac{3\times \text{ \# of triangles in G}}{\text{\# of connected triples of vertices in G}} \tag{2.6}$$

It is also possible to consider the so-called *clustering coefficient* C [8] of a graph. The clustering coefficient gives the probability that two neighbors of a random node are also connected. To do so a *local clustering coefficient* c_i of node i is introduced as the probability that element $a_{jl} = 1$ given that j and l are connected with i. It can be counted up as the number of links e_i existing in the sub-graph G_i (where G_i is the sub-graph formed by considering the neighbors of i) over the maximum number of possible links in G_i, that is $k_i(k_i - 1)/2$:

$$c_i = \frac{2e_i}{k_i(k_i-1)} = \frac{\sum_{j,m} a_{ij}a_{jm}a_{mi}}{k_i(k_i-1)} \ . \tag{2.7}$$

The global clustering coefficient of the entire graph is given by the average of c_i over all the nodes in G:

$$C = \langle c \rangle = \frac{1}{N}\sum_{i\in\mathcal{N}} c_i \ . \tag{2.8}$$

As c_i is defined as a probability we have $0 \leq c_i \leq 1$ and consequently also $0 \leq C \leq 1$. Sometimes it could be useful to consider the clustering coefficient over degree classes c_k or higher orders cycles as order four or five. It's also possible to consider a measure of clustering based on the efficiency (eq.2.4) and defined as *local effciency*:

$$E_{loc} = \frac{1}{N}\sum_{i\in\mathcal{N}} E(G_i) \ . \tag{2.9}$$

2.1.5 Degree correlations

Many real world networks show a complex structure that sometimes cannot be fully explained by the simple degree distribution. This is the case of the so called *correlated* networks in which the probability that a node with degree k is connected to one node with degree k' depends of k. On the contrary, in *uncorrelated* networks such probability is independent of k.

To highlight the correlations in the graph structure the conditional probability $P(k'|k)$, representing the probability that a node with degree k it points to one with degree k', can be introduced. For uncorrelated graphs can be proved [12, 13] that $P(k'|k) = k'P(k')/\langle k\rangle$ thus independent of k. Although $P(k'|k)$ gives enough information to characterize degree correlations its measure in finite size networks cannot be applied due to its sensitivity to noise. To avoid this problem an indirect measure of $P(k'|k)$ is introduced and is defined as the *average nearest neighbors degree* of a node:

$$k_{nn,i} = \frac{1}{k_i}\sum_{j\in\mathcal{N}_i} k_j = \frac{1}{k_i}\sum_{j=1}^{N} a_{ij}k_j \ , \tag{2.10}$$

this expression can be grouped by degree classes k giving $k_{nn}(k)$ that represents the average degree of neighbors of a node with degree k. The evaluation of $k_{nn}(k)$ as a function of k gives the entity of correlations in the graph. For uncorrelated networks $k_{nn}(k)$ is a constant and its values is $\langle k^2\rangle/\langle k\rangle$. Correlations can be evaluated as the numerical value of the slope ν of $k_{nn}(k)$ as a function of k. If ν is positive nodes tent to connect with nodes of similar degree and the graph is said to be *assortative*, otherwise high degree nodes tent to connect with low degree ones and the graph is said *disassortative* [14]. Another method to estimate degree correlations in a graph is to calculate the Pearson correlation coefficient of the degrees of the nodes at the end of each link as proposed in [14, 15].

2.1.6 Motifs and community structures

To fully investigate the internal structure of a graph it is possible to consider complex measures that not only involve a single node (i.e. degree) but group of nodes and how they are connected. A first measure of this type is represented by *motifs* analysis. A motif M is a pattern of connections occurring in a significantly higher number than in randomized versions of the graph. A randomized version of a graph is a graph with the same degree distribution

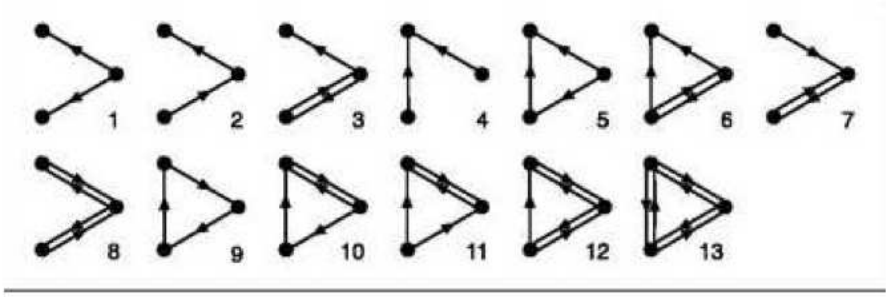

Figure 2.1: All the possible 3-noded motifs on a directed network.

of the original graph but with links randomly rewired. A motif is defined as a n-nodes connected sub-graph both directed or not. To asses the statistical relevance of a motif the *Z-Score* is introduced as:

$$Z_M = \frac{n_M - \langle n_M^{rand} \rangle}{\sigma_{n_M}^{rand}} \ , \tag{2.11}$$

where n_M is the number of times M appears in the original graph G, $\langle n_M^{rand} \rangle$ is the average number of appearances of M in the ensemble of the randomized versions of G and $\sigma_{n_M}^{rand}$ its standard deviation.

Considering larger group of nodes and different connections it is possible to extend the concept of motif to a wider class defined as *community structures*. Communities where firstly introduced in sociology to represent highly connected groups of friend or families in acquaintance networks [10]. A first definition of community in a graph can be as follows: given a graph G a community is a sub-graph G' of G whose nodes are tightly connected. Although this is a formal definition it includes an uncertainty given by the definition of tightly connected nodes giving rise to distinct definition of community in a graph. In the most strict form a group of tightly connected nodes can be seen as a group in which every node is connected to all the others. Such group is defined as a *clique*. Other definitions can be derived considering not only adjacent nodes but just nodes at distance at least n [10].

Another class of definitions are based on the concept that in a community most of the links are directed inside the community and only a small fraction are directed outside it [16].

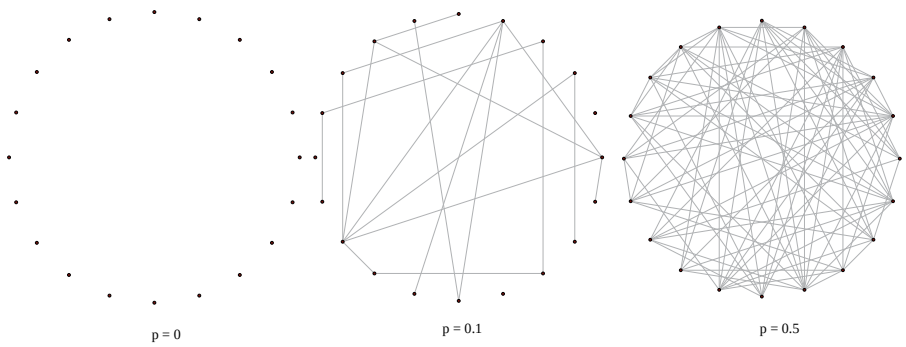

Figure 2.2: Examples of ER graphs generated with three different connection probabilites $p = 0$, $p = 0.1$ and $p = 0.5$

2.2 Networks models

Once the most important characteristics of graphs for complex networks analysis have been introduced, we now focus on the mathematical modeling of real world networks. In this section we present some generic network models with special attention to their properties and construction algorithms.

2.2.1 Erdös and Rényi random graphs

The first approximation of a real network is represented by the so called *random graphs*. The adjective random refers to the non-regular organization of links in the graphs opposite to regular graphs in which each node has a fixed number of links. The mathematical study of this class of graphs was firstly initiated by Solomonoff and Rapoport [17], but the first systematic analysis was proposed by Erdös and Rényi in 1959 [18]. Erdös and Rényi propose a model to generate random graphs with exactly N nodes and K links. We denote such graphs as *Erdös and Rényi* (ER) *random graphs* $G^{ER}_{N,K}$. To generate this kind of graphs it is possible to start with N disconnected nodes and then draw K links between randomly chosen couples of nodes avoiding self and multiple links. It's important to notice that in this case only a possible graph in the statistical

ensemble of all the $G^{ER}_{N,K}$ graphs is created. Another definition (that coincide with the previous for $N \to \infty$) can be obtained considering N nodes and the probability p of connecting two random nodes. The graph can be generated creating N nodes and connecting each possible couple of nodes with probability p. This procedure creates a different ensemble of graphs $G^{ER}_{N,p}$ characterized by different number of links; a graph with K links is obtained with probability $p^K(1-p)^{N(N-1)/2-K}$.

The structural properties of ER graphs vary as a function of p and their behaviour changes at $p_c = 1/N$ corresponding to the critical degree $\langle k \rangle = 1$. In particular it can be proved [5] that for $p < p_c$ the graph almost surely has only components of size $\mathcal{O}(lnN)$. For $p = p_c$ the larger component has size $\mathcal{O}(N^{2/3})$ and for $p > p_c$ a giant component $\mathcal{O}(N)$ appears.

The degree distribution of ER graphs was first studied by Bollobás [19]. In particular a node i has $k = k_i$ links with a binomial probability $P(k_i = k) = C^k_{N-1}p^k(1-p)^{N-1-k}$, where p^k represents the probability that k links are adjacent to the node, $(1-p)^{N-1-k}$ is the probability for the absence of the remaining $N-1-k$ links, and $C^k_{N-1} = \begin{pmatrix} N-1 \\ k \end{pmatrix}$ is the number of different ways of selecting the end points of the k edges. Averaging over degree classes k gives the degree distribution $P(k)$ that for $N \to \infty$ is well approximated by:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \ . \tag{2.12}$$

Given the degree distribution it's also possible to calculate other properties of the ensemble [5]. The diameter of the graph varies around $Diam = \ln N / \ln(pN) = \ln N / \ln\langle k \rangle$. The average path length L consequently scales as $ln(N)$ leading to very small networks also for large values of N. Regarding the clustering coefficient, it can be demonstrated [8] that $C = p = \langle k \rangle / N$. This means that C vanishes in the large N limit.

2.2.2 Watts and Strogatz small world networks

One of the most important properties of real networks is that despite of their huge size the average distance between nodes remains relatively small. This feature is know as *small world* phenomenon. Specifically the term small world refers to graphs with high clustering coefficient and small average path length. The *Watts and Strogatz* (WS) model is a method, proposed by Duncan Watts and Steven Strogatz [8], to create graphs with small average path length and high clustering coefficient. The algorithm starts with a N nodes ring in which

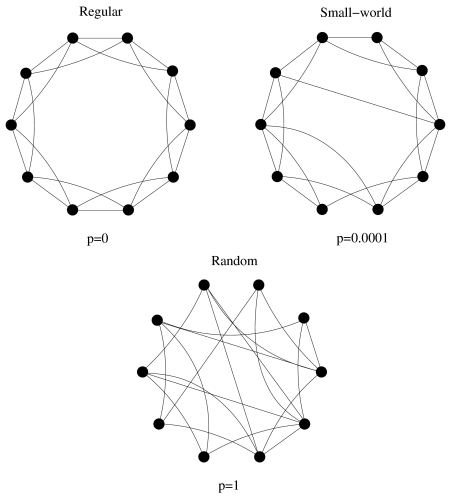

Figure 2.3: Diagram of the random rewiring procedure for interpolating between a one-dimensional lattice and a random network in the Small-world model. Three rewiring probabilities are showed: $p = 0$, $p = 0.0001$ and $p = 1$. The networks have N = 10 nodes and k = 4.

each node is connected with its $2m$ nearest neighbours leading to $K = mN$ links. After this step a rewiring procedure is applied in which with probability p each link (considered only once) is rewired and with probability $1 - p$ is preserved. For $p = 0$ a regular lattice is obtained with high C but large values of L. For $p = 1$ a random graph similar to ER graphs with minimum degree m is obtained and for intermediate values of p there exists a small region in which the graphs show the small world property. The degree distribution in WS graphs also varies between a delta function centered at $2m$ with $p = 0$ to a poissonian distribution similar to ER graphs with $p = 1$. For different values of p the degree distribution reads as [20]:

$$P(k) = \sum_{i=0}^{\min(k-m,m)} \binom{m}{i} (1-p)^i p^{m-i} \frac{(pm)^{k-m-i}}{(k-m-i)!} e^{-pm}$$

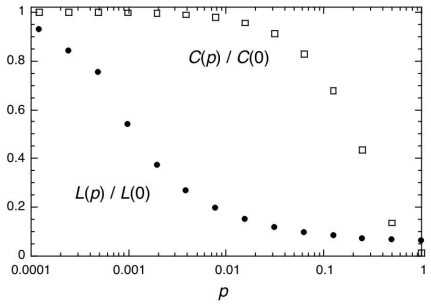

Figure 2.4: Characteristic path length $L(p)$ and clustering coefficient $C(p)$ for the family of randomly rewired graphs described by the small-world model as a function of the rewiring probability p. Original figure from [8].

2.2.3 Scale-Free networks

Recently [129] it has been showed that a huge amount of networks ranging from biological to social and technological networks are characterized by few nodes with very large degree, absolutely improbable in poissonian degree distributions as in the ER graphs. These nodes are called *hubs* and the degree distribution of such networks is well approximated by fat-tailed distributions, specifically in most cases it follows a power-law degree distribution in the form $P(k) \sim Ak^{-\gamma}$ with $2 \leq \gamma \leq 3$. Such graphs are named *Scale-Free* (SF) networks [2, 129] because power-laws have the property of having the same functional form at all scales. Scale free networks represent the most important model in complex networks analysis and due to their peculiar structure they have profound impact on the dynamic that run on top of them.

In the last decade a plethora of models to produce synthetic scale-free networks has been presented, ranging from static [22] to evolving ones [129] or from fitness based [23] to hyperbolic mapping of geometrical spaces [21]. One of the most important is the one proposed by Barabási and Albert [129].

The *Barabási-Albert model* (BA) is based principally on two concepts: the growth and the so-called *preferential attachment*. The model is inspired to the growth of the World Wide Web in which pages with a high number of links gain new connections more rapidly than low-connected pages. The algorithm to create a BA scale-free network $G^{BA}_{N,K}$ is the following: starting with a core of m_0 connected nodes, at each time step $t = 1, 2, 3, 4 \ldots, N - m_0$ add a node

j with $m \leq m_0$ new links. To choose a node i to connect j the preferential attachment mechanism is used. In preferential attachment the probability to choose a node i for a new connection is proportional to its degree k_i:

$$\Pi_{j \to i} = \frac{k_i}{\sum_l k_l} \ . \tag{2.13}$$

In the model every new node has m links thus at time t the network will have $N = m_0 + t$ nodes and $K = mt$ links, the average degree is $\langle k \rangle = 2m$.

The BA model in the infinite time limit $t \to \infty$ produces graphs with power law degree distribution and a fixed $\gamma = 3$. The model has been solved analytically via a mean-field approximation [129, 25], a rate equation approach [26] and a master equation approach [27]. Also in the BA model the mean distance between nodes grows logarithmically with N but, in contrast with ER graphs with a double logarithmic correction so fixed N and K the mean distance L is smaller in BA networks than ER graphs. The clustering coefficient scales as the inverse of the size of the network $C \sim N^{-1}$ leading to a vanishing clustering in the infinite size limit. To obtain scale-free networks with a non zero clustering various models have been presented one of the most important is the so-called *Holme and Kim* model [36] in which the preferential attachment mechanism is used to create the first of the m links and the remaining links are chosen with probability p through a mechanism defined *triad formation* that create triangles in the network choosing as end of a link one neighbor of the node chosen through the preferential attachment.

2.2.4 Generalized Random graphs

Although the presented models can reproduce some of the features of real systems they are specific and sometimes a generalization of the concept of random graphs to graphs with an arbitrary degree distribution could be useful.

The *configuration model* [37] represents one of these generalizations of random graphs allowing to create graphs from a given degree sequence. A degree sequence is a sequence of integer numbers in which each number represents the degree of a node $D = k_1, k_2, k_3, \ldots, k_N$ where $\sum_i k_i = 2K$ with K the number of links in the graph. Note that D can be chosen to approximate the desired $P(k)$ for large N. Graphs created using the configuration model are sampled by the ensemble $G^{conf}_{N,D}$ of all the graphs with N nodes and D degree sequence chosen with equal probability. To create a random graph from D and N it's possible to create N nodes and assign to each node i a number of half-edges equal to its degree k_i and then connect each half-edge randomly. Although this

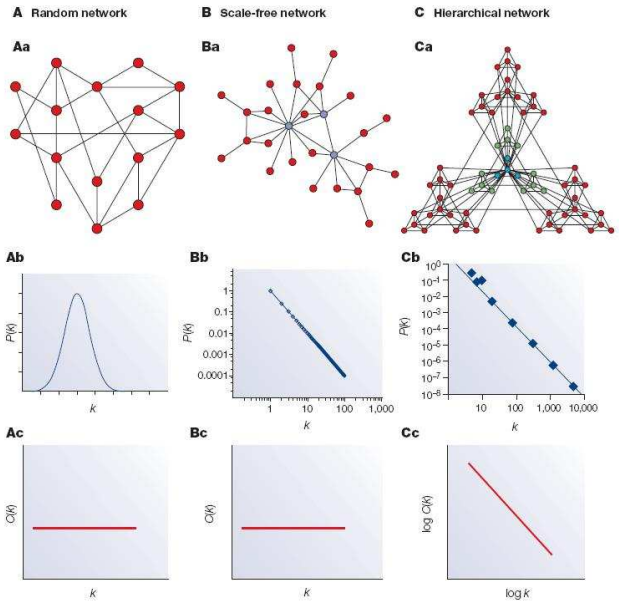

Figure 2.5: Examples of complex networks. *A)* a random graph, *B)* a scale-free network, *C)* a hierarchical network in which a pattern of connections is repeated at different scales. For the three examples the degree distribution $P(k)$ and the clustering coefficient as a function of the degree $C(k)$ is reported. Original figure from [24].

algorithm is quite naive it allows to create a wide range of random graphs from ER graphs to scale-free networks.

Given its simplicity the configuration model permits to calculate some of the topological features analytically. The condition for the formation of a giant component given a degree distribution $P(k)$ has been found by Molloy and Reed [38, 39]:

$$Q = \sum_k k(k-2)P(k) > 0 \ . \tag{2.14}$$

It's worth noting that in case of ER graphs the formula gives the critical threshold at a value equal to $\langle k_c \rangle = 1$ that is the critical degree already observed

in sub-section 2.2.1. An approximate expression for the average path length was found by Newman et al. [40] when $N \leq z_1$ and $z_2 \leq z1$:

$$L = \frac{\ln(N/z_1)}{\ln(z_2/z_1)} + 1 \ , \tag{2.15}$$

with z_1 and z_2 being the mean number of first and second neighbors respectively. The clustering coefficient is given by [28]:

$$C = \frac{\langle k \rangle}{N} \left[\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle^2} \right]^2 = \frac{\langle k \rangle}{N} \left[\left(\frac{\sigma_k}{\langle k \rangle} \right)^2 - \frac{\langle k \rangle - 1}{\langle k \rangle} \right]^2 \ , \tag{2.16}$$

this value is the same that in ER graphs multiplied by an extra factor. It's important to notice that such extra factor can assume high values especially in case of power law degree distributions and although C vanishes in the large N limit, for finite graphs it can be non negligible.

2.2.5 Random Geometric Graphs

A special attention is needed in case of modeling spacial networks, in which nodes can be placed in a D-dimensional space. The most simple model of spacial networks is the *random geometric graph* (RGG) model. A RGG is constructed placing N nodes randomly on a D-torus or a plane and nodes within an interaction radius r are connected. Such graphs represent a fundamental model for many real world applications like sensor networks or cities. All the features of RGGs strictly depend on the density of nodes ρ in the space and the interaction radius r. In particular, considering a d-dimensional space of volume V_d the density ρ is defined as $\rho = N/V_d$ and the mean degree reads as $\langle k \rangle = \rho \pi r^2$. RGGs are characterized by the same degree distribution of ER graphs (see 2.2.1):

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \ , \tag{2.17}$$

and also the size of the giant component as function of $\langle k \rangle$ undergoes a second order phase transition. The clustering coefficient in RGGs can be easily calculated considering the density of nodes and the so-called *excluded volume* $V_{ex} = 2^d V$ [41] :

$$C = \frac{1}{V_{ex}} \int_{V_{ex}} \rho(r) dV. \tag{2.18}$$

Chapter 3

Dynamical models on complex networks

The main goal of complex networks study is to model, describe and predict the different dynamics that can take place on top of real networked systems. Thus, once a short review on the main topological feature of complex networks have been presented we can now focus on dynamical systems that can be seen as a network.

The aim of this chapter is to introduce recent approaches presented in the literature in modeling networked systems. Specifically, in this part we give a brief summary of the concepts needed to understand the results that will be presented in the upcoming chapters. A special focus will be dedicated to diffusion models in complex networks both on a information transmission and epidemic spreading perspective. Also a brief resume of the latest advances in modeling evolutionary dynamics on networks and some basics concepts of percolation theory will be given.

3.1 Percolation Theory

The percolation theory provides a suitable framework to analytically investigate the robustness of a network, i.e., the ability of a network to properly operate even when a fraction of its components is damaged [1].

Strictly speaking, the percolation theory is a general mathematical theory of connectivity and transport in geometrical complex systems. Percolation is of particular interest to physicists as it can be considered the simplest model of

a disordered system capable of experiencing a phase transition. A remarkable aspect of percolation is that many results can be often encapsulated in a small number of simple algebraic relationships. For a comprehensive introduction to the percolation theory the reader is referred to [80].

A standard percolation process can be, in general, of two types: site or bond. Site percolation on a given graph means that the vertices are empty with a given probability f (or occupied with a probability $p = 1 - f$), while bond percolation refers to the existence or not of an edge between two arbitrarily chosen nodes. Once the random deletion (or placement) of nodes or edges is done, several quantities allow the characterization of the network properties. In particular, it is possible to look at the existence and size of the giant component as a function of f, and at the average size and fluctuations in the size of finite components. In this way, it can be defined a critical probability f_c below which the network percolates, i.e., it has a giant component, and a set of critical exponents characterizing the phase transition. The exact value of such a threshold f_c depends on which kind of grid (graph) is considered and its dimension. Percolation theory gives an analytical framework for the study of failures or attacks on a network in general. During the last decades some exact results have been proposed for special types of graphs such as one and two dimensional lattices, cayley trees and a general criterion for study networks robustness. In 1998 Molloy and Reed [75] defined a criterion for the appearance of the giant component in a graph with generic degree distribution $P(k)$ only analyzing its first $\langle k \rangle$ and second moment $\langle k^2 \rangle$. The Molloy and Reed criterion has been used by Cohen et al.[65][66] to give a general form for the percolation threshold f_c both for random failures and intentional attacks

3.2 Evolutionary dynamics

A *game* or *strategic game* is a mathematical framework to model social or economical interactions. Specifically in a strategic game two or more agents (that represent decision makers) are involved. Each agent can choose between two or more different actions. These actions are defined *strategies*. The agents interact through the actions and the goal of each agent is to choose the sequence of actions to maximize its gain (also defined as *payoff*). These simple rules allow to model a wide range of real systems from social to economical and biological ones and, although only a schematic representation of real interaction is used, complex and unexpected behaviors can emerge.

The so-called *game theory* have its origins in the work by J. von Neumann

and O. Morgenstern [29]. Historically the study of game theory is divided in two approaches: the *classical game theory* in which each agent is expected to act as a *rational player* and strategies are fixed and the *evolutionary game theory* in which each agent can change its strategy dynamically and copy successful agents strategies. Both in classical and evolutionary game theory interactions take place in a well-mixed population but complex behaviors can be obtained considering interactions limited to adjacent nodes in a graph.

The most studied game in game theory is the *Prisoner's Dilemma* (PD) [30]. The game is defined in the following way: Two agents namely A and B play repetitively . Each player has two options: to cooperate or to defect and a payoff is associated with each couple of actions. If both players cooperate (defect), both receive one (zero) point. If one player cooperates and the other defects, the cooperator scores zero points while the defector scores b points, with $b > 1$. Table 3.1 depicts a schematic representation of the payoff matrix in the PD.

	Cooperate	Defect
Cooperate	R, R	S, T
Defect	T, S	P, P

Table 3.1: Matrix representation of the Prisoner's Dilemma payoffs. T stands for Temptation to defect, R for Reward for mutual cooperation, P for Punishment for mutual defection and S for Sucker's payoff. in the prisoner's dilemma, the following inequalities must hold: $T > R > P > S$

In a spacial version of the PD interactions are given by a graph as a regular lattice [31]. Each player plays an iterated PD game against its neighbors and a total payoff for each player is the sum of the payoffs accumulated during the encounters with its neighbors. In an evolutionary version at the end of each round of interactions a player can copy the strategy of one of each individual that has earned a higher payoff with probability $1-p$, and, to model occasional irrational moves, the opposite strategy with probability p.

In complex networks versions of the PD [32, 33] an enhancement in the levels of cooperation is usually observed and this is due to the effects of inhomogeneous degree distributions [35, 34].

3.3 Traffic models

One of the most important roles of technological networks is to efficiently transport goods, information, people from one part to another of the network. For this reason traffic and congestion models in complex networks are of extreme interest in computer science, engineering and physics. In particular, information networks such as the Internet or sensor networks are of fundamental importance. The aim of such models is to understand the fundamental characteristics of traffic dynamics and the effects of topology on the network delivery capacity and the onset of congestion.

The first study regarding information traffic in complex networks focused on the transition between a free flow regime, in which a low traffic can be handled by the nodes, and a congested phase in which the size of nodes queues start to grow indefinitely. This first model was proposed by Ohira and Sawatari [92] and two classes of nodes are present: *hosts* which can generate and receive traffic and *routers* that can only store and forward messages. At each time step p new packets are injected in the system with random origins and destinations and each node has an unlimited queue managed with a *first in-first out* (FIFO) policy. Packets follow a shortest path route and in case of two equivalent paths two alternatives are possible: the *deterministic* routing in which the less congested node is chosen and the *probabilistic* routing in which more congested nodes can be chosen with a tunable probability. In the model the topology is a 2D regular lattice with host nodes at the boundaries. Although this model is very different from real communication networks such as the Internet it shows some relevant traffic properties observed in real traffic analysis [42]. The most significant results is related to the sudden transition from free flow to congested state as the traffic p increases. This transition recalls a second order phase transition in statistical mechanics and it's characterized by a critical traffic rate p_c that depends on the routing strategy adopted, demonstrating that an optimal routing is crucial to accommodate high traffic values without congestion.

A plethora of modifications to this model have been presented considering random locations for hosts and routers [93] and traffic dependent packets creation rates [94]. In particular this latter models resemble some traffic control strategies present in actual communication networks [42]. In the same study [94] the authors demonstrate that as a results of this control strategy the system self-organize at the critical point of the transition. The same authors studied the effects of topology on traffic delivery capacity [43] using as a substrate a real Internet topology. The effects of topology on the onset of congestion have

been widely analysed [44, 45, 46, 47] demonstrating that although scale-free networks have small distances between nodes most of the shortest paths pass through main hubs overloading them and promoting the onset of congestion. To avoid the hubs overload a collections of congestion aware routing models have been proposed [48, 49, 50, 51] in which the quality of a channel q_{ij} is expressed as a function of the number of packets n_i currently at one node $q_i = f(n_i)$. Where $f(n)$ is defined as follow: for $n = 0$ is equal to 1 leading to a high quality of the channel and an easy moment of packets, for $n > 0$ takes the form $f(n) = n^{-\xi}$ with $\xi \geq 0$ leading to lower quality as n increases.

The effect of different topology measures on traffic aware models have been investigated by Echenique et al. in [52, 53]. In these latter models a congestion avoiding routing scheme is presented. In particular packets are delivered not considering the topological shortest path but, at each time step the next hop j for a packet with destination t is chosen as the one that minimizes the effective distance δ_j, defined as:

$$\delta_j = hd_{jt} + (1-h)n_j, \quad j = 1, \ldots, k_i \ , \tag{3.1}$$

where d_{jt} is the variation in the distance from the actual node i and t passing through node j (note that this quantity can only be $d_{jt} = -1$ if the node is on the shortest path from i to t, 0 if j is at the same distance from t and 1 if j is further from t), n_j is the size of the queue of j and h is a tunable parameter. When $h = 1$ packets follow the topological shortest path, when $h \neq 1$ a traffic aware routing mechanism is adopted allowing packets to follow longer but less congested paths. At each time steps p packets are introduced in the system and the number $A(t)$ of packets that have not reached their destination at time t is measured. For low traffic values both models (the shortest path, and the traffic aware) show a stationary state for $A(t)$ corresponding to the free flow state. As the traffic p increases two different behaviors for $A(t)$ are observed: in the shortest path model $A(t)$ grows linearly with p and in the traffic aware scheme a rapid and non linear increase of $A(t)$ is observed. To study the phase transition related to this behavior an order parameter $\rho = \lim_{t\to\infty} \frac{A(t+\tau)-A(t)}{\tau p}$ is introduced, where τ is the observation time. Substantially, ρ measures the difference between the number of packets that are injected in the system and the ones who reach their destination. For $\rho = 0$ all the packets are delivered in the time window τ (i.e. the free flow state). Otherwise when $\rho \to 1$ no packet reaches the destination: the system is congested. The study of ρ as a function of p shows important differences between the shortest path and the traffic aware routing. In the shortest path the transition occurs at relatively

low rates and the transition is continuous resembling a second order phase transition. In the traffic aware regime, instead, the congestion arises at higher traffic values but an abrupt transition is present recalling a first order phase transition. An analytical confirmation has been proposed by De Martino et al. [117, 118] in which authors present a minimal traffic model that can be solved analytically. Although the model is very simple it includes all the features of previous models as the dependence from network topology and traffic aware routing schemes. In particular, the congestion aware routing is implemented through a probability η that represented the probability that a packet will be refused if the receiving node is already congested. Analytical results confirm the shift of the critical traffic value p_c in presence of the congestion aware routing and the change in the transition from smooth to abrupt.

All these findings demonstrate the high influence of topology and routing schemes on traffic dynamics that can be exploited to design next generation communication networks.

Another fundamental topic in traffic analysis is represented by the study of fluctuations in mean flow over nodes or links. This kind of studies have been motivated by trying to predict the traffic on desired node (or link) to better design network infrastructure. A first analysis has been conducted in [195, 196] where the authors claim the existence of two universality classes for the relation between the mean flow $\langle f_i \rangle$ on a node and its fluctuations σ_i. The two classes lead to a simple scaling law $\sigma_i \sim \langle f_i \rangle^\alpha$ where α can take only two values $1/2$ or 1. The authors propose two models based on random diffusion and shortest path to explain the different behaviors and conclude that the two universality classes are due to the effects of external and internal dynamic fluctuations respectively. A different approach is exposed in [91]. In this case the authors propose a model based on queueing theory with exponential arrival and service time and claim that the scaling exponent α can vary continuously between $1/2 \leq \alpha \leq 1$.

3.4 Epidemic spreading

The study of how a disease spreads in a population is a fundamental topic in medical research. Since the 20th century attracted a lot of attention from mathematicians and, nowadays the mathematical modeling of infectious diseases is a key concept in epidemiology. Physicists and engineers entered the field when the similarities between the spreading of a disease and a percolation process were en-lighted [54].

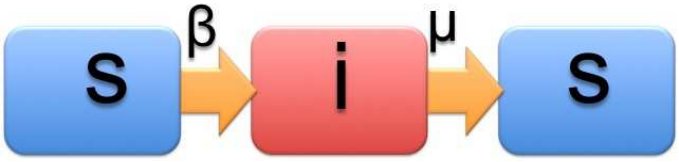

Figure 3.1: Schematization of an SIS process with β as infection probability and μ as recovery rate

The first application of epidemic modeling on complex interaction topologies is due to Pastor-Satorras and Vespignani [163, 164] that in a couple of seminal papers analysed the effects of network connections on the rate and diffusion patterns of a disease.

A fundamental building block in modeling infectious diseases is represented by the so-called *compartmental models*, in which the population is divided into different groups each representing a possible state of the disease [128, 160]. In the most simple case population can be divided into two groups: *susceptible* (S) healthy persons that can catch the disease if in contact with infected individuals and *infected* (I) persons that currently have the disease and can transmit it to the others. Within this framework, or adding additional states like the recovered (R) persons that have been infected and now are cured (or died), it is possible to model a variety of different diseases.

One of the simplest epidemiological models describe diseases that can only be caught once and end up in a immunization or death of the infected is the SIR. The model is based on two parameters, the transmission rate β, and the recovery rate μ. At the beginning of the spread an initial seed I_0 of infected individuals is inserted in the population and a susceptible individual i if in contact with an infected j is infected with probability β:

$$S(i) + I(j) \xrightarrow{\beta} I(i) + I(j) \tag{3.2}$$

and an infected becomes recovered at rate μ:

$$I(i) \xrightarrow{\mu} R(i) \ . \tag{3.3}$$

It's also possible to model diseases that don't give immunization to their survivors, such as the tuberculosis and the common cold. These diseases are well described by the SIS model in which an infected returns in the S state at

rate μ. In the simplest case the spreading of the disease is considered much faster than mean lifespan of an individual so birth and death rates are not taken into account but it's possible to model more complicated scenarios in which births and deaths are considered or other classes such as *latent* (L, individuals that are infected but are not infectious).

The two presented models (SIR and SIS) although very similar lead to a totally different behavior. In the SIS two possible steady states are possible: one with $I = 0$ in which the disease is absorbed by the system and no real outbreak in the population takes place or an endemic state $I > 0$ in which the infected population reaches a stationary state and the disease propagates indefinitely. In the SIR model at the final state the number of infected is always zero and in this case there are two possible outcomes: the disease didn't produce an outbreak and the final recovered population is near the value of the initial seed, or the disease propagated to a finite fraction of the population. Even though the dynamical behavior of the two models is very different, in both cases the two parameters β and μ (or more correctly their ratio $\sigma = \beta/\mu$) play an important role in the presence or not of an epidemic outbreak. In particular we are interested in predicting the critical point σ_c at which the *epidemic transition* from the absorbing phase in which the disease cannot create a finite outbreak and, to the endemic phase in which a finite fraction of infected (or removed in SIR) occurs. To get some initial insight on the value of the critical point and the nature of the epidemic transition it's possible to consider a simple scenario defined as *homogeneous mixing*.

In the first approximation, both the SIR and the SIS models are considered within the homogeneous mixing hypothesis [128], meaning that the contacts between individuals are chosen randomly from the entire population. Although this strong approximation doesn't consider any geographical or local detail, it permits to represent the system as a set of ordinary differential equations for the densities of individuals in each class. For the SIR model the equations are:

$$\begin{aligned} \frac{ds(t)}{dt} &= -\beta\overline{k}\rho(t)s(t) \\ \frac{d\rho(t)}{dt} &= -\mu\rho(t) + \beta\overline{k}\rho(t)s(t) \\ \frac{dr(t)}{dt} &= \mu\rho(t) \end{aligned} \tag{3.4}$$

where $\overline{k}$ is the number of contacts in the unit time (that is fixed for all individuals) and $s(t)$, $\rho(t)$, $r(t)$ are, respectively, the fraction of susceptible,

infected, and removed individuals at time t. Note that the normalization condition $s(t) + \rho(t) + r(t) = 1 \ \forall t$ also hold. Equations 3.4 can be explained in the following way: susceptibles become infected at a rate that is proportional to the infection probability β, the densities of infected, susceptibles and the number of contacts per unit time; infected decay into recovered at rate μ. It's important to notice that β and μ are fixed and equal for all the contacts. The system in eqs. 3.4 can be solved analytically and predicts the presence of a non-zero epidemic threshold β_c for the outbreak of the disease. In particular by considering the so-called *epidemic incidence* (indicated as $r_\infty = \lim_{t\to\infty} r(t)$) if $\beta > \beta_c$ the value of r_∞ assumes a finite value, otherwise r_∞ is infinitesimally small in the very large population limit. Solving the model for a small infection seed as initial conditions ($s(0) \simeq 1$, $\rho(0) \simeq 0$, $r(0) = 0$) leads to a simple expression for the epidemic threshold $\beta_c = \overline{k}^{-1}$ (a complete analytical derivation of the results can be found in [56]). Specifically, it can be proved that β_c acts as a critical point in phase transitions, where r_∞ is the order parameter and β the control parameter. Another important relation can be established with a very fundamental concept in epidemiology: the *basic reproductive number* R_0 defined as the number of secondary infections generated by an infected individual in an otherwise fully susceptible population and obviously only values of R_0 larger than one imply the survival of the disease. In the case of homogeneous mixing R_0 is given by $R_0 = \beta\mu\overline{k}^{-1}$.

Although the homogeneous mixing hypothesis has good analytical properties in order to obtain realistic previsions it must be abandoned and a contact patterns must be considered. Recent studies in sexually transmitted diseases have revealed the high heterogeneity of sexual contacts networks that profoundly affects the epidemic dynamics [128, 154, 155].

A first step toward the consideration of contact networks came from Grassberger [54] that demonstrated the equivalence between a spreading process on a network and a percolation-like process on the same graph. The first results on highly skewed degree distribution is due to Moreno, Pastor-Satorras and Vespignani that via a mean-field analysis solved a modified version of SIR on heterogeneous graphs with a generic degree distribution $P(k)$ and a finite average connectivity $\langle k \rangle$. Let's consider the quantities $s_k(t)$, $\rho_k(t)$ and $r_k(t)$, which are the densities of susceptible, infected, and removed nodes in the degree class k at time t. We can also write the normalization condition as:

$$s_k(t) + \rho_k(t) + r_k(t) = 1 \ , \tag{3.5}$$

and express the global values of the epidemic incidence by the average over

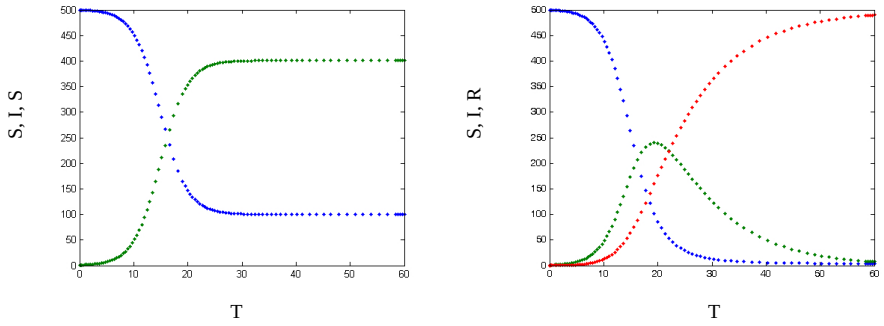

Figure 3.2: Temporal evolution of the number of Susceptible (blue), Infected (green) and Recovered (red) in the SIS (left) and SIR (right) models respectively. Original picture from *http* : *//en.wikipedia.org/*.

the various connectivity classes: $r_\infty = \lim_{t\to\infty} r(t)$, with $r(t) = \sum_k P(k) r_k(t)$. At the mean-field level, these densities satisfy the same set of coupled differential equations as in Eqs. (3.4), but differentiated by connectivity classes:

$$\begin{aligned} \frac{ds_k(t)}{dt} &= -\beta k s_k(t)\Theta(t) \\ \frac{d\rho_k(t)}{dt} &= -\rho_k(t) + \beta k s_k(t)\Theta(t) \\ \frac{dr_k(t)}{dt} &= \rho_k(t) \end{aligned} \tag{3.6}$$

where μ is set to one, and $\Theta(t)$ represents the probability that any given link points to an infected site [163, 164]:

$$\Theta(t) = \frac{\sum_k kP(k)\rho_k(t)}{\langle k\rangle}. \tag{3.7}$$

Solving the system 3.6 in a similar way than 3.4 (see [159] for details) it's possible to obtain a condition for the epidemic threshold $1/\langle k\rangle \sum_k kP(k)(\beta k) = \beta\langle k^2\rangle/\langle k\rangle > 1$, leading to the value for the epidemic threshold:

$$\beta_c = \frac{\langle k\rangle}{\langle k^2\rangle} \tag{3.8}$$

This latter result has profound implications in highly heterogeneous network. In fact for graphs in which $\langle k^2 \rangle < \infty$ as ER graphs the threshold has a finite value and a standard phase transition is observed, but for graphs with highly fluctuating degree distributions $\langle k^2 \rangle$ can assume high values and in some cases like scale free networks with $2 < \gamma \leq 3$ can diverge leading to a vanishing epidemic threshold for the $N \to \infty$ limit. Although a real zero threshold is impossible in practice eq. 3.8 states that also in finite networks the threshold can be very low and will be smaller in scale-free networks than in ER random graphs of equal sizes. This behavior can be understood analysing the value of the basic reproductive number which can be calculated as: $R_0 = \beta \mu^{-1} \langle k^2 \rangle / \langle k \rangle$ in which higher values of $\langle k^2 \rangle$ leads almost surely to $R_0 > 1$.

Recently a more sophisticated class of models have been presented that incorporate a complex interaction scheme and represent well the behavior of large scale societies. These models are defined as *metapopulation models* and have been intensively studied in ecology, populations and migrations dynamics [59, 60, 61]. Metapopulation models rely on assumption that the population under study is structured and localized in discrete patches or subpopulations and connected by some degree of migration. In this class of models each node represents not and individual but a subpopulation and subpopulations are connected through individual mobility. Within each sub-population individuals interact in a homogeneous mixing fashion and they can move from one population to another following some kind of diffusion rule. This framework results very useful in modeling epidemic dynamics in structured populations [57, 58, 128]. It's important to notice that in this case the arrival of the disease in a new subpopulation and the entire evolution of the epidemic process is governed by the mobility rates and coupling between subpopulations.

A very simple metapopulation model of spreading in heterogeneous networks has been presented in [141]. The authors propose a Reaction - Diffusion system with only two types of particles namely A and B that can diffuse at a constant rate $D_{AB} = 1$ and react in the following way:

$$B \xrightarrow{\mu} A \tag{3.9}$$

$$B + A \xrightarrow{\beta} 2B \ . \tag{3.10}$$

Where μ and β represent transition probabilities from one state to another. The system evolution can be represented in terms of fraction of particles A (or B) in nodes with degree k:

$$\rho_{A,k} = \frac{1}{V_k} \sum_{i|k_i=k} a_i \tag{3.11}$$

$$\rho_{B,k} = \frac{1}{V_k} \sum_{i|k_i=k} b_i \ . \tag{3.12}$$

Where a_i and b_i are the number of A and B particles in node i and V_k is the number of nodes with degree k. In this way it's possible to derive an expression for the density of A (B) particles at the stationary state as:

$$\rho_{A,k} = \frac{k}{\langle k \rangle}(\rho_A + \mu\rho_B - \beta\Gamma) \ . \tag{3.13}$$

With $\Gamma = \sum_k P(k)\Gamma_k$ and $\Gamma_k = \rho_{A,k}\rho_{B_k}$. Eq. 3.13 can be studied in a similar fashion like the SIR model (see [141] for the detailed analysis) and an expression for the critical total particle density ρ_c over which B particles are not absorbed by the system is:

$$\rho_c = \frac{\langle k \rangle^2}{\langle k^2 \rangle} \frac{\mu}{\beta} \ . \tag{3.14}$$

This latter result indicates that also in this case for high heterogeneous networks a vanishing critical point is present and the effects of heterogeneity in degree distribution play a fundamental role in shaping spreading dynamics.

As the previous model has profound analogies with a SIS model a more elaborate metapopulation model with a real SIR dynamics has been presented by the same authors [142] and an expression for the critical mobility rate p_c at which the disease invades the whole system is given. In this case individuals can belong to one of the three classes: S, I and R. Inside each node a homogeneous mixing version of the SIR takes place with the basic reproductive number given by $R_0 = \beta/\mu$. A mobility rate p_k proportional to nodes degree and a mean population per node $\overline{N}$ are considered. Solving the model via a mean-field treatment (see [142]) the critical mobility rate p_c reads as:

$$p_c = \frac{1}{\overline{N}} \frac{\langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle} \frac{\mu R_0^2}{2(R_0 - 1)^2} \ . \tag{3.15}$$

Part II

Mobile and Sensor Networks

Introduction

A sensor network consists of a collection of nodes deployed in an environment that cooperate to perform a task. Each node, which is equipped with a radio transceiver, a micro-controller and a set of sensors, shares data to reach the common objective. Sensor networks provide a framework in which, exploiting the collaborative processing capabilities, several problems can be faced and solved in a new way. However, it comes along with several challenges such as limited processing, storage and communication capabilities as well as limited energy supply and bandwidth. Performing a partial computation locally on each node, and exploiting inter-node cooperation, is the ideal way to use sensor networks. Unfortunately, this *modus-operandi* is highly constrained by the reduced hardware capabilities as well as by the limited energy resources that makes communication extremely unreliable as well as expensive in terms of node life-time. As a consequence, the availability of a mechanism to build distributed robust connectivity topologies, where robustness is meant against random node failures and intentional node attacks, is crucial.

Sensor networks can be of interest to different areas of application, ranging from environmental monitoring [64, 83], civil infrastructures [70, 74], medical care [76, 78] to home and office applications [71, 79]. In each field, the deployment of a sensor network has provided interesting advantages. For instance, in the context of environmental monitor the introduction of a sensor network made it possible to keep environments intrinsically threatening for human beings [83] under surveillance, or in the context of medical care it made it possible to remotely monitor the health condition of patients by continuously extracting clinical relevant information [76].

Regardless to the specific application, for a sensor network in order to properly operate, information must be shared across the network allowing for data dissemination and data aggregation. Indeed, a big effort has been done by the research community to develop efficient topology discovery and control algo-

rithms able to achieve that. Strictly speaking, the topology discovery aims to infer the topological structure of the network for management purpose, while the topology control aims to maintain some desired network properties in order to improve the performance of networking services such as routing. In particular, regarding the topology control problem, the majority of works available in the literature address this problem in terms of per-node transmission power in order to increase the life-time of the sensor network [62, 63, 73]. Some contributions focus their attention on the fault-tollerance aspects in terms of network deployment or power assignment [69, 72, 82].

In this second part of the work we present the work carried out in the sensor and mobile networks analysis and control. Specifically our interest will address relevant topics in the field: how to create a self-organized robust topology for a static sensor network and the study of the conditions for the emergence of a collaborative behavior on a class of moving agents.

In chapter 4 a novel topology control algorithm is proposed. The main idea is to design a robust connectivity topology by exploiting percolation theory principles applied to complex networks. In detail, a mechanisms to build an arbitrary topology over a geographical environment is proposed. In addition, a robust distribution against random failures and intentional attacks has been exploited.

Once we have exploited the possibilities for robust static sensor networks we can move one step further considering an enhanced class of networks the so-called *mobile networks*. As the number and performances of mobile devices we use creates new ways of communication and interaction arise. In particular, future internet architecture will a backbone-less architecture in which each node can both produce and handle traffic creating a network of self configuring and dynamically changing devices. Mobile networks are also a good model to study moving unmanned objects behavior as robot or mobile sensors. In this case communication can be crucial to survey or operate in an unknown environment. In both the presented examples: future internet and robots interaction a fundamental role is played by the cooperation between individuals and collective behaviors. Cooperation in these cases requires that part of the individuals capacities (such as battery energy or bandwidth) must be shared among the individuals to obtain a global result (i.e. the whole system connectivity or the ability to solve a complex task). For these reasons the study of the conditions that permits cooperation to be a convenient and winning strategy is crucial for these infrastructures functioning.

In chapter 5 we present a very simplified model for the study of the emergence of cooperation in a group of mobile agents that play a social game as

the Prisoner's dilemma. Especially we use concepts borrowed from random geometric graphs and game theory to numerically investigate the conditions that allows cooperation to be an evolutionary stable strategy in a dynamically chancing interaction network.

Chapter 4

Robust Topologies For Sensor Networks Applications

In this chapter we describe a distributed algorithm for the generation of highly robust network topologies where nodes are randomly distributed on a 2D-plane. Nodes can be easily seen as elements of a sensor network that have been deployed in a geographical space. The main idea under the proposed algorithm is to exploit percolation theory principles to design a very robust topology both to random failures and intentional attacks. To do so we define a specific degree distribution and then let the nodes self-configure according to it.

The rest of the chapter is organized as follows. In Section 4.1 the proposed algorithm is described. In Section 4.2 a numerical analysis to corroborate the analytical results is given. Finally, in Section 4.3 conclusions are drawn.

4.1 The Proposed Algorithm

The simple idea underlying this work is that well-known techniques in the field of complex networks robustness can be suitably applied to a geographical environment. In detail, this work proposes a way of reproducing an arbitrary degree distribution $P(k)$ on a geographical space where nodes are characterized by limited visibility. In particular, a degree distribution with properties of robustness against both faults and attacks is desired. Indeed, the multi-modal distribution proposed by [81] turns out to have these properties. As a result, a robust topology for sensor network can achieved.

The following scenario is considered:

- Nodes are uniformly distributed in a closed 2-Dimensional plane of side L and area $A = L^2$.

- Nodes have a limited radius of interaction r defined as a fraction of L,

Now, given a sensor network consisting of N nodes, the number of neighbors (degree) of a generic node i is $\langle k_i \rangle = \rho \pi r^2$, where $\rho = N/L^2$ is the density of nodes deployment.

According to the given scenario the proposed algorithm works as follow: i) N nodes are distributed uniformly on a square of side L ii) as each node i starts operating, it extracts a integer number k from a selected distribution iii) then i tries to make k connections with the nodes in its visibility radius r, iv) to assure the full connectivity of all the nodes an additional step, i.e., the connectivity maintenance step, is introduced. Note that, a node might not be able to establish the desired number of connections, indeed this is highly influenced by the density of deployment ρ along with the fixed radius r. Nonetheless, a good approximation of the distribution can always be reached for reasonable values of ρ and r. Indeed, this is the case for a realistic sensor network scenario.

As far as the connectivity maintenance step is concerned, the idea is to exploit a consensus algorithm by which nodes share their ID within their visibility neighborhood, i.e., node within its range of visibility. From an algorithmic perspective, each node broadcasts its ID to its neighbors, if a node receives a lower ID it starts sharing the received lower number. Periodically, each node check IDs within visibility neighborhood. If one of these nodes k holds a lower ID, then node i creates a new connection to k and starts sharing k's ID. This step permits to obtain a connected network only adding few links to the original distribution, and if executed periodically to readapt network topology to failures and damages. Note that, even though the connectivity maintenance step is required, from a practical standpoint this can be avoided by performing a proper choice of ρ and r.

Note that, a few parameters regarding the degree distribution are required for the algorithm in order to properly operate. Moreover, as these parameters are fixed, they can be directly hardcoded into each node.

At this point, being a technique for constructing an arbitrary distribution over a geographical space available, the analytical evaluation of the best form for the $P(k)$ is faced.

4.1.1 Optimal degree distribution for network robustness

A robust multimodal degree distribution has been proposed in [81]. In this work, the authors show that a network which maximizes the value of the threshold f_T, defined as $f_T = f_r + f_a$ with f_r the percolation threshold for random node removal and f_a the threshold for targeting node removal, can be obtained by exploiting the following functional form:

$$P(k) = \sum_{i=1}^{m} r_i \delta(k - k_i) = \sum_{i=1}^{m} r_1 a^{-(i-1)} \delta(k - k_i) \tag{4.1}$$

where $k_i = k_1 b^{-(i-1)}$ with k_1 min degree of the network and $\delta(x)$ Dirac's delta function. In addition, the authors show an inter-dependency among all the parameters of Eq. 4.1 leading to a model that depends only on N and m, i.e, the number of nodes and the number of distinct modes in the distribution respectively. In detail, the model is characterized by three different quantities, namely: a that represents the fraction of nodes having different degrees and is larger than 1, b that controls the values of the degrees, and k_1 that is the smallest degree in the network. As far as the other parameters r_1 and r_m are concerned, they can be obtaiend from the normalization condition:

$$\sum_{i=1}^{m} r_i = r_1 \sum_{i=1}^{m} a^{-(i-1)} = 1 \tag{4.2}$$

as follows:

$$r_1 = \frac{1 - a^{-1}}{1 - a^{-m}} \text{ or } r_m = \frac{a - 1}{a^m - 1} \tag{4.3}$$

and

$$\frac{a - 1}{a^m - 1} = \frac{q}{N} \tag{4.4}$$

$$r_m = \frac{q}{N} = N^{\alpha - 1} \text{ with } 0 < \alpha < 0.25 \tag{4.5}$$

with q the number of nodes with the highest degree k_m. Authors of ref. [81] also demonstrate that the mean degree $\langle k \rangle$ is:

$$\langle k \rangle = \sum_{i=1}^{m} k_i r_i = k_1 r_1 \sum_{i=1}^{m} (ab)^{-(i-1)} \tag{4.6}$$

leading to a general form for parameters a and b

$$ab \sim N^{(1/2 - \alpha)/(m-1)} \tag{4.7}$$

from eq.4.7 it is possible calculate

$$a \sim N^{(1-\alpha)/(m-1)} \tag{4.8}$$

and

$$b \sim N^{1/2(m-1)} \tag{4.9}$$

Eqs. 4.8 and 4.9 give the dependency of a and b, once fixed the number of nodes N, from m and α leading to a two parameters model, on which it is possible to calculate the optimal value of the percolation threshold f_T^{opt} defined as the sum $f_T^{opt} = f_a^{opt} + f_r^{opt}$ of the two percolation thresholds for nodes attacks and random failures respectively. it is possible to demonstrate that optimal values of f_T are obtained for $\alpha = 0$ and $m = 2$. Note that, as the f_T^{opt} is a linear combination of two factors, a slightly different behavior, i.e., higher robustness to random node failures or higher robustness to intentional node attacks, can be obtained with a proper choice of the two parameters α and m.

An example of network topology created with the proposed algorithm when exploiting the multi-modal distribution described so far is given in Fig. 4.1. It can be noticed that the obtained topologies are characterized by a high number of triangles which guarantee robustness. At the same time, the degree of the most connected nodes is kept sufficiently low which allows to both mitigate the impact of intentional attacks and limit the effect of random failures.

4.2 Numerical Analysis

The proposed algorithm has been investigated through numerical simulations. Two aspects of interest have been investigated: the robustness to random node failures and the robustness to node attacks. The first aims to evaluate the capability of the sensor network to properly operate even when suddenly some nodes stop working, while the second investigates the resistance of the network when in presence of organized attack aiming to destabilize the normal operating conditions.

The following indexes of quality have been considered: i) the number of components ii) the size of the giant component iii) the percentage of network disconnected. The first index gives an information about the overall connectivity of the network, the second one gives an idea about the remaining operability, while the last one gives an information about the number of nodes still functioning.

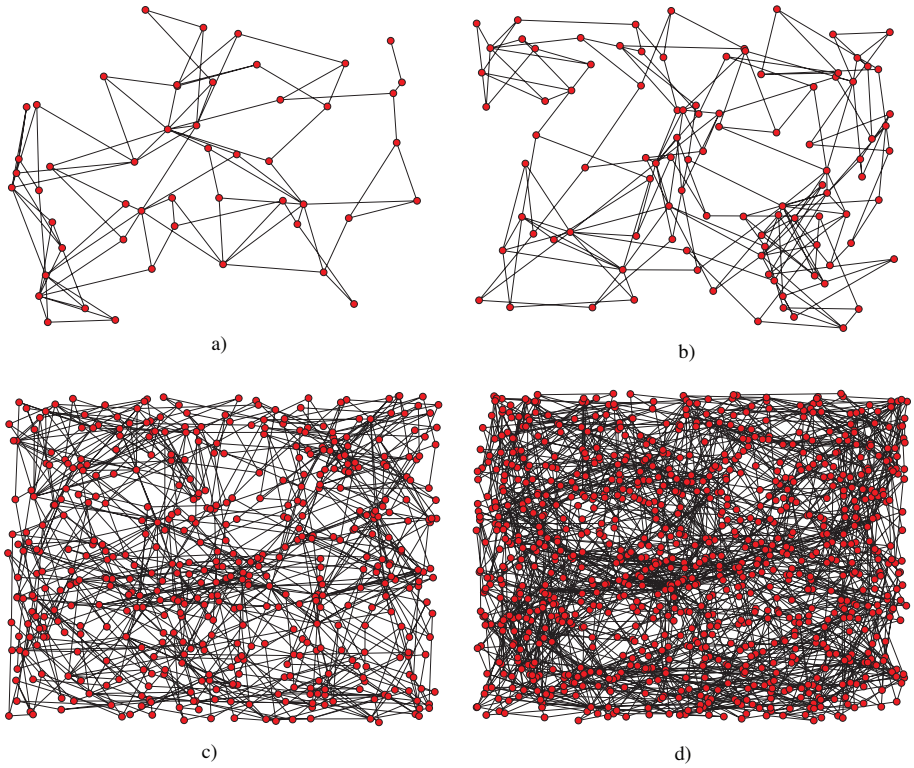

Figure 4.1: Four examples of network produced by the proposed algorithm with different number of nodes N, a) $N = 50$, b) $N = 100$, c) $N = 500$ and d) $N = 1000$, with $m = 3$, $k_1 = 3$ and $\alpha = 0$.

Moreover, a comparison against a null-model has been performed. Such a null-model is built starting from the network produced by the proposed algorithm by keeping the same constraints on the number of nodes, the visibility radius r but introducing a randomized version of the link connections leading to a Poisson degree distribution. As a result, a random network topology is achieved.

Several network configurations have been analyzed. In the following only results regarding a network composed by 2500 nodes deployed in an geographical space with side $L = 10$ and density $\rho = 2$ are shown.

Fig. 4.2-a) shows the degree distribution $P(k)$ obtained for the proposed model with the parameter $m = 3$. It can be noticed the presence of three peaks, respectively for $k = 3, 22, 33$, representing the three modes of the distribution. The two remaining spare peaks can be explained by the limited visibility r of nodes. Indeed, these two peaks would tend to the closest ones on the line if the radius r were sufficiently big. Note that for $m \to \infty$ the distribution $P(k)$ tends to a scale-free distribution [81]. On the other hand, Fig. 4.2-b) describes the degree distribution $P(k)$ obtained for the null-model, which is, as expected, a Poisson distribution.

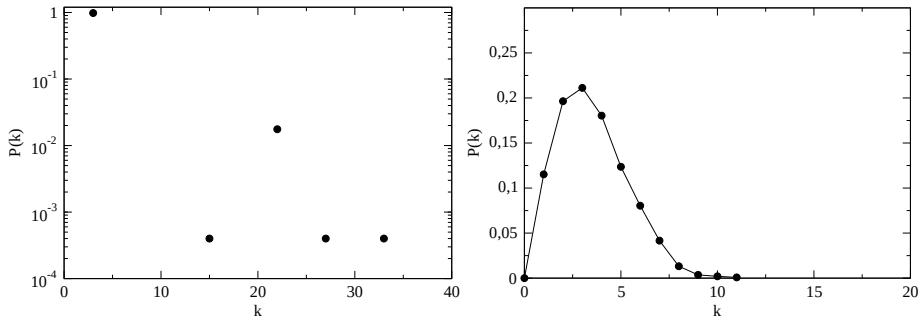

Figure 4.2: Degree distribution of the proposed model a), and of the null-model b). Parameters setting: $N = 2500, m = 3$, $k_1 = 3$, $\alpha = 0$.

Table 4.1 gives a synoptical overview of the conducted analysis. In particular, it can be noticed that when considering two networks with a comparable number of nodes and links the proposed model turns out to be more robust. This can be explained by the higher value of the clustering coefficient $\langle c \rangle$ leading to an higher number of triangles in the network that are known to be the most robust structure against random failures. Moreover, another interesting aspect can be pointed out: both the characteristic path length and diameter values are lower for the proposed model. Indeed, this is a good property for a sensor network as it implies a lower consumption to spread data over the network.

Fig. 4.3 shows the number of connected components (CC) for both the proposed model (circles) and the null-model (squares) when varying the fraction of removed links. In detail, Fig. 4.3-a) represents the behavior of the models against random node failures, while Fig. 4.3-b) depicts the same behavior against intentional attacks. In both cases, the proposed model outperforms

Table 4.1: Main topological features of the proposed model against the null-model.

Model	N	E	$\langle k \rangle$	k_{max}	$\langle d \rangle$	d_{max}	$\langle c \rangle$
Proposed Model	2500	4201	3.36	33	5.99	13	0.004766
Null Model	2500	4277	3.42	11	7.24	15	0.001279

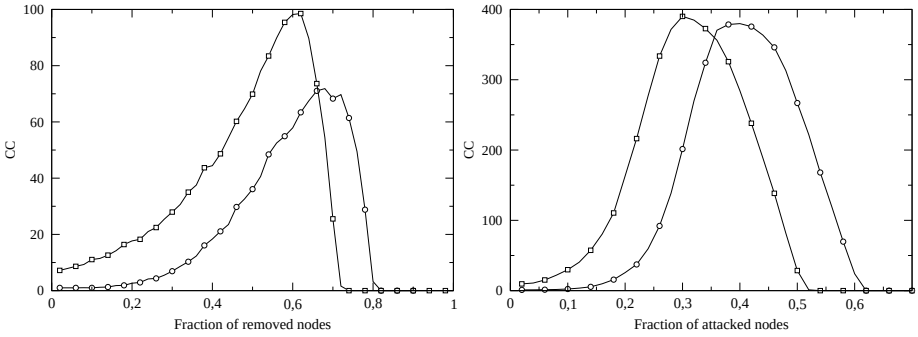

Figure 4.3: Number of connected components (CC) vs. fraction of removed nodes for the proposed model (circles) and the null-model (squares) in case of random node failures a) and intentional node attacks b). Parameters setting: $N = 2500, m = 3$, $k_1 = 3$, $\alpha = 0$.

the null-model, i.e., the network starts to break down after a higher fraction of node (approx. 20%). Note that, isolated nodes are not counted as components.

Fig. 4.4 shows the size of the giant component for both the proposed model (circles) and the null-model (squares) when varying the fraction of removed links. Also in this case, the proposed model outperforms the null-model. In particular, the size of the biggest component decreases almost linearly with the fraction of removed nodes in the case of random nodes removal.

Finally, Fig. 4.5 shows the fraction of disconnected nodes for both the model (circles) and the null-model (squares) when varying the fraction of removed links. As before, the performance of the proposed model is significantly better than the null-model.

An additional analysis of the behavior of proposed technique has been successively carried out. In particular the following aspects have been investigated: the rate of growth of the number of links with the respect to the number of

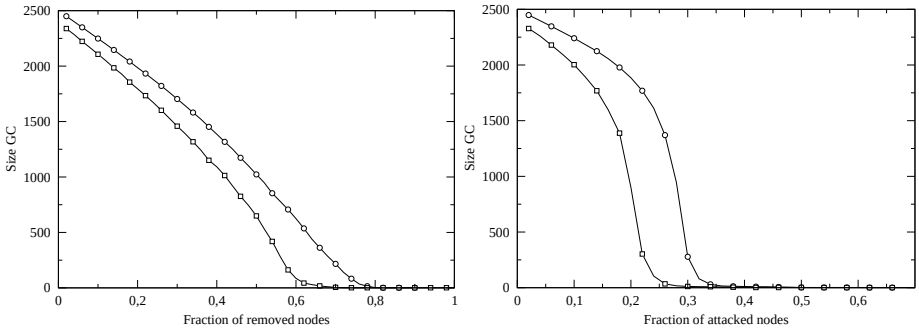

Figure 4.4: Size of the giant component (Size GC) vs. fraction of removed nodes for the proposed model (circles) and the null-model (squares) in case of random node failures a) and intentional node attacks b). Parameters setting: $N = 2500, m = 3$, $k_1 = 3$, $\alpha = 0$.

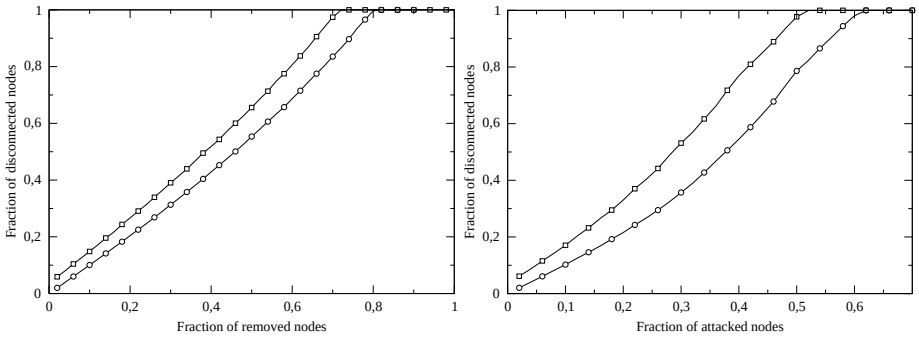

Figure 4.5: Fraction of disconnected nodes vs. fraction of removed nodes for the proposed model (circles) and the null-model (squares) in case of random node failures a) and intentional node attacks b). Parameters setting: $N = 2500, m = 3$, $k_1 = 3$, $\alpha = 0$.

nodes, the fraction of isolated nodes resulting from the removal of a fraction of nodes and the variation of the tree indexes previously introduced, i.e., the number of components, the size of the giant component and the percentage of network disconnected, when varying the value of the parameter m.

Fig. 4.6-a) shows the rate of growth of the number of links with respect

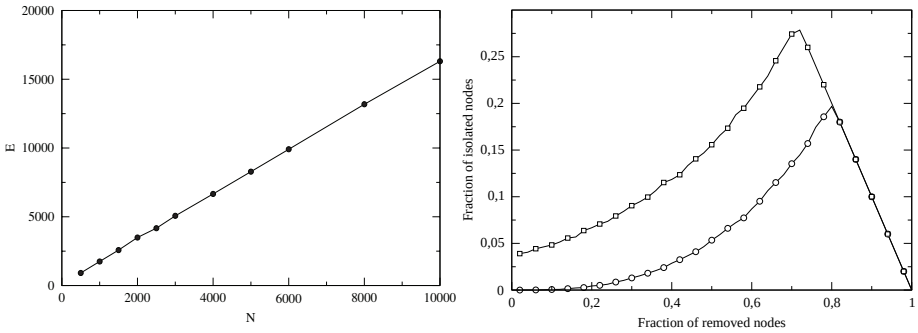

Figure 4.6: a) Growth rate of the number of links vs. the number of nodes. b) Fraction of isolated nodes resulting from the removal of a fraction of nodes in case of random node failures.

to the number of nodes. it can be easily noticed that the number of links increases linearly with the number of nodes. This is indeed a good property of the algorithm as the higher vis the number of links the higher is the power consumption of the network leading to a good scalability. Fig. 4.6-b) illustrates the fraction of isolated nodes resulting as a consequence of the removal of a fraction of nodes. This is another interesting property of the algorithm. In fact, it points out that only a negligible percentage of nodes are affected by the removal of other nodes. In other words, by removing a node the connectivity of its neighbors is not significantly influenced.

Figs. 4.7 shows how the tree indexes change when varying the value of the parameter m. Note that, this result is referred to the proposed model against random node failures. According to the theoretical results, the higher is the value of the parameter m the better is the performance against random failures as the scale-free characterization of the degree distribution becomes more and more notable. Indeed, this is in agreement with the results obtained in [81] as the percolation threshold is not influenced by the variation of the parameter m, but at the same time other characteristics, such as the number of connected components, are positively influenced in the case of random node failures.

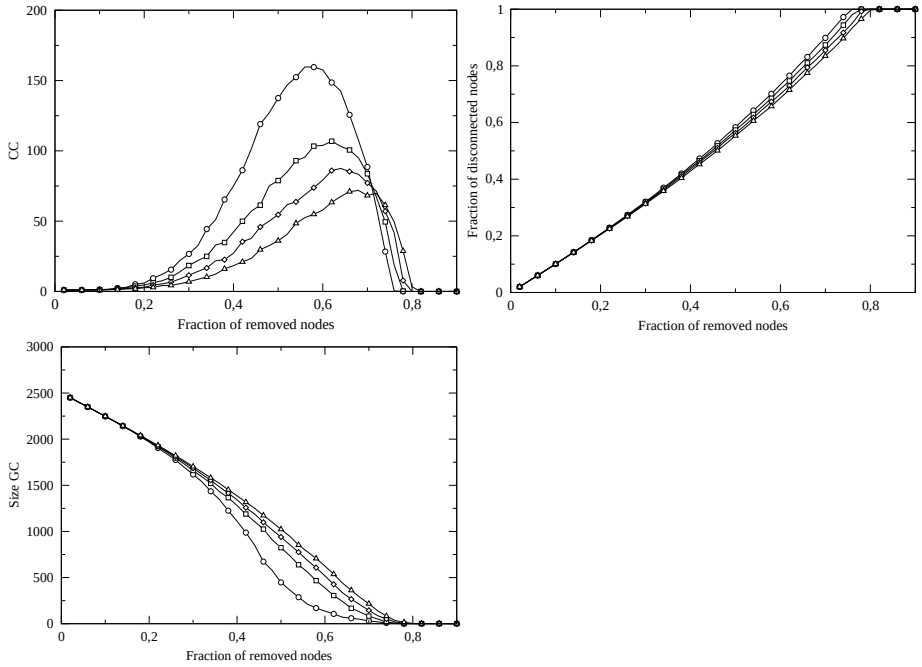

Figure 4.7: a) Number of connected components, b) Size of the giant component and c) Fraction of isolated nodes vs fraction of removed nodes in the proposed model with different values of m. $m = 3$ (circles), $m = 5$ (squares), $m = 7$ (diamonds), $m = 10$ (triangles). Parameters setting: $N = 2500$, $k_1 = 3$, $\alpha = 0$.

4.3 Conclusions

In this chapter, a novel topology control algorithm has been proposed. Indeed, the availability of a connectivity topology algorithm able to properly operate even when in presence of random failures of nodes drastically increases the robustness as well as the operability of a sensor network.

In detail, an algorithm to build an arbitrary topology over a geographical environment is proposed. In addition, a robust degree distribution against random failures and intentional attacks has been exploited [81]. The properties of the resulting model have been analytically characterized by exploiting the percolation theory and the related results have been corroborated by numerical

simulations. In particular, three different indexes of quality have been investigated, namely the number of connected components, the size of the giant component and the fraction of disconnected nodes in the network. Moreover, a comparison against a randomized version of the network (null-model) has been performed. According to these results, the proposed topology control technique has turned out to be very effective as it always outperforms the null-model in terms of connectivity maintenance against both random node failures and intentional node attacks.

To conclude the proposed algorithm is distributed and easy to implement on-board each node. It requires a limited number of messages in order to build the topology and the number of links scales linearly with the size of the network. Moreover, even though the algorithm has been implemented only in a 2-dimensional plane, there is no additional cost to extend it to a n-dimensional space, as the topology construction relies only on the Euclidian distance.

Several challenges still remain for future work. An extension where a node independently sets its radius of visibility r might be investigated in order to reduce the energy consumption. In addition, a dynamical network rewiring process able to reconnect the network anytime two or more components arise might be studied. Finally, an enhanced scenario where mobility is taken into account for some nodes might be of interest.

Chapter 5

Cooperation in a Mobile Agents Scenario

An open question in biology and social sciences is to understand how cooperation emerges in a population of selfish individuals. A theoretical framework that has shed some light into this long-standing problem is evolutionary game theory [172, 173]. Through the development and the study of different social dilemmas, scientists have been able to elucidate some of the mechanisms that enable cooperative behavior in populations.

One of the most studied games is the Prisoner's Dilemma (PD) in which players can cooperate (C) or defect (D) and receive a different payoff according to their and their opponent strategy. While a population of individuals playing a PD game does not support cooperation if they are well-mixed, the existence of a spatial structure gives as a result that cooperation survives under certain conditions as cooperative clusters can emerge in the system [172, 173].

Recent works have shown that the cooperative behavior is actually enhanced when the individuals play on a ER random graph. Further improvement in the global level of cooperation is obtained if the network of contacts is scale-free [174, 175, 176, 177]. The reason is that cooperators are fixed in the highly connected nodes, turning also into cooperators their neighborhood and guaranteeing in this way their long-time success. Additionally, several works have explored different rewiring mechanisms that allow an improvement in the average level of cooperation in the system [178, 179, 180]. Interestingly, social dilemmas can also be used to generate highly cooperative networks by implementing a growth mechanism in which the newcomers are attracted to already

existing nodes with a probability that depends on the nodes' benefits [181].

In spite of the relative large body of work that has been accumulated in the last few years, there are situations of practical relevance that remain less explored. This is the case of models where individuals can move and change their neighborhood continuously by encountering different game's partners as time goes on. Highly changing environments can be found in a number of social situations and the study of how cooperative levels are affected by the inherent mobility of the system's constituents can shed light on the general question of how cooperation emerges. Furthermore, the insight gained can be used to design cooperation-based protocols for communication between wireless devices such as robots [182]. Recently, a few works have dealt with this kind of situation [183, 184, 185]. However, the models were limited to the case in which individuals are allowed to move on the sites of a 2D regular lattice.

In this chapter, we consider the less-constrained -perhaps more realistic- case in which a set of Prisoner's Dilemma players unconditionally move on a two dimensional plane. We explore under which conditions cooperation is sustained. In particular, we inspect the robustness of the average level of cooperation in the population under variation of the game parameters and of the mobility rules. Our results show that cooperation is actually promoted provided that players do not move too fast and that cooperation is not too expensive. Additionally, at variance with other cases, the dynamics of the system exhibits only two stable attractors -those in which the whole population plays with one of the two possible strategies.

the rest of the chapter is organized as follows. In section 5.1 the dynamical model is presented. Then the results of the numerical simulations and a brief discussion are given in section 5.2 and finally in section 5.3 the conclusions are drawn.

5.1 The model

In our model, we consider N agents (individuals) moving in a square plane of size L with periodic boundary conditions, and playing a game on the instantaneous network of contacts. The three main ingredients of the model are: the rules of the motion, the definition of the graph of interactions, and the rules of the evolutionary game.

5.1.1 Motion

Each agent moves at time t with a velocity $\mathbf{v}_i(t)$ $(i = 1, 2, \ldots, N)$. We assume that the modulus of the velocities of the individuals is constant in time, and equal for all the agents, while the individuals are allowed to change their direction of motion, $\theta_i(t)$. Hence we can write the velocities as: $\mathbf{v}_i(t) = (v\cos\theta_i(t), v\sin\theta_i(t))$. The individuals are initially assigned a random position in the square and a random direction of motion. At each time step they update their positions and velocity according to the following dynamical rules:

$$\boldsymbol{x_i(t+1) = x_i(t) + v_i(t)} \tag{5.1}$$

$$\theta_i(t+1) = \eta_i \tag{5.2}$$

where $\mathbf{x}_i(t)$ is the position of the i-th agent in the plane at time t and η_i are N independent random variables chosen at each time with uniform probability in the interval $[-\pi; \pi]$.

5.1.2 Network of interactions

At each time step we consider that the neighborhood of a given agent i is made up by all the individuals j which are within an Euclidean distance d_{ij} less than some threshold r. In what follows, without loss of generality, we set $r = 1$. Therefore, the instant network of contacts is defined as the graph formed by nodes centered at all the N circles of radius 1 together with the links between those agents in the neighborhood of each individual. Note that as agents move every time step, the network of contacts, and hence the adjacency matrix of the graph is continuously changing, not only because the number of contacts an individual has may change, but also due to the fact that the neighbors are not always the same. The topological features of the graph defined above depend on several parameters. For instance, the mean degree of the graph can be written as $\langle k\rangle = \rho\pi r^2 = \rho\pi$ where $\rho = N/L^2$ is the density of agents. For small values of ρ, the graph is composed by several components and there may also exist isolated individuals. On the contrary, when $\rho > \rho_c$ a unique giant component appears [41] (for our system with periodic boundary conditions $\rho_c \sim 1.43$). An example of an interaction network with $N = 10^3$ agents is depicted in fig. 5.1. Long range links are due to the periodic boundary conditions.

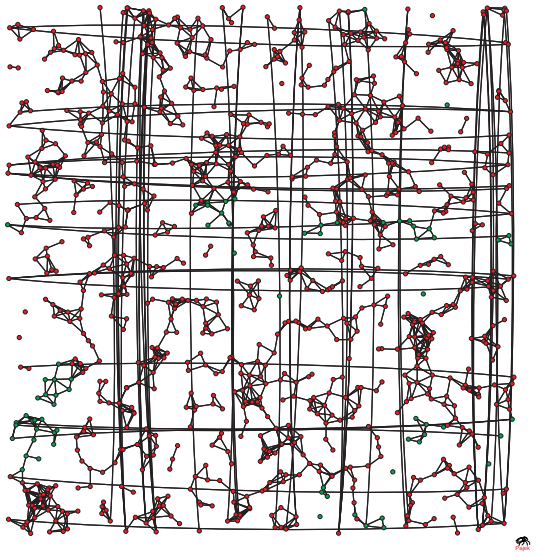

Figure 5.1: A typical interaction network with $N = 10^3$ agents. Cooperators agents are depicted in green and defectors agents are in red. Long range links are due to the periodic boundary conditions.

5.1.3 Evolutionary dynamics

As the rules governing the evolutionary dynamics, we assume that individuals interact by playing the Prisoner's Dilemma (PD) game. Initially, players adopt one of the two available strategies, namely to cooperate or to defect, with the same probability 1/2. At every round of the game all the agents play once with all their corresponding instant neighbors. The results of a game translate into the following payoffs: both agents receive R under mutual cooperation and P under mutual defection, while a cooperator receives S when confronted to a defector, which in turn receives T. These four payoffs are ordered as $T > R > P \geq S$ in the PD game so that defection is the best choice, regardless of the opponent strategy. As usual in recent studies, we choose the PD payoffs as $R = 1$, $P = S = 0$, and $T = b > 1$. Once the agents have played with all their neighbors, they accumulate the payoffs obtained in each game, and

depending on their total payoffs and on the payoffs of the first neighbors, they decide whether or not to keep playing with the same strategy for the next round robin. In this process, an agent i picks up at random one of its neighbors, say j, and compare their respective payoffs P_i and P_j. If $P_i > P_j$, nothing happens and i keeps playing with the same strategy. On the contrary, if $P_j > P_i$, agent i adopts the strategy of j with a probability proportional to the payoff difference:

$$\Pi_{ij} = \frac{P_j - P_i}{\max\{k_j, k_i\}b} \, , \tag{5.3}$$

where k_i and k_j are the number of instant neighbors of i and j respectively (*i.e.* the number of agents inside the circles of radius r centered at i and j respectively). This process of strategy updating is done synchronously for all the agents of the system and is a finite population analogue of replicator dynamics. When finished, the payoffs are reset to zero, so that repeated games are not considered.

The movement and game dynamics might in general be correlated, and the influence of the agents movement on the performance of the PD dynamics depends on the ratio between their corresponding time scales. Here, we consider the situation in which both movement and evolutionary dynamics have the same time scale. Therefore, at each time step, the following sequence is performed: *(i)* the agents perform a new movement in the two-dimensional space, *(ii)* establish the new network of contacts (determined by the radius r of interaction) and *(iii)* they play a round of the PD game, accumulating the payoffs and finally updating their corresponding strategies accordingly. After this latter step, the players move again. The process is repeated until a stationary state is reached. Here, a stationary state is one in which no further changes of strategies are possible.

5.2 Results and Discussions

We have performed extensive numerical simulations of the model for various values of the agent density ρ and velocity v, and different values of the game parameter b. Let us first note that for the limiting case in which $v = 0$, the results point out that the average level of cooperation is different from zero, as one might expect from the fact that the underlying network of contacts has a Poisson degree distribution. Indeed, the graph corresponds to a random geometric graph [41], a network having the same $P(k)$ as an ER random graph, but with a higher clustering coefficient. This latter feature leads to a further

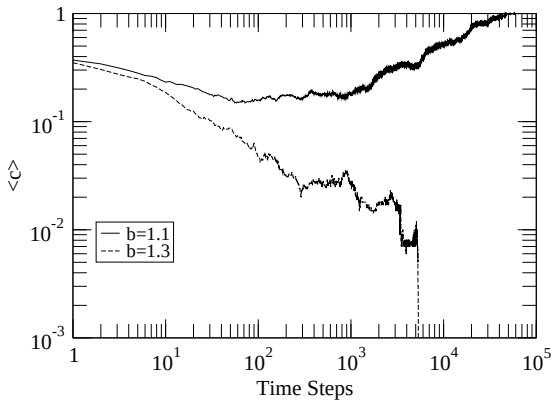

Figure 5.2: Average level of cooperation, $\langle c \rangle$, as a function of time (Monte Carlo steps) for $v = 0.01$ and two different values of b, $b = 1.1$ $b = 1.3$, as indicated. Other model parameters have been fixed to $\rho = 1.30$ and $N = 10^3$ agents.

increment of the average level of cooperation, as it has been shown that a network with a high clustering coefficient promotes cooperation [186, 187].

Let us now focus on the case $v \neq 0$. The first difference that arises with respect to the case in which agents do not move is that the dynamics of the system only have two attractors. Namely, the asymptotic state (i.e., when the probability that any player changes its strategy is zero) is either a fully cooperative network (all-C) or a network in which all the individuals end up playing as defectors (all-D). This behavior is illustrated in Fig. 5.2, where we have reported the average level of cooperation $\langle c \rangle$ in a population of $N = 10^3$ individuals as a function of time, for $v = 0.01$ and for two different values of b. Starting from a configuration in which individuals are cooperators or defectors with the same probability, the average level of cooperation slowly evolves to one of the two asymptotic states: all-C or all-D. It is also worth stressing that the system reaches those states more slowly than in static settings (i.e., when $v = 0$). Specifically, it appears that the system spends a considerable time in metastable states (flat regions in the figure) that are followed by a sudden decrease (or increase) of the average level of cooperation.

The evolution of the system depends on the density of players. In Fig. 5.3,

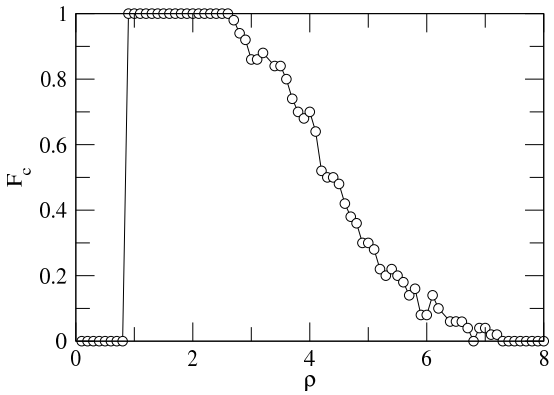

Figure 5.3: Fraction of realizations in which the system ends up in an all-C configuration, F_c, as a function of the density of players ρ for a fixed value of $b = 1.1$ and $v = 0.01$. The system is made up of $N = 10^3$ agents. The results are averages taken over 100 different realizations.

we have represented the dependence of the fraction of realizations, F_c in which the population ends up in an all-C configuration as a function of the density ρ for $b = 1.1$ and $v = 0.01$. There are two limits for which $F_c = 0$. At low values of the density, the agents are too spread in the 2D plane. As a result, cooperators unsuccessfully strive to survive and get extinguished given the low chance they have to form clusters -the only mechanism that can enforce their success. On the contrary, for large values of ρ the population is quite dense and, locally, the agents' neighborhoods resemble a well-mixed population in which more or less everybody interacts with everybody and therefore defection is the only possible asymptotic state. Values of ρ between these two limiting cases confer to cooperators a chance to survive. Interestingly, there is a region of the density of players, $0.9 \lesssim \rho \lesssim 3$ which is optimal for cooperative behavior. Beyond this region F_c decays exponentially with ρ reaching zero at $\rho \approx 7$.

Up to now, we have analyzed the behavior of the system for small values of the velocity of the agents and of the temptation to defect. Figure 5.4 summarizes the results obtained for a wider range of model parameters (v and b) in a population of $N = 10^3$ agents and $\rho = 1.3$. The results are averages taken over 100 realizations of the model. The phase diagram shows a relative wide region of the model parameters in which cooperative behavior survives. For a

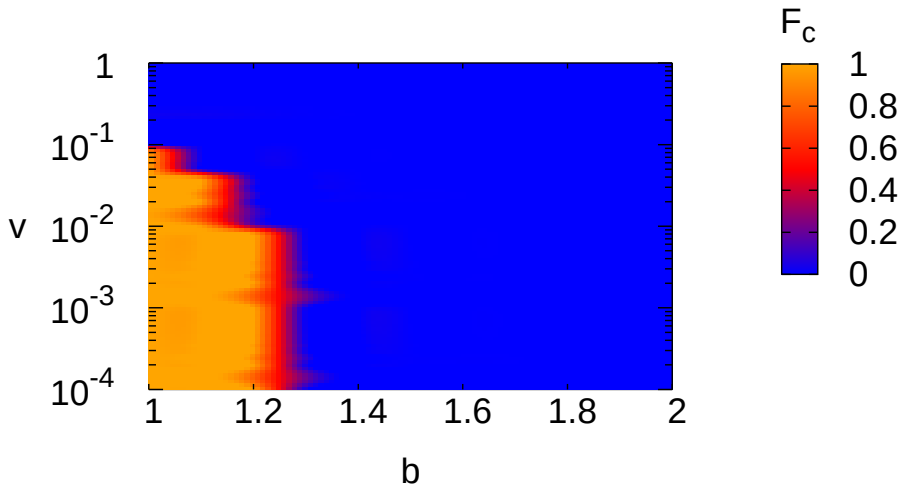

Figure 5.4: The color code shows the fraction of realizations in which the whole system is made up of cooperators, F_c, as a function of the velocity at which the agents move (v) and the temptation to defect (b). The Y-axis is in log scale for clarity. The rest of parameters are $N = 10^3$ agents and $\rho = 1.30$. Each point is an average over 100 different realizations

fixed value of v, this region is bounded by a maximum value of the temptation to defect close to $b = 1.3$, which decreases as the velocity at which players move increases. Furthermore, when b is kept fixed, increasing the value of v is not always beneficial for the survival of cooperation. In fact, when the individuals move too fast, they change their environment quite often and quickly, then increasing the likelihood to meet each time step a completely different set of players. In other words, when the velocity is increased beyond a certain value, the well-mixed hypothesis applies to the whole population of players, thus leading to the extinction of cooperation in the long time limit.

Figure 5.5 sheds more light on the dependence of the fraction of cooperators with respect to the velocity of the agents. There we have represented the layer corresponding to $b = 1.1$ in Fig. 5.4. As can be seen from the figure, for low

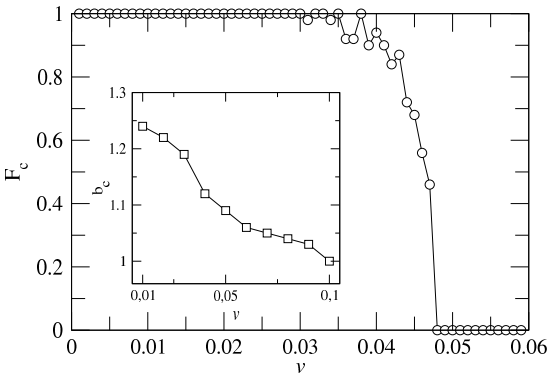

Figure 5.5: Fraction of realizations ending up in an all-C configuration as a function of the velocity v of the agents for $b = 1.1$. The inset shows the smallest value of the temptation to defect, b_c, for which the probability of achieving a fully cooperator asymptotic state is zero, as a function of v. In both cases, $N = 10^3$ agents, $\rho = 1.30$, and results correspond to averages over 100 realizations.

values of v all the realizations lead the system to a configuration in which all strategists are cooperators. As the PD players move faster, the probability of achieving such a configuration decreases and gets zero for values of v close to 0.05. From that point on, the all-C asymptotic state is never realized. This latter point also depends on the specific value of b. The inset of Fig. 5.5, represents the smallest values of the temptation to defect, b_c, for which in all the realizations performed the system ended up in the all defectors state as a function of v. The results show that beyond $v \approx 0.1$, cooperation never survives in a population of moving agents irrespective of b.

5.3 Conclusions

Concluding in this chapter, we have studied the effects of mobility on a population of Prisoner's Dilemma players that are able to move in a two-dimensional plane. Numerical simulations of the model show that a fully cooperative system is sustained when both the temptation to defect and the velocity of the agents are not too high. Although cooperation is extinguished for a wide region of the

parameter space, our results show that mobility have a positive effect on the emergence of cooperation. As a matter of fact, as soon as $v \neq 0$, the mobility of the agents provokes the spread of the winning strategy to the whole population, leading the system to a global attractor in which all players share the surviving strategy. In other words, the movement of individuals prevents the coexistence of different strategies in the long time limit. Namely, for small (and fixed) values of b cooperation prevails at low velocities, while defection succeeds for larger v. Our results are relevant for the design of new cooperation-based protocols aimed at motion coordination among wireless devices and for other communication processes based on game theoretical models [182].

Part III

Traffic Processes in Technological Networks

Introduction

Communication networks [188] are nowadays subject of intense research as modern society increasingly depends on them. In this third part we investigate some interesting models for the study of traffic properties on such networks.

The first studies on the field have dealt with the architecture of these systems, showing that the systems' topological features [188, 189, 190] are at the root of the critical behavior of several dynamical processes taking place on top of them [1, 188]. On the other hand, models for traffic and information flow on complex networks have been recently investigated as a way to improve our understanding on key issues such as the scalability, robustness, performance and dynamics of technological networks [1, 188]. In particular, much effort has been invested in finding what are the conditions for an efficient performance of communication networks, the latter being measured as the ability of the system to avoid congestion and reduce transit times [191, 192, 193, 194].

Actual studies are mainly distinguished in two areas. One focusing on the traffic properties in the free flow regime as main communication infrastructures adopt control mechanisms to prevent the onset of the congestion (i.e. congestion avoidance in TCP protocol [42]). These studies address problems as the structural evolution of these networks [85, 86] or the navigability [87, 88, 89] and the dynamical features of traffic [90, 91]. Another area of interest is represented by the study of mechanisms and techniques aimed to delay the arise of congestion. These mechanisms are intended to operate before traffic control protocols to allow the network to handle a larger quantity of traffic before congestion control become active. Many studies try to optimize traffic conditions by designing efficient routing strategies [95, 96, 97, 98, 99, 101, 100] that on one hand, provide with short delivery times and, on the other hand, avoid the onset of the congested state causing the failure of information flow. It has been shown that finding the best suited strategy depends strongly on two main features: the topological patterns of the particular network and the load of information

on top of it. Regarding the first of these two issues, a number of routing mechanisms have been studied on different structures [101, 102, 103, 104] allowing to design resilient network backbones [105, 106, 107]. On the second issue many congestion-aware routing protocols have been proposed, based on biased random walks [113], shortest-path [114, 115] and efficient-path [116] routings. In addition to the design of efficient routing protocols, several strategies to avoid congestion have been implemented. Remarkable examples of these strategies are the implementation of incoming flow rejection [117, 118] and the packet-dropping mechanisms [119] for avoiding the congestion of single nodes.

In this part we will cover the two aspects of traffic dynamics: the study of the traffic properties in the free flow regime and mechanisms to delay the onset of the congested state.

Predicting fluctuations in traffic intensity is a crucial issue in communication network design and maintenance as hardware costs and resource planning are strictly influenced by variations on the mean traffic level. The aim of chapter 6 is to investigate the relationship between mean flow on a node and its variance. Specifically, we present a traffic model based on random diffusion that can give some insights on the value of standard deviation in single nodes and then study the scaling relationship between mean traffic and its variations.

The problem of delaying the appearance of congestion is tackled in chapter 7 in which an analytic model for traffic simulation is presented and then a series of self-adaptive strategies are proposed. The main idea behind these adaptive behavior is that nodes located in different positions on the network, based on their degree and congestion level of their neighbors, can experience different traffic values and then the optimal behavior can be different. We propose a minimal traffic model in which nodes can chose if accept or not an incoming packet based on their congestion state. As a further improvement we consider a situation in which nodes can collaborate with their neighbors and define a local optimal strategy.

Chapter 6

Flow Fluctuations on Complex Networks

The relationship between the fluctuations σ and the average flux $\langle f \rangle$ in traffic dynamics on complex networks is a controversial issue that has received a lot of attention very recently [195, 197]. The authors of Refs. [195, 196] claimed the existence of the relation $\sigma \sim \langle f \rangle^{\alpha}$, with real communication networks belonging to one of two universality classes, the first one characterized by an exponent value $\alpha = 1/2$, the second one by $\alpha = 1$. The authors of [197] questioned the existence of the two universality classes. They numerically showed that there is a wide spectrum of possible values for α, depending on parameters such as the persistence of packets in the network, the duration of the time window during which statistics are recorded, and the rate of service at the nodes' queues [197].

In this chapter, we propose a model for traffic in complex networks, the Random Diffusion (RD) model, that is amenable to analytical solution. The model predicts the existence of a simple law that relates the fluctuations at a node i, σ_i to the average traffic flow f_i, depending on the delicate balance of three quantities: (i) the variation in the number of packets in the network, (ii) the degree of the node i, and (iii) the length of the time window in which measures of traffic flow are performed. Notwithstanding its simplicity, the RD model is able to capture the essential ingredients determining the scaling of fluctuations empirically observed for traffic flow in real complex networks. More important, we also show that the hypothesis of a *power-law scaling* of flow fluctuations has to be abandoned under certain conditions. Results of numerical simulations of a traffic-aware model and analysis of real data of

Internet flow confirm our theoretical findings.

6.1 Random diffusion model

In the random diffusion (RD) model we represent packets of information as w random walkers traveling in a network made up of N nodes and K links. Under the assumption that the packets are not interacting, it follows that the average number of walkers λ_i at a node i is given, in the stationary regime, by [198, 199]

$$\lambda_i(w) = \frac{k_i}{2K} w \ . \tag{6.1}$$

Let us assume that the total observation time T is divided into time-windows of equal length. Each window is made of M time units. A window represents the minimal resolution for measurements of the flux in a node and its fluctuations, being the first the result of accumulating the number of packets traveling through the node during the M time units. The average number of packets $\langle f_i \rangle$ processed by node i in a time window is measured, together with its standard deviation σ_i. These are the two quantities monitored in Refs.[195, 197] for real systems and in the numerical simulations of network traffic models. The main interest is to investigate the dependence of σ_i with $\langle f_i \rangle$. In particular, we want to verify whether a power-law relation $\sigma_i \sim \langle f_i \rangle^\alpha$ holds, and what factors determine the exponent α. In the RD model we can consider two possible situations: either the number of packets in the network is constant over the whole period of time T, namely $w = W$, or it can vary from one time window to the other. In the latter situation, we assume that the probability $F(w)$ of having w walkers on the network in a window of length M is equally distributed in the range $[W - \delta, W + \delta]$, i.e.,

$$F(w) = \frac{1}{2\delta + 1} \ , \tag{6.2}$$

with $1 \leq \delta \leq W$. To find an expression for the average number of packets $\langle f_i \rangle$ flowing through a given node i, we first calculate the probability $P_i(n)$ that, after M time steps, n packets have visited node i.

In the case $w = W$, due to the fact that the packets are not interacting, the arrival of walkers at a node is a Poisson process. Therefore, after a period of M time units, the mean number of packets (the average flux) at a node i is $\langle f_i \rangle = \lambda_i(w) M$, and the probability of having n packets reads

$$P_i(n) = \mathrm{e}^{-\lambda_i(w) M} \frac{(\lambda_i(w) M)^n}{n!} \ , \tag{6.3}$$

with $\sigma = \sqrt{\lambda_i(w)M} = \sqrt{\langle f_i \rangle}$. Thus, the scaling exponent is $\alpha = 1/2$.

In the more general case in which the number w is distributed as in Eq. (6.2), the probablity $P_i(n)$ is

$$P_i(n) = \sum_{j=0}^{j=2\delta} \frac{\mathrm{e}^{-\frac{k_i}{2K}(W-\delta+j)M}}{2\delta+1} \frac{[\frac{k_i}{2K}(W-\delta+j)M]^n}{n!} \,. \tag{6.4}$$

Calculating first and second moments of $P_i(n)$ one obtains

$$\langle f_i \rangle = \sum_{n=0}^{\infty} nP_i(n) = \frac{k_i W M}{2K} \,, \tag{6.5}$$

$$\langle f_i^2 \rangle = \sum_{n=0}^{\infty} n^2 P_i(n) = \langle f_i \rangle^2 (1 + \frac{\delta^2}{W^2}) + \langle f_i \rangle \,. \tag{6.6}$$

Finally, the standard deviation can be expressed as a function of $\langle f_i \rangle$ as

$$\sigma_i^2 = \langle f_i \rangle \left(1 + \langle f_i \rangle \frac{\delta^2}{W^2} \right) \,. \tag{6.7}$$

6.2 Discussion

The above derivation provides an understanding of the origins of Eq. (6.7), proposed in [195], and shows that the relation between σ_i and $\langle f_i \rangle$ depends on the concurrent effects of three factors, namely: *(i)* the noise δ associated to the fluctuations in the number of packets in the network from time window to time window; *(ii)* the length M of the time window; and *(iii)* the degree of the node k_i (since $\langle f_i \rangle$ depends on k_i). Consequently, real traffic rarely falls in either of the two limiting cases of Eq. (6.7), i.e., $\sigma \sim \langle f \rangle^\alpha$ with $\alpha = 1/2$ or 1.

Expression (6.7) contains all the behaviors previously observed in Refs. [195, 197], and also predicts new dependencies that can be tested to be valid in more refined traffic models as well as in real data. In fact, if the three quantities δ, M and k_i are such that

$$\frac{k_i M \delta^2}{2KW} \ll 1 \,, \tag{6.8}$$

expression (6.7) reduces to a power-law scaling $\sigma \sim \langle f \rangle^\alpha$ with exponent $\alpha = 1/2$. On the contrary, whenever the ratio $\frac{k_i M \delta^2}{2KW}$ is not negligible anymore, the exponent α differs from $1/2$ and approaches 1. In other words, it may well

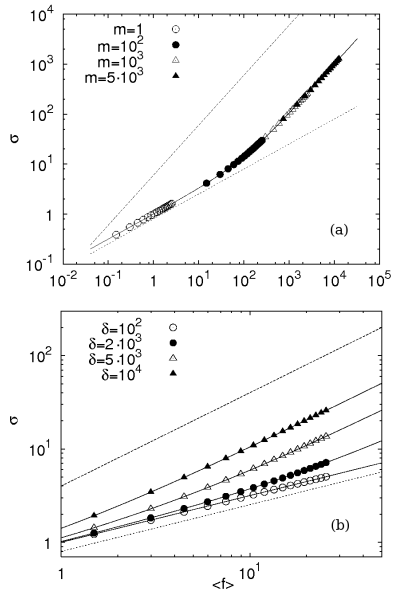

Figure 6.1: Flow fluctuation σ as a function of $\langle f \rangle$ for the RD model with various parameter values. In panel (a), $\delta = 10^3$ and $W = 10^4$. In panel (b), W has the same value while M has been fixed to 10. In both figures, points correspond to the solution of Eq. (6.7) for different values of k_i $(1 \ldots 18)$. The total number of links is $K = 33500$. Dashed lines are guides to the eyes and correspond to $\sigma \sim \langle f \rangle^\alpha$, with $\alpha = 1/2$ (lower curves) and $\alpha = 1$ (upper curves). See the text for further details.

be the case in which, even for small values of the noise parameter δ, a large value of M cancels out the effect of the ratio $\frac{\delta}{W}$ being too small in Eq. (6.7). This behavior was already explored in [197] by means of numerical simulations. However, the fact that the ratio in formula (6.8) depends quadratically on δ and only linearly on k_i and M, has gone unnoticed. The RD model puts such dependence on solid theoretical grounds, and also reveals the role played by the other two parameters M and k_i on the observed scaling.

In Fig. 6.1 we plot the dependence of σ with $\langle f \rangle$ in the RD model for several values of the parameters M and δ. Panel (a) corresponds to the case in which

the ratio $\frac{\delta}{W} = 10^{-1}$ is fixed and the length of the time windows used to measure the flow of packets through different nodes is varied. For each value of M, we have superimposed the results obtained for nodes with different connectivity values, ranging from $k_i = 1$ to $k_i = 18$. If one follows the arguments given in [195], a value of $\alpha = 1/2$ should be expected for this choice of δ/W. Instead, as shown in the figure, $\sigma \sim \langle f \rangle^{\frac{1}{2}}$ only for small values of M, while the scaling exponent approaches 1 as M is increased. This means that, whenever the temporal resolution in the measurements is not small enough and packets are counted and accumulated over long periods, α tends to 1.

A novel striking feature revealed by law (6.7), and not revealed in previous studies, is the dependence with the degree of the nodes. An example of the effects of node degrees is shown in Fig. 6.1*a*. It turns out that, for some values of M (e.g. $M = 10^2$ in the figure), the fluctuations at lowly connected nodes are characterized by an exponent $\alpha = 1/2$, whereas for highly connected nodes the exponent turns out to be $\alpha = 1$. Hence, there is not a single exponent characterizing the fluctuations at *every* node of the network, regardless of its connectivity. This is again a clear indication that a power-law behaviour, $\sigma \sim \langle f \rangle^{\alpha}$, even with non-universal exponents ranging in $[1/2, 1]$, is not the most general situation when characterizing the flow fluctuations for a whole network [195, 197]. Admittedly, α is not constant for every possible choice of the parameters δ, W and M along the whole set of k_i values. This effect is particularly relevant for highly heterogeneous networks like the Internet, where degree classes span several decades. In these kinds of networks, one should therefore expect different scaling laws depending on whether the packets are flowing through lowly or highly connected nodes.

The influence of the noise level on α for a fixed time window length ($M = 10$) is depicted in Fig. 6.1*b*. When δ is small, so that the number of packets in the network from one time frame to the following does not change significantly, $\alpha = 1/2$. On the contrary, when δ is sufficiently large, the exponent is 1. This is more in consonance with the results in [195], where the dependence with the noise level was addressed only for a low value of M, getting that as δ increases $\alpha \to 1$. On the other hand, we observe again that fixing M and varying δ does not guarantee the existence of a unique exponent for the scaling of fluctuations in traffic flow, though in this case the dependence is smoother than that observed in Fig. 6.1*a*.

In the following we show that expression (6.7) predicted by the RD model is indeed valid for more elaborated traffic models, and that the RD approximation captures the phenomenology of real communication systems. We report the

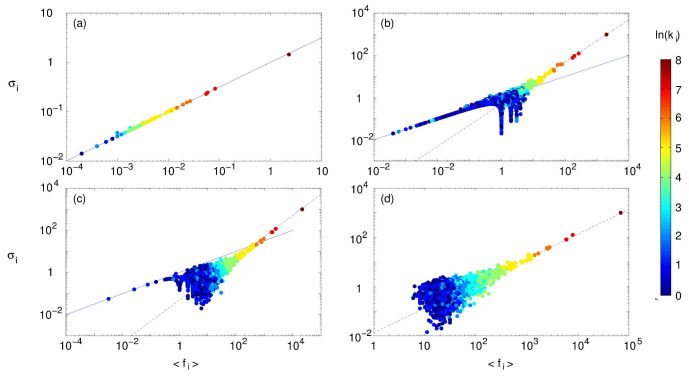

Figure 6.2: Flow fluctuation σ as a function of $\langle f \rangle$ from numerical simulations of the Internet traffic model (see text for details) on synthetic scale-free networks with $N = 10^4$ nodes, $K = 37551$ links, and degree exponent $\gamma = 2.2$. Different panels correspond to different values of M, respectively $M = 1, 5 \times 10^2, 35 \times 10^3, 10^5$. Color-coded values represent the logarithm of node degree. The continuous line is the curve $y = x^{0.5}$, while the dashed line is $y \sim x$.

results obtained on top of synthetic scale-free (SF) networks with $N = 10^4$ nodes and power-law degree distributions $p_k \sim k^{-\gamma}$, with an exponent $\gamma = 2.2$ as the one empirically observed for the Internet at the autonomous system level [189]. However, we stress that since the topological properties of the underlying graph only enter into Eq. (6.7) through the degree of the nodes k_i and the total number of links in the network, K, the results hold for any graph with an arbitrary degree distribution p_k as our own simulations using SF networks, random graphs and a real autonomous system map of the Internet [189] reveal.

On the other hand, to mimic the way packets flow in real communication networks, we consider a dynamical model that is able to simulate Internet's most important dynamical characteristics [192, 193]. The dynamics of the packets is simulated as follows. Each node represents a router with an infinite size buffer. The delivery of packets is made following a First In First Out (FIFO) policy. At each time step, p new packets are introduced in the system with randomly chosen sources and destinations [1]. Packets routing is based

[1] We have checked that the behavior observed does not depend on the creation rate of

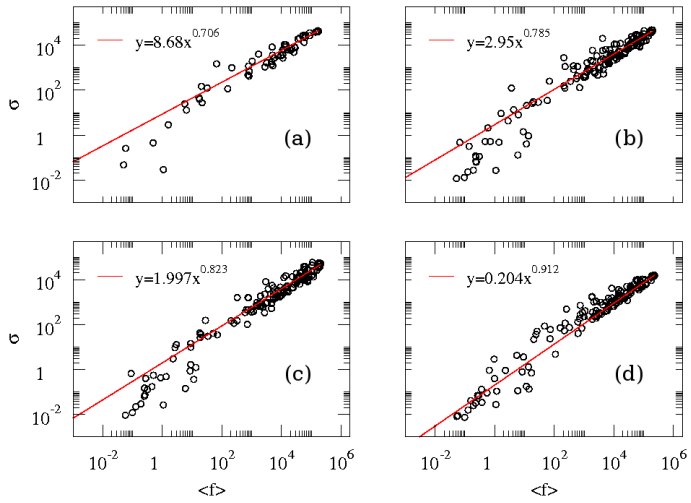

Figure 6.3: Flow fluctuation σ as a function of $\langle f \rangle$ for the Abilene Interfaces. The values of M used in each panel are: $M = 5$ (a), $M = 30$ (b), $M = 60$ (c), and $M = 720$ (d). Time is in minutes. The value of α for each M is also reported. Averages are taken over one month of data corresponding to the period between January 11 to February 11 of 2006.

on a traffic-aware scheme [192, 193] in which the path followed by a packet is that that minimizes the effective distance $d^i_{\text{eff}} = hd_i + (1-h)c_i$, where d_i is the distance between node i and the packet destination, c_i is the number of packets in i's queue, and h is a tunable parameter that accounts for the degree of traffic awareness incorporated in the delivery algorithm [192, 193]. It is worth recalling that $h = 1$ recovers a shortest-path delivery protocol, mimicking most of the actual Internet routing mechanisms.

Figure 6.2 shows σ as a function of $\langle f \rangle$ obtained through extensive numerical simulations of the traffic model with $h = 1$ and $p = 2$. Different panels in the figure correspond to different values of the time-window length M. The results indicate that the main responsible of the value of α (interpolating between the two extreme $\alpha = 1/2$ and $\alpha = 1$) is the interplay between the node degree and

packets, by considering uniform, exponential and power law distribution functions (in all cases, the distributions are characterized by the same average number of packets p).

the time resolution used to record the flux of packets, exactly as predicted by the scaling law (6.7) obtained in the RD model. In fact, Fig. 6.2*a* corresponds to the choice of parameters for which formula (6.8) holds for all values of k_i, leading to $\alpha = 1/2$. On the contrary, when M is large enough and the other parameters are kept fixed as in Fig. 6.2*d*, relation (6.8) is not satisfied whatever the value of k_i used, hence giving an exponent $\alpha = 1$. Finally, the breakdown of the scaling law $\sigma \sim \langle f \rangle^{\alpha}$ anticipated by the RD model is captured in Figs. 6.2*b* and *c*, where it is clearly revealed that there is not a unique exponent characterizing the flow of packets through *every* node of the network. Indeed, there is a crossover from $\sigma \sim \langle f \rangle^{1/2}$ for lowly connected nodes to $\sigma \sim \langle f \rangle$ for the highly connected ones. We also note that a similar behavior is observed (figures not shown) when traffic-aware routings ($h < 1$) are taken into account.

We have also analyzed the data corresponding to the traffic between routers of the Abilene backbone network [2]. As the data collected for the routers in the backbone correspond only to the flow between them, this backbone network can be viewed as an isolated communication system where the routers create, delivery and receive data packets. Therefore, the measures effectively correspond to a small network handling a large amount of traffic and whith all its nodes having a similar degree. For this reason, we are not able to observe here the dependence with the node degree. However, at variance with the analysis performed in [197], we have varied the length of the time windows used to extract the flux and its deviation [3]. Once again, the results, depicted in Fig. 6.3, show that the exponent α is not universal and radically depends on M. Note that, although the lower bound of $\alpha = 0.706 > 1/2$ is determined by the minimal resolution ($M = 5$ minutes) of the raw data, further increasing M will recover the upper bound $\alpha = 1$.

As a further test we analyzed data produced by a network simulator software NS2. NS2 is a discrete event simulator that provides substantial support for simulation of TCP, routing, and multicast protocols over wired and wireless (local and satellite) networks. It permits to fully reproduce most of the characteristics of a real computer network, implementing the standard TCP/IP stack and several types of connection media. To collect the data we firstly generate

[2] Data publicly available at http://abilene.internet2.edu.

[3] We are implicitly assuming that packets are uncorrelated. This approximation seems not to crude as the lifetime of packets in the Internet is at most of several seconds, while the minimal resolution of the raw data is 5 minutes. After all, the 5 minutes time window can also be considered as cumulative data. We have, however, redone the calculations taking nonconsecutive time windows to further avoid possible correlations with no qualitative changes in the results.

a BA network with $N = 1000$ nodes, then ran the NS2 simulator with randomly choosen origins and destantions at constant packets creation rate (no fluctuations in system's arrivals) and with all the TCP/IP stacks protocols activated. While the simulations ran, we collect the flow data on each node with the highest resolution made possible by the simulator (about 10^{-3} seconds) and comparable with traffic dynamics time scale (about 10^{-4} seconds). Then we computed for each node the mean flow $\langle f_i \rangle$ and its fluctuations σ_i. Results, presented in Fig. 6.4, show that all exponents α between $0.5 \leq \alpha \leq 1$ are recovered (as prediceted by the theoretical model) and, as no external fluctuations are presents and the sampling time scale is comparable to dynamics time scale, the observed behavoir is a truly effect of nodes degrees heterogeneity

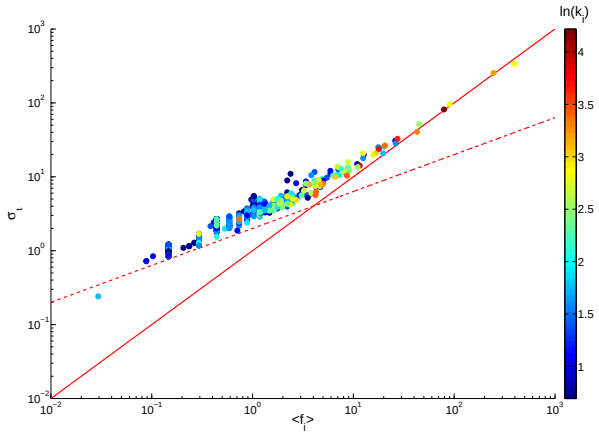

Figure 6.4: Flow fluctuation σ as a function of $\langle f \rangle$ from NS2 simulations for a BA Network with N=1000 and $m_0 = 2$. The dashed line is the curve $y = x^{0.5}$, while the continuous line is $y \sim x$.

6.3 Conclusions

Concluding, in this chapter we have derived a theoretical law for the dependence of fluctuations with the mean traffic in a network. Such a dependence is governed by three factors: one related to the dynamics, one related to the topol-

ogy, and one of statistical nature. More importantly, the theoretical law reveals that the previously claimed power-law scaling (with universal or non-universal exponents) has to be abandoned. Our numerical results and the analysis of real data confirm that, even in the presence of correlations between packets, one cannot assume a single exponent to characterize the fluctuations of traffic for the whole network. Finally, we note that the scaling breakdown predicted here is amenable to experimental confirmation by measuring the traffic flow in large communication networks so to capture the predicted (topological) effects of degree heterogeneity.

Chapter 7

Empathy Minimizes Congestion in Communication Networks

Lot of the recent literature has tackled the critical properties of jamming and congestion transitions in communication networks [92, 93, 94, 95, 96, 97, 98, 99, 101, 100]. It has been shown that finding the best suited strategy depends strongly on two main features: the topological patterns of the particular network and the load of information on top of it. Regarding the first of these two issues, a number of routing mechanisms have been studied on different structures [102, 103, 101, 104] allowing to design resilient network backbones [105, 106, 107].

Many of the routing policies proposed so far rely on the (static) structural properties of the communication network. Examples of such policies are biased random walks [108, 109], shortest-path [110, 111] and efficient-path [112] schemes. These routing mechanisms can be conveniently reformulated to incorporate the information about the dynamical state of the system, *i.e.* the congestion state of routers. This allow to dynamically change the paths followed by information packets in order to bypass those over-congested routes. In this line, congestion-aware schemes have significantly improved the performance of or the addition of a router memory to avoid packets getting trapped between two adjacent nodes [120].

All the above studies have assumed that both network topology and the mechanisms to avoid congestion are static (*i.e.* neither topology nor the routing strategies change). However, this approach neglects that, even for the same graph, the optimal routing policy depends strongly on the state of congestion

of the system [114, 115, 117, 118, 121, 122]. Therefore, in order to balance correctly the congestion in a communication system it seems appropriate to allow the elements (routers) to switch to the best suited strategy to avoid congestion given the instant traffic conditions. In this chapter, we propose an adaptive mechanism that allows nodes to choose their individual strategies instead of imposing a common policy. In this adaptive protocol routers exploit their local information about the congestion state of the system to decide whether to accept incoming packets. First, in section 7.1, we introduce a minimal routing model without any adaptive mechanism that allow us to unveil the role of the rejection mechanism when it is externally tuned. In section 7.2 we will consider that each router can adopt its own rejection strategy and make some analytical derivations about the optimal strategic configuration to avoid the congestion onset. In section 7.3 we will implement our first adaptive mechanism and show that when nodes are allowed to dynamically adapt their own strategy while only being aware of their own congestion state (myopic case), the onset of congestion is shifted to a larger critical load (with respect of the static algorithm introduced in section 7.1). This improvement is due to the self-organization of the strategies of nodes into degree-correlated configurations. However, we will show that the delay of the onset of congestion comes together with a sharp, first-order like, transition that provides no dynamical signals about the onset of congestion. Finally, in section 7.4 we show that when nodes are allowed to know the congestion state of its nearest neighbors and empathize with them, it is possible to recover the former large critical load together with a smooth phase transition, avoiding the uncertain scenario of the myopic adaptive model. More importantly, we will show that tuning conveniently the degree of empathy between routers it is possible to recover, through a local mechanism, both the congestion levels and the rejection patterns provided by the global minimization introduced in section 7.2.

7.1 Minimal traffic model

Let us start by introducing the minimal traffic model in which the adaptive algorithm will be implemented below. In this model, we consider the transfer of information packets between adjacent routers as a probabilistic event. Inspired by [117, 118] we consider a set of stochastic equations for describing the time evolution of the queue length of the nodes at some time t, $q^t = \{q_i^t\}$. The queue length of a given node, q_i^t, can either increase or decrease due to several events. First, at each time step with some probability p a new packet is generated

being added to the queue of the node. Second, at each time step each node tries to send a packet in its queue to any of its first neighbors. This packet can be rejected by the chosen neighbor with some probability η. If the packet is accepted, it may be removed from the system with certain probability μ. These two latter events mimic the effects, although with some important differences, of an active queue control strategy as the random early detection (RED) [123] present on Internet routers and the arrival of the packet to its final destination respectively. Following the above ingredients we can write the time-discrete Markov chain of the minimal traffic model as:

$$\begin{aligned} q_i^{t+1} = q_i^t + p \quad &+ \quad \sum_{j=1}^{N} \frac{\Theta(q_j^t) A_{ji}}{k_j} (1-\mu)(1-\eta) \\ &- \quad \Theta(q_i^t) \sum_{j=1}^{N} \frac{A_{ij}}{k_i} (1-\eta) \ , \end{aligned} \tag{7.1}$$

where A_{ij} represents the $(i,\ j)$ term of the adjacency matrix of the network substrate and $\Theta(x)$ is the Heaviside step function ($\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ otherwise). Since our network is undirected and unweighted, the adjacency matrix is defined as $A_{ij} = A_{ji} = 1$ if nodes i and j are connected and $A_{ij} = A_{ji} = 0$ otherwise. The quantity k_i is the degree of a node i ($\sum_j A_{ij} = k_i$), *i.e.* the number of routers connected to it. Two terms in the right-hand-side of equation (7.1) account for the incoming flow of packets that arrive to the queue of node i, namely, p (accounting for the external load of packets) and the first sum (accounting for the arrival of packets from its first neighbors). On the other hand, the second sum in equation (7.1) accounts for the probability that a packet from i is delivered to a first neighbor.

The set of equations (7.1) are solved starting from a zero congestion state: $q_i^0 = 0\ \forall i$. The evolution of the system is monitored by means of the following order parameter [95]:

$$\rho = \lim_{T \to \infty} \frac{Q(t+T) - Q(t)}{pT} \ , \tag{7.2}$$

where $Q(t)$ is the sum of all the queue lengths at time step t, $Q(t) = \sum_{i=1}^{N} q_i^t$. The value of the above order parameter is bounded ($0 \leq \rho \leq 1$) and describes the dynamical regime in which the system ends up. Namely, $\rho = 0$ indicates that the system is able to balance the incoming flow of new packets with a successful delivery of the old ones. In this case the system is said to operate

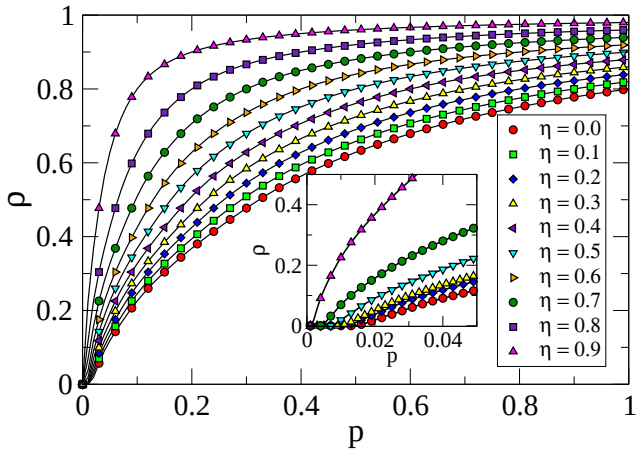

Figure 7.1: Phase diagrams, $\rho(p)$, of the minimal traffic model using different values of the rejection rate η. The inset shows the existence of different critical values p_c when varying η

in the *free-flow regime*. Instead, when $\rho > 0$ the above balance is not fulfilled and the queues of the nodes increase their size in time at a rate $\rho \cdot p$. In this latter situation the system is in the *congested phase*.

We have studied the behavior of the order parameter ρ taking the rate of packet creation p as the control parameter. The arrival-to-destination probability is set to $\mu = 0.2$ as the usual value found in the Internet [84]. The corresponding phase diagrams are shown in Fig. 7.1 for several values of the rejection probability η using a SF network of $N = 5000$ with $P(k) \sim k^{-2.2}$. As observed in the figure, the transition from free-flow to congestion occurs in a smooth way at low values of p being the critical point $p_c = 0.02$ for $\eta = 0$ (no rejection). However, as the rejection rate η increases the value of p_c decreases and ρ increases faster (see inset in Fig. 7.1).

7.2 Analytical approximation of global congestion minimization

The above results question the convenience of implementing a rejection mechanism in routing models. However, the bad performance of this rejection mechanism relies on the homogeneous distribution of the rejection rates across the routers of the network. We now explore the general situation in which the individual rejection rates are independent. Therefore the set of equations (7.1) transforms into:

$$\begin{aligned} q_i^{t+1} = q_i^t + p \quad &+ \quad \sum_{j=1}^{N} \frac{\Theta(q_j^t) A_{ji}}{k_j} (1-\mu)(1-\eta_i) \\ &- \quad \Theta(q_i^t) \sum_{j=1}^{N} \frac{A_{ij}}{k_i} (1-\eta_j) \; . \end{aligned} \tag{7.3}$$

This new set of equations is now used to determine the optimal set $\{\eta_i\}$ so that congestion is minimized for a given value of p. To this aim, we first use two assumptions: *(i)* the node has reached a stationary state, $q_i^{t+1} = q_i^t \; \forall i$, and *(ii)* the queue length of nodes is nonzero, $\Theta(q_i^t) = 1$. These provisos admitted, equations (7.3) turn into the following set of equations for the rejection rates of the routers $\{\eta_i\}$:

$$0 = p + \sum_{j=1}^{N} \frac{A_{ji}}{k_j} (1-\mu)(1-\eta) - \sum_{j=1}^{N} \frac{A_{ij}}{k_i} (1-\eta) \; . \tag{7.4}$$

Now we make use of the annealed approximation of the adjacency matrix [124, 125, 126]:

$$A_{ij} = A_{ji} = \frac{k_i k_j}{N \langle k \rangle} \; , \tag{7.5}$$

where $\langle k \rangle$ is the average degree of the network ($\langle k \rangle \simeq 4$ in our case). Introducing the annealed expression (7.5) into equations (7.4) we obtain:

$$k_i (1-\eta_i) = \frac{1}{1-\mu} \left[\langle k(1-\eta) \rangle - p \langle k \rangle \right] \; , \tag{7.6}$$

where $\langle k(1-\eta) \rangle = \sum_j k_j (1-\eta_j)/N$. Equation (7.6) clearly shows that the larger the degree of a router the larger its rejection rate. Therefore, from this expression we observe that a non-homogeneous distribution of rejection rates

across the routers is beneficial to assure the free-flow condition (and thus to delay the onset of congestion). We can calculate the expression of the rejection rate by computing the value of $\langle k(1-\eta)\rangle$. From equation (7.6) we obtain:

$$\langle k(1-\eta)\rangle = \frac{1}{1-\mu}\left[\langle k(1-\eta)\rangle - p\langle k\rangle\right] , \tag{7.7}$$

and finally we have:

$$\langle k(1-\eta)\rangle = \frac{p}{\mu}\langle k\rangle . \tag{7.8}$$

Therefore, the rejection rate of a node with connectivity k_i reads:

$$\eta_i = 1 - \frac{p\langle k\rangle}{\mu k_i} . \tag{7.9}$$

As anticipated above, expression (7.9) shows that the rejection rates of nodes should depend on their degrees rather than being externally set to a constant value. In Fig. 7.2 we apply equation (7.9) to plot the rejection patterns corresponding to different values of the external load p. As shown η_i decreases with p and increases with k_i.

The assumptions made in order to obtain equation (7.9) point out that the validity of this expression, for all the nodes, should be restricted to the proximity of the critical point p_c. First, for $p < p_c$ many of the queues are zero (invalidating assumption (*ii*)) thus making the rejection rate imposed by equation (7.9) too restrictive for the real traffic conditions. On the other hand, for $p > p_c$ assumption (*i*) does not hold for all the nodes. This is manifested by the prediction of negative rejection rates, $\eta_i < 0$, in equation (7.9) for those nodes with low connectivity. In practice, the impossibility of displaying negative rejection rates fix their rejection rate to $\eta_i = 0$. However, those nodes with large enough connectivity can still avoid congestion by means of positive rejection rates as described in equation (7.9) (see Fig. 7.2). Following these arguments, we can estimate the exact value of p_c as the maximum value of p for which $\eta_i \geq 0$ for all the nodes in the network. In particular, given that, for a given p, the value of η_i increases with k_i we obtain p_c imposing in equation (7.9) that those nodes with the minimum connectivity, $k_i = k_{min}$, have $\eta_i = 0$. Since in our case $k_{min} = 2$ and $\langle k\rangle \simeq 4$ we obtain $p_c \simeq 0.1$. Therefore, by externally fixing the rejection rate of each node we can assure the permanence in the free-flow phase up to $p_c \simeq 0.1$.

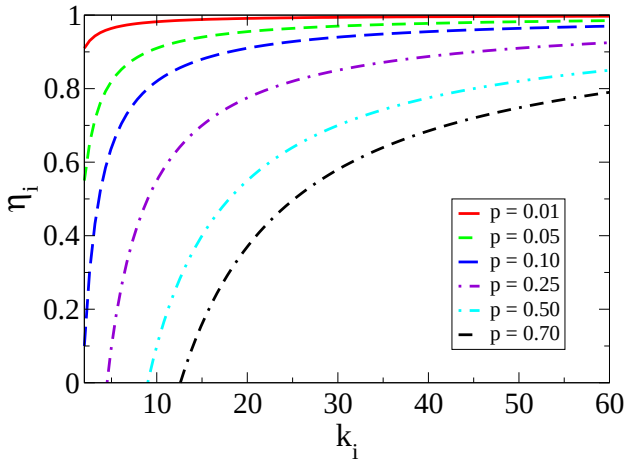

Figure 7.2: Rejection rates of nodes as a function of their degree $\eta_i(k_i)$ as obtained from equation (7.9). The curves correspond to different values of the external load of information p.

7.3 Myopic adaptability

The minimal traffic model introduced in section 7.1 shows that system's performance deteriorates as soon as rejection rates are uniformly set in the system. However, in section 7.2 we have shown that a non-uniform configuration for the rejection rates shifts the critical load to larger values. However, this non-uniform configuration has been externally imposed and derived analytically following different assumptions. A correct derivation of the optimal configuration would imply, on one hand, a more sophisticated calculation and, on the other hand, a complete knowledge of the architecture of the network. This latter condition makes unrealistic the external tuning of the individual rejection rates.

In order to overcome the need of global knowledge about the topology of the network we now introduce an adaptive scheme based solely on the local information available to nodes. In this adaptive setting we will allow nodes to

choose their own rejection rate so that the dynamical state of a node will be described by both q_i^t and η_i^t:

$$\begin{aligned} q_i^{t+1} = q_i^t \quad &+ \quad p + \sum_{j=1}^{N} \frac{\Theta(q_j^t) A_{ji}}{k_j} (1-\mu) \left[1 - \eta_i^t\right] \\ &- \quad \Theta(q_i^t) \sum_{j=1}^{N} \frac{A_{ij}}{k_i} \left[1 - \eta_j^t\right] \end{aligned} \tag{7.10}$$

The individual choice of each instant value η_i^t aims at operating at the optimal regime as given by the external parameters p and μ. To this aim, each node chooses its own rejection rate for the following time-step attempting to reach an optimal queue length, $q^{opt} = p/\mu$, so that traffic is homogeneously distributed across the network. To this end, a node raises or decreases its own rejection rate depending on the deviation of its instant queue length from the optimal queue, $\Delta_i^t = q_i^t - q^{opt}$. This rationale mimics a myopic behavior by which, regardless of the congestion state of the system, nodes are allowed to close the door to new packets while decreasing their respective queues. To incorporate this adaptive behavior we couple equations (7.10) with the following equations for the set $\{\eta_i^t\}$:

$$\eta_i^{t+1} = \frac{1}{1 + \exp\left(-\beta \Delta_i^t\right)} \, . \tag{7.11}$$

This evolution rule takes the form of the saturated Fermi function so that congested nodes, $q_i > q^{opt}$, will tend to total rejection, $\eta_i^{t+1} \to 1$, whereas those under-congested will open the door to new packets, $\eta_i^{t+1} \to 0$. The velocity of the transition from these two regimes is controlled by β so that it controls the reactivity of nodes to congestion. while $\eta_i^{t+1} = 0.5$ will be adopted when $q_i^t = q^{opt}$.

The adaptive equations (7.11) allow for abrupt changes in the rejection rates between two consecutive time steps. Thus, we also explore a different formulation:

$$\eta_i^{t+1} = \eta_i^t + \beta \Delta_i^t \, , \tag{7.12}$$

in which the rejection rates evolve smoothly. Rule (7.12) is completed by assuring that η_i remains bounded so that $0 \leq \eta_i \leq 1$. In the above equation (7.12), β acts as the inverse of the time between two consecutive time steps of the adaptive dynamics. Therefore, in the continuous time approximation of equation (7.12), the derivative of the rejection rates is equal to the difference

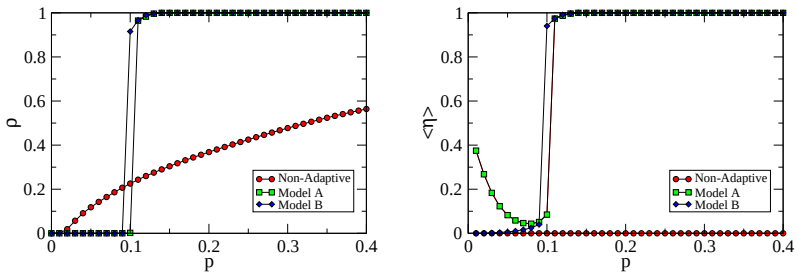

Figure 7.3: (Left) Phase diagram $\rho(p)$ for the myopic routing models A (squares) and B (diamonds) and for the minimal routing model (circles). (Right) Average rejection rate $\langle\eta\rangle$ as a function of p of the former three routing schemes.

between the instant queue length and its optimal value, *i.e.* $\Delta_i^t = q_i^t - q^{opt}$. Note that in this setting when $q_i^t = q^{opt}$ a router will adopt $\eta_i^{t+1} = 0$.

In the following we will use the two formulations for the myopic adaptive model and show that the results are qualitatively the same. Namely, we will call model A to equations (7.10) and (7.11), and model B to the formulation using equations (7.10) and (7.12). Note that in both models the parameter β controls the reaction speed of nodes to congestion. In this direction, our numerics have shown that by changing β one basically controls the duration of the transient time before the stationary distribution of the rejection rates is reached. In the following, we set $\beta = 10$ and $\beta = 10^{-2}$ in models A and B respectively.

In the left panel of Fig. 7.3 we show the phase diagram, $\rho(p)$, of the myopic adaptive model with the two formulations. As observed, in both formulations the myopic model displays an abrupt, first-order like, transition from the free-flow to the congested state. Moreover, in Fig. 7.3 we have also plotted the phase diagram of the minimal model when $\eta = 0$, *i.e.* its most congestion-resilient version, to show the improvement of myopic adaptability by shifting the jamming transition from $p_c = 0.02$ to $p_c \simeq 0.1$. This value for the critical load is exactly the same as the one predicted in section 7.2 using the analytical approximation with global knowledge. Thus, the myopic adaptive model, equals the delay predicted by minimizing congestion globally.

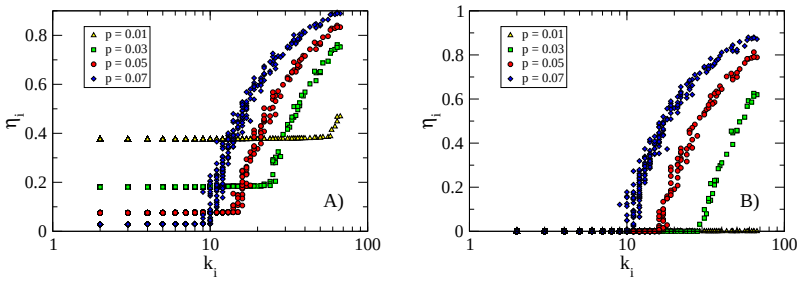

Figure 7.4: Distribution of the individual rejection rates η_i across degree-classes for several values of p in the myopic routing models A (left) and B (right).

To analyze the roots of the resilience of the myopic adaptive routing to congestion we have plotted in the right panel of Fig. 7.3 the mean value of the rejection rate, $\langle \eta \rangle = \sum_{i=1}^{N} \eta_i$. In this case we observe that models A and B display the same pattern after the sharp transition to congestion, *i.e.* the sudden closing of all the doors in the network thus causing the abrupt transition to $\rho \simeq 1$ as soon as $p > p_c$. On the other hand, the configurations adopted by both models before the onset of congestion, $p < p_c$, are quite different: While in model B $\langle \eta \rangle \simeq 0$, for model A a significant part of the population adopts $\eta_i > 0$. Surprisingly, in this latter setting the average rejection rate decreases as we approach the critical point, p_c.

To have a deeper insight about the microscopic configurations that allow to delay the onset of congestion we show in Fig. 7.4 the set of individual rejection rates of nodes $\{\eta_i\}$ ranked according to their degrees. In both models A and B, the correlation between η_i and k_i is clear since all the routers within the same degree-class display similar rejection rates. First, in model A we observe that for $p = 0.01$ the system self-organizes homogeneously around $\eta \simeq 0.4$. However, when p increases the rejection rates of low-degree classes decreases while hubs start to close their doors progressively as p increases. For model B the microscopic configurations adopted as p increases are similar regarding the behavior of high-degree nodes. However, in this latter scenario low-degree nodes remain accepting incoming packets up to the congested state. These two figures show that the two different internal dynamics (showing different microscopic organizations) lead to the same macroscopic result: the delay of

the onset of congestion.

Let us highlight that the delay of the congestion onset in this myopic adaptive setting again contradicts the results obtained for the minimal routing model in which, even a small (homogeneously distributed across routers) rejection rate leads to an increase of the congestion in the system. Quite on the contrary, the myopic adaptive model points out the same idea concluded from the global minimization of congestion: a hierarchical (degree-based) organization of the rejection rates by the system is strongly beneficial to avoid the congestion of the system. However, it is important to note that the self-adopted strategies differ strongly from those configurations found in section 7.2. From figure 7.4 it becomes evident that the strategies adopted by the myopic adaptive settings are clearly different than the ones obtained from equation (7.9) by minimizing congestion using global knowledge. Although in equation (7.9) the value of the rejection rate increases with the degree of the node (as in the myopic settings) in this case the evolution with p is basically different. Thus, although the critical load has been shifted to the same value as in the one found in section 7.2, the self-organized patterns of the rejection rates in the myopic settings reveal a clearly different scenario.

7.4 Empathetic adaptability

The myopic adaptive setting has improved remarkably the resilience to congestion without the need of tuning any external parameters. However, the existence of an abrupt phase transition, again as found in [114, 115, 111, 120], demands for further improvements. The main goal in order to soften such abrupt transition is to avoid that all the nodes close their doors due to its own congestion by incorporating an empathetic behavior based on the local knowledge about the dynamical state of their neighbors. This empathetic behavior should motivate congested nodes to open their doors when detecting an hyper-congested state in its surroundings. To this aim we take model B (note that the generalization of model A can be done in the same way) and reformulate its equations as follows:

$$\eta_i^{t+1} = \eta_i^t + \beta \left[(1-\alpha)\Delta_i^t - \alpha \langle \Delta_j^t \rangle_{\Gamma_i} \right] \ . \tag{7.13}$$

In the above equations we introduce a new term accounting for the average level of congestion in the neighborhood, Γ_i, of a node i,

$$\langle \Delta_j^t \rangle_{\Gamma_i} = \sum_{j=1}^{N} \frac{A_{ij}}{k_i} \Delta_j^t \ . \tag{7.14}$$

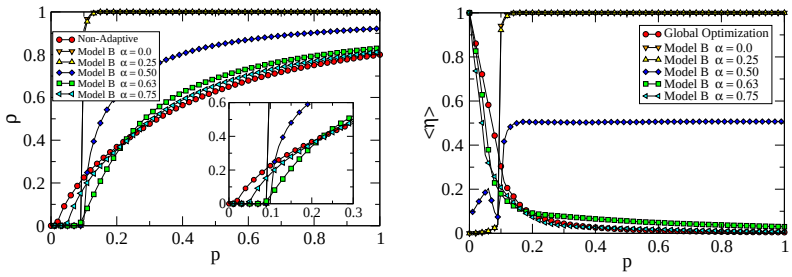

Figure 7.5: (Left) Phase diagram $\rho(p)$ of the empathetic routing model for several values of the empathy parameter α. The phase diagram of the minimal routing model (circles) is also plotted for the sake of comparison. (Right) For the same empathy parameters we show the average rejection rate $\langle\eta\rangle$ as a function of p. The function $\langle\eta\rangle(p)$ obtained analytically from global minimization and computed from equation (7.9) is also shown.

The relative importance that nodes assign to the local level of congestion in their neighborhoods with respect to their own state is controlled by the parameter α. In particular, when $\alpha = 0$ we recover the myopic setting whereas for $\alpha = 1$ routers behave "altruistically" and their decisions are based solely on their neighbor's state of congestion. Thus, the parameter α measures the degree of empathy of routers.

In the left panel of Fig. 7.5 we plot the phase diagrams for several values of α together with that of the minimal non-adaptive routing model for the sake of comparison. From our simulations we found that for $\alpha < 0.5$ the phase-transition is similar to that of the myopic adaptive model ($\alpha = 0$), *i.e.* having a critical load of $p_c \simeq 0.1$ followed by a first-order transition to full congestion. However, from the figure we observe that when $\alpha > 0.5$ the transition to congestion occurs smoothly, thus recovering the behavior of the minimal model. On the other hand, the value of p_c also decreases with α (thus anticipating the onset of congestion) although it remains close to the original value $p = 0.1$ until $\alpha \simeq 0.63$. For this latter value $\alpha = 0.63$ we identify the optimal operation point since the onset of congestion is delayed as much as in the myopic model but displaying a second-order phase transition reaching similar levels of congestion to those observed in the minimal model.

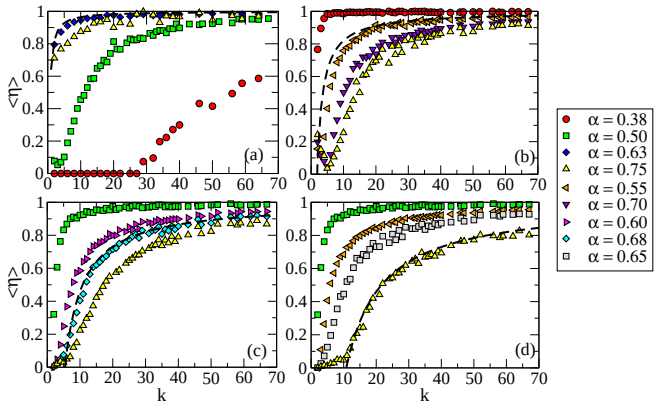

Figure 7.6: Distribution of the mean rejection rates $\langle\eta\rangle$ across degree-classes of the empathetic adaptive model for several values of the empathy parameter α compared with the global minimization prediction (dashed line). Different traffic values p are presented: (a) $p = 0.02$ (free-flow regime), (b) $p = 0.1$ (critical point), (c) $p = 0.3$ and (d) $p = 0.6$ (congested state)

In order to unveil what is the strategy adopted by the system to reach such optimal phase diagram we have computed the average level of rejection rate as a function of p for the relevant values of α. In the right panel of Fig. 7.5 we observe that those curves corresponding to $\alpha > 0.5$ are quite different from those obtained in Fig. 7.3 for the myopic adaptive setting. In particular, when $p \ll p_c$ the empathetic adaptability shows a large amount of rejection. However, as p increases the average rejection rate decreases monotonously. This high-rejecting behavior for $p < p_c$, was not observed in the myopic scheme. Quite on the contrary, it was shown that nearly all the doors were open in the sub-critical regime. However, the high rate of rejection observed in the right panel of Fig. 7.5 is due to the large degree of empathy ($\alpha > 0.5$) and the existence of a number of under-congested nodes, $\Delta_i < 0$, in the sub-critical regime. Under these low traffic conditions, most nodes will close partially their doors when detecting under-congested neighborhoods, $\langle\Delta_j^t\rangle_{\Gamma_i} < 0$, in order to benefit from the availability of neighbors to handle their packets. This situation is highly dynamical and most of the nodes experiment large fluctuations in

their rejection rates until the system equilibrium is reached. This microscopic scenario, although clearly different from that of the myopic setting, is able to delay the onset of congestion in an efficient way. As p approaches p_c and for $p > p_c$ we observe in Fig. 7.5 that $\langle\eta\rangle$ decreases to 0 as p increases when $\alpha > 0.5$. This decrease is due to both the large number of over-congested neighborhoods, $\langle\Delta_j^t\rangle_{\Gamma_i} > 0$, surrounding routers in the super-critical regime and their large degree of empathy. This large degree of empathy prevents from the sudden door closing in when $p > p_c$, thus favoring a smooth transition to congestion reaching similar congestion levels to those observed in the minimal routing model in the super-critical regime.

At variance with the myopic adaptive setting, the monotonous decrease of $\langle\eta\rangle(p)$ from $\langle\eta\rangle = 1$ at $p = 0$ points out a similar behavior to that obtained by means of global minimization of congestion. As shown in the right panel of Fig. 7.5 the theoretical estimation of $\langle\eta\rangle(p)$ (circles) follows the same trend as the self-adopted patterns for $\alpha > 0.5$. To analyze in detail the similarity between the empathetic setting and the microscopic patterns predicted by global minimization of congestion we plot in Fig. 7.6 the average value of the rejection rate as a function of the degree k of the nodes, $\langle\eta\rangle(k)$, for several values of p and α. The panels correspond to (a) $p = 0.02$ (free-flow regime), (b) $p = 0.1$ (critical point), (c) $p = 0.3$ and (d) $p = 0.6$ (congested state). The shape of each curve $\langle\eta\rangle(k)$ behaves similarly to the theoretical one. More importantly, for each value of p there is one value of α, α^{opt}, for which the curve $\langle\eta\rangle(k)$ fits perfectly the prediction made by global minimization of congestion. The precise value of α^{opt} depends on p. In particular, for $p = 0.02$ we find $\alpha^{opt} \simeq 0.63$, for $p = 0.1$ we obtain $\alpha^{opt} \simeq 0.55$, for $p = 0.3$ we have $\alpha^{opt} \simeq 0.68$ and, finally, for $p = 0.6$ the value found is $\alpha^{opt} \simeq 0.75$. Moreover, from the left panel of Fig. 7.5, we observe that the values found for α^{opt} are those for which congestion, $\rho(p)$, is minimum. This result points out that empathetic adaptability is able to avoid congestion by means of only local information as much as global minimization does.

7.5 Conclusions

We have studied a novel mechanism that allow routers to adapt their individual strategies based on their local knowledge about congestion. Although in our approach nodes can only decide either to refuse or to accept incoming packets from their first neighbors, we obtain a variety of dynamical behaviors. First, we have analyzed the situation when no individual adaptability is allowed. This

allow us to show that whenever a small level of rejection is applied indistinctly to all the nodes, one obtains a worse overall behavior than when all incoming flows are accepted by the routers. Then, we have considered that routers can have different rejection rates and derived analytically their patterns to minimize congestion, considering global knowledge of the network topology. With these globally optimized patterns the resilience to congestion of the system can be enhanced significantly. Besides, these patterns reveal a dependence of the rejection rate and the degree of the router and its mean value decays with the incoming load of packets.

After the global minimization of congestion we have studied the situation in which nodes self-adjust their own rejection rate dynamically depending on their instant level of congestion (myopic setting). In this case we have shown that the critical load of the network is shifted to a value similar to that found analytically by means of global minimization of congestion. This improvement is again achieved by a proper distribution of the rejection rates according to the degrees of the routers. However, in the adaptive case, such degree-correlated configuration is self-tuned by the system and differs from that obtained analytically. As usual in congestion-aware routing schemes, such delay in the congestion onset comes together with an abrupt transition from the free-flow phase to the congested one that prevents from having any warnings of the approach to the onset of congestion. For this reason, we have finally explored the situation in which routers also consider the congestion state of their first neighbors to adapt their rejection rates. We have shown that when nodes empathize with the congestion state of their neighbors, thus not rejecting packets from them when they detect an over-congested neighborhood, the shift in the critical load (obtained through global minimization and the myopic adaptability) is preserved and followed by a smooth congestion transition. Moreover, the analysis of the microscopic patterns of rejection rates when empathy is the mechanism at work point out a similar organization to that obtained from global minimization. In particular it is possible to find the degree of empathy that perfectly agrees with the analytical estimation of the rejection pattern that minimize congestion for a give information load.

In summary, we have shown that allowing routers to adapt their own strategies together with a certain degree of local empathy is strongly beneficial to the behavior of complex communication systems. Moreover, the improvement shown when local empathy is at work is similar to that obtained by minimizing congestion by means of a global knowledge of the network topology. Thus, the empathetic strategy represents, a remarkable example of how local rules can achieve levels of functioning as optimal as those obtained with global knowl-

edge of the system. Besides, our results open the relevant question about how local empathy can be naturally tuned as a function of the external inputs.

Part IV

Epidemic-like Dynamics on Complex Networks

Introduction

The problem of modeling how diseases spread among individuals has been intensively studied for many years [55, 128, 143, 152, 160]. The development of mathematical models to guide our understanding of the disease dynamics has allowed to address important issues such as immunization and vaccination policies [128, 153, 144]. Scientist's approaches to problems in epidemiology involve statistical physics, the theory of phase transitions and critical phenomena [170], which have been extremely helpful to grasp the macroscopic behavior of epidemic outbreaks [14, 130, 138, 146, 155, 158, 159, 163, 164]. The main artifice of this success has been the Mean-Field (MF) approximation, where local homogeneities of the ensemble are used to average the system, reducing degrees of freedom. It consists in consider all the nodes have the same dynamical properties, thus the mean behavior of the system is representative for all the nodes and assumes that fluctuations can be neglected.

The study of complex networks [1, 3, 4] has provided new grounds to the understanding of contagion dynamics. Scale-free networks have demonstrated to be a good approximation of the connection patterns of sexual contacts [154], the Internet [165], as well as other social, technological and biological networks [133] and thus their structure must be taken into account when considering a spreading scenario. Scale-free networks [1, 4, 129] are characterized by the presence of hubs, which are responsible for several striking properties for the propagation of information, rumors or infections [14, 130, 146, 155, 159, 163]. To consider this peculiar structure a modified version of the MF approximation the so-called *heterogeneous mean field* (HMF) is introduced [163]. In HMF nodes with the same degree are considered to have the same dynamical behavior and are thus grouped in degree classes. The system behavior is then obtained averaging over all the degree classes. The main accomplishment of the HMF approach is that it can analytically predicts the critical rate β_c at which the disease spreads, i.e. the epidemic threshold.

Theoretical modeling of how diseases spread in complex networks is largely based on the assumption that the propagation is driven by reaction processes, in the sense that the transmission occurs from every infected through all its neighbors at each time step, producing a diffusion of the epidemics on the network. However, this approach overlooks the notion that the network substrate is a fixed snapshot of all the possible connections between nodes, which does not imply that all nodes are concurrently active [148]. Many networks observed in nature [1, 4], including those in society, biology and technology, have nodes that temporally interact only with a subset of its neighbors [127, 162]. For instance, hub proteins do not always interact with all their neighbor proteins at the same time [151], just as individuals in a social network [154] do not interact simultaneously with all of their acquaintances. Likewise, Internet connections being utilized at a given time depends on the specific traffic and routing protocols. Given that transport is one of the most common functions of networked systems, a proper consideration of this issue will irreparably affect how a given dynamical process evolves. Starting from this assumption it's also possible consider a more complex scenario in which not only traffic shapes the actual connection that can be exploited by an epidemics but also a backward loop in introduced and the knowledge of an existing epidemic can alter the normal traffic activity.

In this forth part of the work we will focus on some open problems in epidemic modeling on complex networks and we will try to solve some of the problems stated above. Specifically we will address the following questions. The HMF although very effective in the prediction of the epidemic threshold β_c cannot be used to study the behavior of the system away of the critical point and cannot reproduce the microscopical behavior of single nodes. Is it possible to have some analytical insights of the single nodes state and study the entire epidemic phase diagram? Recent works propose very accurate models that include several classes of mobility but all rely on the assumption that propagation is similar to a reaction process. Is it possible to consider an epidemic spreading model in which interactions are not reactive but driven by traffic processes that better reproduce real world cases? And, in this latter case, can we model the impact of human responses to an epidemic spreading?

To answer the first of these questions in chapter 8 we introduce a microscopical analytical treatment of the problem of how a contact-based disease spreads on a complex network. To do so, we extend the classical HMF through a probabilistic discrete-time Markov chains approach that can be used to study the overall behavior of the system in the whole phase diagram and faithfully reproduce the single nodes infection probability. In chapter 9 we consider a

scenario in which interactions between individuals are not fully reactive (i.e. in the case of the classical HMF)but only a subset of the connections are considered simultaneously and are driven by a traffic process. We show that, in this case, the epidemic threshold β_c is strictly related to the amount of traffic on the network and the previous results of a vanishing threshold can be obtained as a particular case of random diffusion and high traffic values. Finally, in chapter 10 we will answer the third question: how human responses impact on the evolution of the spreading. We will present a meta-population model in which the contagion dynamics is shaped by traffic flows and individuals have a knowledge of the epidemic state in each node adapting their behavior to the epidemic conditions. We will show that in many cases trying to avoid infected nodes produces an increment in terms of number of infected populations and overall epidemic incidence. We will also present the evidence that traffic reduction policies are almost ineffective.

Chapter 8

Modeling Epidemic Spreading in Complex Networks

In this chapter, we present a theoretical framework for contact-based spreading of diseases in complex networks. This formulation, Microscopic Markov-Chain Approach (MMCA), is based on probabilistic discrete-time Markov chains, generalizes existing HMF approaches and applies to weighted and unweighted complex networks [148]. Within this context, in addition to capturing the global dynamics of the different contact models and its associated critical behavior, it is now possible to quantify the *microscopic dynamics* at the individual level by computing the probability that any node is infected in the asymptotic regime. Extensive Monte-Carlo (MC) simulations corroborate that the formalism here introduced reproduces correctly the *whole* phase diagram for model and real-world networks. Moreover, we capitalize on this approach to address how the spreading dynamics depends on the number of contacts actually used by a node to propagate the disease.

8.1 Microscopic Markov-Chain Approach to disease spreading

The critical properties of an epidemic outbreak in SF networks can be addressed using the HMF prescription [163, 164, 155, 159, 14, 130, 146, 158]. It consists of coarse-grained vertices within degree classes and considers that all nodes in a degree class have the same dynamical properties; the approach also assumes that fluctuations can be neglected. Specifically, if β is the rate (probability per

unit time) at which the disease spreads, it follows that the epidemic threshold in uncorrelated SF networks is given [163] by $\beta_c = \langle k \rangle / \langle k^2 \rangle$, leading to $\beta_c \to 0$ as $N \to \infty$ when $2 < \gamma \leq 3$.

MF approaches are extremely useful to assess the critical properties of epidemic models, however they are not designed to give information about the probability of individual nodes but about classes of nodes. Then, questions concerning the probability that a given node be infected are not well posed in this framework. To obtain more details at the individual level of description, one has to rely on MC simulations, which have also been used to validate the results obtained using MF methods. Restricting the scope of epidemiological models to those based in two states [152, 143, 160] −susceptible (S) and infected (I)−, the current theory concentrates on two specific situations, the contact process [156, 134, 150, 135, 136, 132] (CP) and the reactive process [145, 137, 141] (RP). A CP stands for a dynamical process that involves an individual stochastic contagion per infected node per unit time, while in the RP there are as many stochastic contagions per unit time as neighbors a node has. This latter process underlies the abstraction of the SIS model [152, 143, 160]. However, in real situations, the number of stochastic contacts per unit time is surely a variable of the problem itself [148]. In this chapter we develop a microscopic model, based on Markov-Chains, to cope with the concurrency problem in the spreading of epidemics.

8.1.1 Contact-based epidemic spreading models

Let us suppose we have a complex network, undirected or directed, made up of N nodes, whose connections are represented by the entries $\{a_{ij}\}$ of an N-by-N adjacency matrix $\mathbf{A}$, where $a_{ij} \in \{0, 1\}$. Unlike standard HMF approaches, our formalism allows the analysis of weighted networks, thus we denote by $\{w_{ij}\}$ the non-negative weights ($w_{ij} \geq 0$) of the connections between nodes, being $w_i = \sum_j w_{ij}$ the total output strength [131] of node i. The above quantities completely define the structure of the underlying graph. The dynamics we consider is a discrete two-state contact-based process, where every node is either in a susceptible (S) or infected (I) state. Each node of the network represents an individual (or a place, a city, an airport, etc.) and each edge is a connection along which the infection spreads. At each time step, an infected node makes a number λ of trials to transmit the disease to its neighbors with probability β per unit time, and then has a probability μ of recovering to the susceptible state. This forms a Markov chain where the probability of a node being infected depends only on the last time step, hence the name Microscopic

Markov-Chain Approach (MMCA). After some transient time, the previous dynamics sets the system into a stationary state in which the average density of infected individuals, ρ, defines the prevalence of the disease.

We are interested in the probability $p_i(t)$ that any given node i is infected at time step t. We denote by r_{ij} the probability that a node i is in contact with a node j, defining a matrix $\mathbf{R}$. These entries represent the probabilities that existing links in the network are used to transmit the infection. If i and j are not connected, then $r_{ij} = 0$. With these definitions, the discrete-time version of the evolution of the probability of infection of any node i reads

$$p_i(t+1) = (1 - q_i(t))(1 - p_i(t)) + (1 - \mu)p_i(t) + \mu(1 - q_i(t))p_i(t)\,, \qquad (8.1)$$

where $q_i(t)$ is the probability of node i not being infected by any neighbor at time t,

$$q_i(t) = \prod_{j=1}^{N} (1 - \beta r_{ji} p_j(t))\,. \qquad (8.2)$$

The first term on the right hand side of eq. (8.1) is the probability that node i is susceptible $(1 - p_i(t))$ and is infected $(1 - q_i(t))$ by at least a neighbor. The second term stands for the probability that node i is infected at time t and does not recover, and finally the last term takes into account the probability that an infected node recovers $(\mu p_i(t))$ but is re-infected by at least a neighbor $(1-q_i(t))$. Within this formulation, we are assuming the most general situation in which recovery and infection occur on the same time scales, allowing then reinfection of individuals during a discrete time window (for instance, one MC step). This formulation generalizes previous approximations where one time step reinfections can not occur.

The formulation so far relies on the assumption that the probabilities of being infected p_i are independent random variables. This hypothesis turns out to be valid in the vast majority of complex networks because the inherent topological disorder makes dynamical correlations not persistent. The dynamical system (8.1, 8.2) corresponds to a family of possible models, parameterized by the explicit form of the contact probabilities r_{ij}. Without loss of generality, it is instructive to think of these probabilities as the transition probabilities of random walkers on the network. The general case is represented by λ_i random walkers leaving node i at each time step:

$$r_{ij} = 1 - \left(1 - \frac{w_{ij}}{w_i}\right)^{\lambda_i}\,. \qquad (8.3)$$

The Contact Process (CP) corresponds to a model dynamics of one contact per unit time, $\lambda_i = 1$, $\forall i$ in eq. (8.3) thus $r_{ij} = w_{ij}/w_i$[1]. In the Reactive Process (RP) all neighbors are contacted, which corresponds, in this description, to set the limit $\lambda_i \to \infty$, $\forall i$ resulting on $r_{ij} = a_{ij}$ regardless of whether the network is weighted or not. Other prescriptions for λ_i conform the spectrum of models that can be obtained using this unified framework. The phase diagram of every model is simply obtained solving the system formed by eq. (8.1) for $i = 1, \ldots, N$ at the stationary state,

$$p_i = (1 - q_i) + (1 - \mu) p_i q_i \,, \tag{8.4}$$

$$q_i = \prod_{j=1}^{N} (1 - \beta r_{ji} p_j) \,. \tag{8.5}$$

This equation has always the trivial solution $p_i = 0$, $\forall i = 1, \ldots, N$. Other non-trivial solutions are reflected as non zero fixed points of eqs. (8.4) and (8.5), and can be easily computed numerically by iteration. The macroscopic order parameter is given by the expected fraction of infected nodes ρ, computed as

$$\rho = \frac{1}{N} \sum_{i=1}^{N} p_i \,. \tag{8.6}$$

8.1.2 Numerical results

To show the validity of the MMCA model here discussed, we have performed MC simulations on different SF networks for RP. In Figure 8.1 the phase diagram of the system obtained by MC simulations is compared with the numerical solution of eqs. (8.4) and (8.5). To model the epidemic dynamics on the described topologies we incorporate a SIS model in which, at each time step, each node can be susceptible or infected. Each simulation starts with a fraction ρ_0 of randomly chosen infected individuals ($\rho_0 = 0.05$ in our simulations), and time is discretized in time-steps. At each time step an infected node i infects with the same probability β all its neighbors and recovers at a rate μ. The simulation runs until a stationary state for the density of susceptible individuals, $\rho(t)$ is reached. The agreement between both curves is matchless. Moreover, the formalism also captures the microscopic dynamics as given by the p_i's, see

[1] Strictly speaking, when $\lambda = 1$, our model is not exactly the standard CP, since in that case reinfections are not considered. However, we will refer to it as a CP since only one neighbor is contacted at each time step and the critical points of both variants are the same.

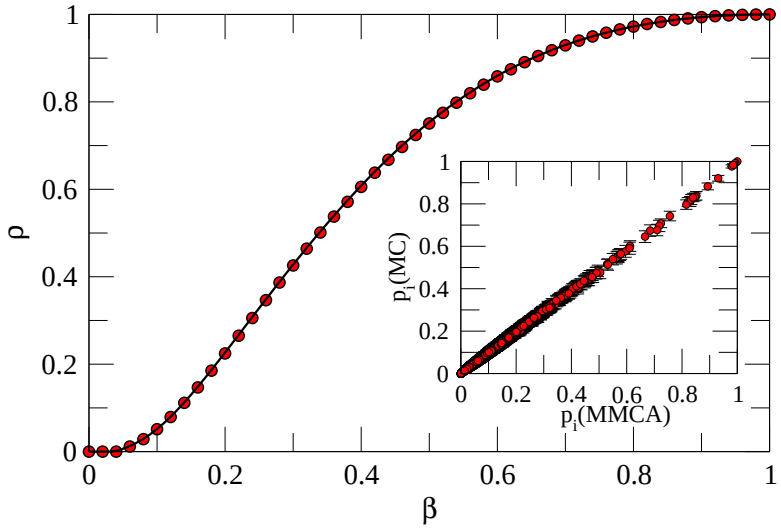

Figure 8.1: Average fraction of infected nodes ρ as a function of the infection rate β for $N = 10^4$. Lines stand for the MMCA solutions (with $\lambda = \infty$) and symbols correspond to MC simulations of the SIS model on top of random scale-free networks with $\gamma = 2.7$ (error bars are smaller than the size of the symbol). In the inset, scatter plot for the probability that a node is infected using results of MC simulations (the y-axis) and the solutions (x-axis) of eqs. (8.4) and (8.5). Both results have been obtained for $\mu = 1$, the inset is for $\beta = 0.1$.

the inset of Fig. 8.1. While the computational cost of the MC simulations is considerably large, the numerical solution of the fix point eqs. (8.4) and (8.5), by iteration, is fast and accurate.

In Fig. 8.2 we analyze our formalism on top of the airports network data set, composed of passenger flights operating in the time period November 1, 2000, to October 31, 2001 compiled by OAG Worldwide (Downers Grove, IL) and analyzed previously by Prof. Amaral's group [149]. It consists of 3618 nodes (airports) and 14142 links, we used the weighted network in our analysis. Airports corresponding to a metropolitan area have been collapsed into one node in the original database. We show the density of infected individuals ρ as a function of β for different values of λ. Both the critical points and the

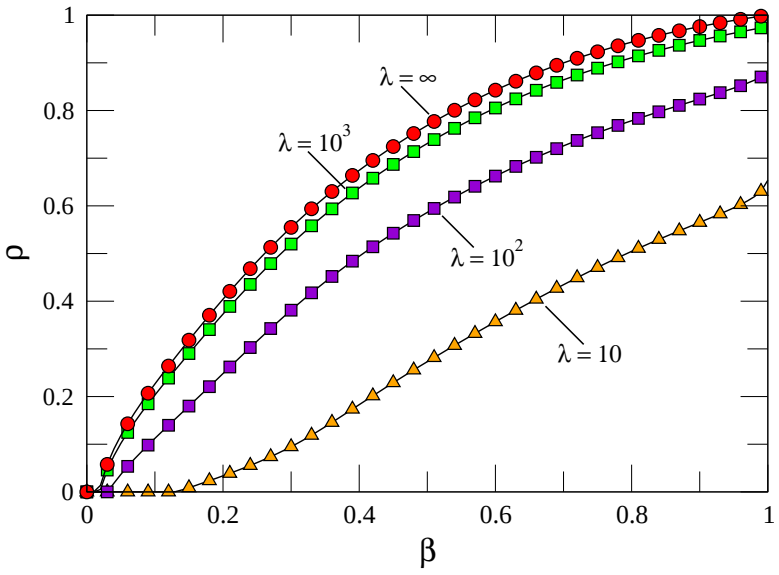

Figure 8.2: Density of infected individuals ρ as a function of β for different values of λ in the air transportation network [149]. The smallest epidemic threshold and largest incidence is obtained for the RP, in which the matrix R corresponds to the adjacency matrix. This implies that the SIS on unweighted networks is a worst case scenario for the epidemic spreading in real weighted networks. ρ is calculated according to eq. (8.6) once the p_i's are obtained, μ is set to 1.

shape of the $\rho - \beta$ phase diagrams greatly change at varying the number of stochastic contacts (λ). We observe a moderate disease prevalence in the case of small values of λ, even for large values of the spreading rate β. In contrast, when the number of trials is of order 10^3 the situation is akin to a RP.

Finally, we compare the results of the formalism for different random scale-free networks satisfying $P(k) \sim k^{-\gamma}$, which have been generated using the configuration model [1, 4] with a fixed size of $N = 10^4$ nodes. Fig. 8.3 shows the phase diagram for $\mu = 1$ and several values of the exponent γ, both below and above $\gamma = 3$. Symbols correspond to MC simulations, whereas dotted lines represent the results obtained using the analytical approximation. As it can

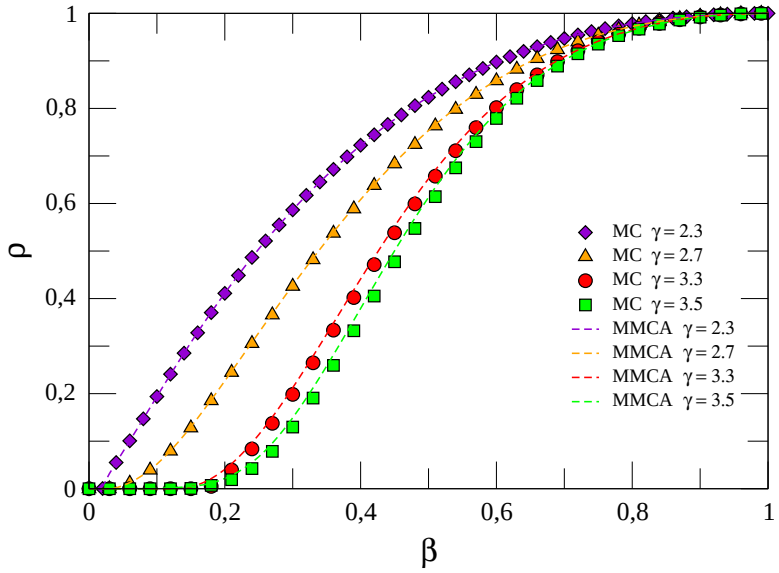

Figure 8.3: Phase diagram for the SIS model ($\lambda = \infty$) in a random scale free network for different γ's. The networks size is $N = 10^4$ nodes and $\mu = 1$. MC results are averages over 10^2 realizations. Dashed lines corresponds to the theoretical prediction and symbols to MC results.

be seen, the agreement between both methods is remarkable, even for values of $\gamma < 2.5$ where structural changes are extremely relevant [169]. On the other hand, one may explore the dependency with the system size while fixing the degree distribution exponent γ. This is what is shown in Fig. 8.4, where we have depicted the phase diagram for networks with $\gamma = 2.7$ for several system sizes ranging from $N = 500$ to $N = 10^5$. Except for $N = 500$, where MC results have a large standard deviation close to the critical point, the agreement is again excellent in the whole range of β values.

8.1.3 Epidemic Threshold

Let us now assume the existence of a critical point β_c for fixed values of μ and λ_i such that $\rho = 0$ if $\beta < \beta_c$ and $\rho > 0$ when $\beta > \beta_c$. The calculation of this

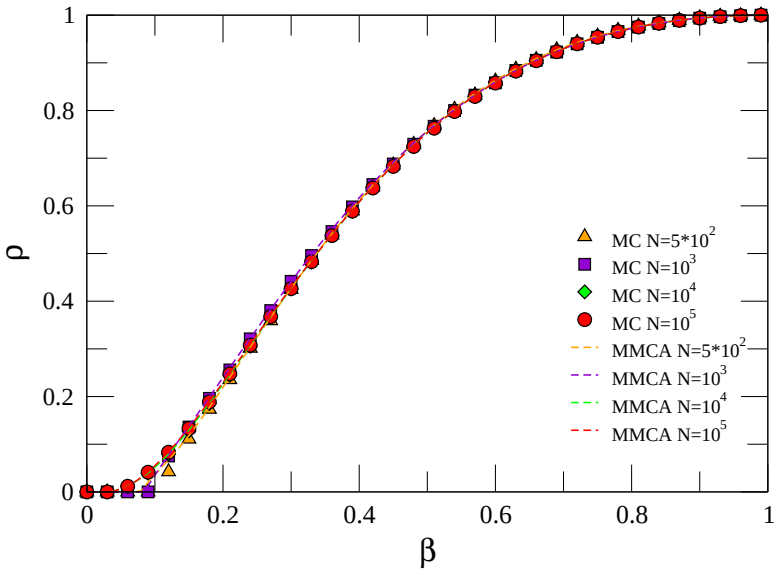

Figure 8.4: Phase diagram for the SIS model ($\lambda = \infty$) in a random scale free network for different system sizes as indicated. The networks have a power-law degree distribution with an exponent $\gamma = 2.7$ and $\mu = 1$. MC results are averages over 10^2 realizations.

critical point is performed by considering that when $\beta \to \beta_c$, the probabilities $p_i \approx \epsilon_i$, where $0 \leq \epsilon_i \ll 1$, and then after substitution in eq. (8.2) one gets

$$q_i \approx 1 - \beta \sum_{j=1}^{N} r_{ji} \epsilon_j \,. \tag{8.7}$$

Inserting eq. (8.7) in eq. (8.4), and neglecting second order terms in ϵ we get

$$\sum_{j=1}^{N} \left(r_{ji} - \frac{\mu}{\beta} \delta_{ji} \right) \epsilon_j = 0 \,, \qquad \forall i = 1, \ldots, N \,, \tag{8.8}$$

where δ_{ij} stands for the Kronecker delta. The system (8.8) has non trivial solutions if and only if μ/β is an eigenvalue of the matrix $\mathbf{R}$. Since we are

looking for the onset of the epidemic, the lowest value of β satisfying (8.8) is

$$\beta_c = \frac{\mu}{\Lambda_{\max}}\,, \tag{8.9}$$

where $\Lambda_{\max}$ is the largest eigenvalue of the matrix $\mathbf{R}$. Equation (8.9) defines the epidemic threshold of the disease spreading process.

With the previous development it is worth analyzing the two limiting cases of CP and RP above. In the first case, one must take into account that the matrix $\mathbf{R}$ is a transition matrix whose maximum eigenvalue is always $\Lambda_{\max} = 1$. Thus the trivial result that the only non-zero solution corresponds to $\beta_c = \mu$. For the RP corresponding to the SIS spreading process usually adopted [163], the classical result for uncorrelated SF networks is recovered because, in this case, the largest eigenvalue [139, 166] is $\Lambda_{\max} = \langle k^2\rangle/\langle k\rangle$.

8.1.4 Mesoscopic equations at the critical point

Once the general framework given by the dynamical system (8.1, 8.2) has been proposed, it is instructive to approximate it using the hypotheses underlying HMF. These hypotheses consist of: i) coarse-graining the system in classes of node by degree, assuming that the dynamical properties within each class are the same, and ii) neglecting fluctuations. To obtain the mesoscopic description we consider the second order approximation of eqs. (8.4) and (8.5), and proceed as in the previous section. Therefore,

$$q_i \approx 1 - \beta \sum_j r_{ji}\epsilon_j + \beta^2 \sum_{j<l} r_{ji} r_{li} \epsilon_j \epsilon_l\,. \tag{8.10}$$

After substitution in (8.4) and reordering terms one gets

$$0 = -\mu\epsilon_i + \beta(1-\epsilon_i)\sum_j r_{ji}\epsilon_j + \mu\beta\epsilon_i \sum_j r_{ji}\epsilon_j - \beta^2 \sum_{j<l} r_{ji} r_{li}\epsilon_j\epsilon_l\,, \tag{8.11}$$

which are the equations governing the dynamics of the contact-based epidemic spreading process at the microscopic level. It is possible to write eq. (8.11) at the commonly used mesoscopic (degree class) level for unweighted, undirected heterogeneous networks. The interactions then take place between classes of nodes. Defining the average density of infected nodes with degree k as $\rho_k = \frac{1}{N_k}\sum_{k_i=k} p_i$, where N_k is the number of nodes with degree k and the sum runs over the set of nodes of degree k, we obtain the generalized HMF equation near criticality.

To simplify the notation, we define the function

$$R_\lambda(x) = 1 - (1-x)^\lambda \,. \tag{8.12}$$

Thus, the values of r_{ji} may be expressed as

- Weighted networks:

$$r_{ji} = R_\lambda\left(\frac{w_{ji}}{w_j}\right) . \tag{8.13}$$

- Unweighted networks:

$$r_{ji} = R_\lambda\left(\frac{a_{ji}}{k_j}\right) = a_{ji} R_\lambda\left(\frac{1}{k_j}\right) = a_{ji} R_\lambda(k_j^{-1}) \,. \tag{8.14}$$

Homogeneous networks

For homogeneous unweighted undirected networks, $\epsilon_i = \epsilon$ and $k_i \approx \langle k\rangle$ for all nodes. Thus, $\rho = \frac{1}{N}\sum_j \epsilon_j = \epsilon$ and

$$0 = -\mu\rho + \beta\rho(1-\rho)\sum_j r_{ji} + \mu\beta\rho^2 \sum_j r_{ji} - \beta^2\rho^2 \sum_{j<l} r_{ji} r_{li} \,. \tag{8.15}$$

The terms involving values of r_{ji} are

$$\begin{aligned} r_{ji} &\approx a_{ji} R_\lambda(\langle k\rangle^{-1}) \,, & (8.16)\\ \sum_j r_{ji} &\approx \langle k\rangle R_\lambda(\langle k\rangle^{-1}) \,, & (8.17)\\ \sum_{j<l} r_{ji} r_{li} &\approx \frac{1}{2}\langle k\rangle(\langle k\rangle - 1) R_\lambda(\langle k\rangle^{-1})^2 \,. & (8.18) \end{aligned}$$

Now, eq. (8.15) becomes

$$\begin{aligned} 0 \;=\; & -\mu\rho + \beta\rho(1-\rho)\langle k\rangle R_\lambda(\langle k\rangle^{-1}) \\ & + \mu\beta\rho^2\langle k\rangle R_\lambda(\langle k\rangle^{-1}) - \beta^2\rho^2\frac{1}{2}\langle k\rangle(\langle k\rangle - 1) R_\lambda(\langle k\rangle^{-1})^2 \,, \end{aligned} \tag{8.19}$$

which may be considered as the MF approximation of our model for homogeneous networks.

If $\lambda = 1$ then $R_1(\langle k\rangle^{-1}) = \frac{1}{\langle k\rangle}$ and eq. (8.19) becomes

$$0 = -\mu\rho + \beta\rho(1-\rho) + \mu\beta\rho^2 - \frac{\langle k\rangle - 1}{2\langle k\rangle}\beta^2\rho^2 \,. \tag{8.20}$$

If $\lambda \to \infty$ then $R_\infty(\langle k\rangle^{-1}) = 1$ and eq. (8.19) reads

$$0 = -\mu\rho + \beta\rho(1-\rho)\langle k\rangle + \mu\beta\rho^2\langle k\rangle - \frac{1}{2}\beta^2\rho^2\langle k\rangle(\langle k\rangle - 1)\,. \tag{8.21}$$

In both cases, the first two terms correspond to the standard CP and RP models (previously reported in the literature) respectively, and the additional terms are second order contributions corresponding to reinfections and multiple infections.

Heterogeneous networks

Now we will concentrate on the class of heterogeneous unweighted undirected networks completely specified by their degree distribution $P(k)$ and by the conditional probability $P(k'|k)$ that a node of degree k is connected to a node of degree k'. Of course, the normalization conditions $\sum_k P(k) = 1$ and $\sum_{k'} P(k'|k) = 1$ must be fulfilled. In this case, the average number of links that goes from a node of degree k to nodes of degree k' is $kP(k'|k)$.

In these heterogeneous networks it is supposed that all nodes of the same degree behave equally, thus $\epsilon_i = \epsilon_j$ if $k_i = k_j$, and the density ρ_k of infected nodes of degree k is given by $\rho_k = \frac{1}{N_k}\sum_{i\in K}\epsilon_i = \epsilon_j\,, \quad \forall j \in K$, where $N_k = P(k)N$ is the expected number of nodes with degree k. Here we have made use of K to denote the set of nodes with degree k. This notation allows to group the sums by the degrees of the nodes. For instance, if the degree of node i is $k_i = k$ then

$$\sum_j a_{ji}\epsilon_j = \sum_{k'}\sum_{j\in K'} a_{ji}\rho_{k'} = \sum_{k'}\rho_{k'}\sum_{j\in K'} a_{ij} = \sum_{k'}\rho_{k'}kP(k'|k) = k\sum_{k'}P(k'|k)\rho_{k'}\,. \tag{8.22}$$

Now, let us find the mean field equation for heterogeneous networks. First we substitute equation 8.14 in equation 8.11

$$\begin{aligned} 0 \;=\; & -\mu\epsilon_i + \beta(1-\epsilon_i)\sum_j a_{ji}R_\lambda(k_j^{-1})\epsilon_j \\ & + \mu\beta\epsilon_i\sum_j a_{ji}R_\lambda(k_j^{-1})\epsilon_j - \beta^2\sum_{j<l} a_{ji}a_{li}R_\lambda(k_j^{-1})R_\lambda(k_l^{-1})\epsilon_j\epsilon_l \end{aligned} \tag{8.23}$$

It is convenient to analize separately the summatory terms in equation 8.23,

supposing node i has degree k:

$$\begin{aligned}\sum_j a_{ji} R_\lambda(k_j^{-1})\epsilon_j &= \sum_{k'}\sum_{j\in K'} a_{ji} R_\lambda(k'^{-1})\rho_{k'} \\ &= \sum_{k'} R_\lambda(k'^{-1})\rho_{k'} \sum_{j\in K'} a_{ij} \\ &= k\sum_{k'} P(k'|k) R_\lambda(k'^{-1})\rho_{k'}\,, \end{aligned} \tag{8.24}$$

$$\begin{aligned}&\sum_{j<l} a_{ji}a_{li} R_\lambda(k_j^{-1}) R_\lambda(k_l^{-1})\epsilon_j\epsilon_l \\ &= \frac{1}{2}\sum_j\sum_l a_{ji}a_{li} R_\lambda(k_j^{-1}) R_\lambda(k_l^{-1})\epsilon_j\epsilon_l - \frac{1}{2}\sum_j a_{ji}^2 R_\lambda(k_j^{-1})^2\epsilon_j^2 \\ &= \frac{1}{2}\sum_{k'}\sum_{k''}\sum_{j\in K'}\sum_{l\in K''} a_{ji}a_{li} R_\lambda(k'^{-1}) R_\lambda(k''^{-1})\rho_{k'}\rho_{k''} - \frac{1}{2}\sum_{k'}\sum_{j\in K'} a_{ji}^2 R_\lambda(k'^{-1})^2\rho_{k'}^2 \\ &= \frac{1}{2}\sum_{k'}\sum_{k''} R_\lambda(k'^{-1}) R_\lambda(k''^{-1})\rho_{k'}\rho_{k''}\sum_{j\in K'} a_{ij}\sum_{l\in K''} a_{il} - \frac{1}{2}\sum_{k'} R_\lambda(k'^{-1})^2\rho_{k'}^2\sum_{j\in K'} a_{ij}^2 \\ &= \frac{1}{2}k^2\sum_{k'}\sum_{k''} R_\lambda(k'^{-1}) R_\lambda(k''^{-1}) P(k'|k)P(k''|k)\rho_{k'}\rho_{k''} \\ &\quad - \frac{1}{2}k\sum_{k'} R_\lambda(k'^{-1})^2 P(k'|k)\rho_{k'}^2\,. \end{aligned} \tag{8.25}$$

Substitution in equation 8.23 leads to the generalized HMF equation

$$\begin{aligned} 0 &= -\mu\rho_k + \beta k(1-\rho_k)\sum_{k'} P(k'|k) R_\lambda(k'^{-1})\rho_{k'} \\ &\quad + \mu\beta k\rho_k \sum_{k'} P(k'|k) R_\lambda(k'^{-1})\rho_{k'} \\ &\quad - \frac{1}{2}\beta^2 k^2 \sum_{k'}\sum_{k''} R_\lambda(k'^{-1}) R_\lambda(k''^{-1}) P(k'|k)P(k''|k)\rho_{k'}\rho_{k''} \\ &\quad + \frac{1}{2}\beta^2 k\sum_{k'} R_\lambda(k'^{-1})^2 P(k'|k)\rho_{k'}^2\,. \end{aligned} \tag{8.26}$$

If $\lambda = 1$ then $R_1(k^{-1}) = \frac{1}{k}$ and eq. (8.26) becomes

$$\begin{aligned} 0 = &-\mu\rho_k + \beta k(1-\rho_k)\sum_{k'}\frac{1}{k'}P(k'|k)\rho_{k'} \\ &+ \mu\beta k\rho_k\sum_{k'}\frac{1}{k'}P(k'|k)\rho_{k'} + \frac{1}{2}\beta^2 k\sum_{k'}\frac{1}{k'^2}P(k'|k)\rho_{k'}^2 \\ &- \frac{1}{2}\beta^2k^2\left(\sum_{k'}\frac{1}{k'}P(k'|k)\rho_{k'}\right)^2 . \end{aligned} \tag{8.27}$$

If $\lambda \to \infty$ then $R_\infty(k^{-1}) = 1$ and eq. (8.26) reads

$$\begin{aligned} 0 = &-\mu\rho_k + \beta k(1-\rho_k)\sum_{k'}P(k'|k)\rho_{k'} \\ &+ \mu\beta k\rho_k\sum_{k'}P(k'|k)\rho_{k'} + \frac{1}{2}\beta^2 k\sum_{k'}P(k'|k)\rho_{k'}^2 \\ &- \frac{1}{2}\beta^2k^2\left(\sum_{k'}P(k'|k)\rho_{k'}\right)^2 . \end{aligned} \tag{8.28}$$

Again, the first two terms in both cases correspond to the standard CP and RP HMF equations respectively, and the additional terms are second order contributions corresponding to reinfections and multiple infections.

8.2 Conclusions

In this chapter we have presented a novel framework to study disease spreading in networks. By defining a set of discrete-time equations for the probability of individual nodes to be infected, we construct a dynamical system that generalizes from an individual contact process to the classical case in which all connections are concurrently used, for any complex topology. The whole phase diagram of the system can be found solving the equations at the stationary state. The numerical solution of the analytic equations overcomes the computational cost of MC simulations. Moreover, the formalism allows to gain insight on the behavior of the critical epidemic threshold for different values of the probability of contacting a fraction λ of neighbors per time step. The MMCA model deals with infections driven by direct contacts between nodes, but not with traffic situations where nodes transmit the epidemics by flow communication with others [158]. In this latter case, the routing protocol of traffic

between nodes is absolutely relevant and can change the critical point of the epidemic spreading. In the next chapters, we develop a framework in the scope of MF theories to cope with the problem of assessing the impact of epidemics when the routing of traffic is considered.

Chapter 9

Traffic-driven epidemic spreading in complex networks

In this chapter we investigate the outcome of an epidemic spreading process driven by transport instead of diffusion. To this end, we analyzed a paradigmatic abstraction of epidemic contagion, the SIS model, which assumes that contagion occurs through the eventual contact or transmission between connected partners that are using their connections at the time of propagation. This is achieved by considering a quantized interaction at each time step. Mathematically, we set up the model in a flow scenario where contagion is carried by interaction packets traveling across the network. We consider two possible scenarios that encompass most of real traffic situations: i) unbounded delivery rate, and ii) bounded delivery rate, of packets per unit time. We derive the equation governing the critical threshold for epidemic spreading in SF networks, which embeds, as a particular case, previous theoretical findings. For unbounded delivery rate, it is shown that the epidemic threshold decreases in *finite* SF networks when traffic flow increases. In the bounded case, nodes accumulate packets at their queues when traffic flow overcomes the maximal delivery rate, i.e. when congestion arises. From this moment on, the results show that both the epidemic threshold and the infection prevalence are bounded due to congestion.

9.1 The model

In the first place, two different types of SF networks are generated. On one hand, we build random uncorrelated SF networks using the configuration model [1, 4]. On the other hand, small-world, SF and highly clustered networks – all properties found in many real-world networks [1, 4] such as the Internet – are also generated using a class of recently developed network models [168, 88], in which nearby nodes in a hidden metric space are connected. This metric space can represent social, geographical or any other relevant distance between the nodes of the simulated networks. Specifically, in the model currently at study, nodes are uniformly distributed in a one-dimensional circle by assigning them a random polar angle θ distributed uniformly in the interval $[0, 2\pi)$ and assigned an expected degree k. The expected degrees of the nodes are then drawn from some distribution $x(k)$ and the network is completed by connecting two nodes with hidden coordinates (θ, k) and (θ', k') with probability $r(\theta, k, \theta', k') = \left(1 + \frac{d(\theta,\theta')}{\eta' k k'}\right)^{-\alpha}$, where $\eta' = (\alpha-1)/2\langle k\rangle$, $d(\theta, \theta')$ is the geodesic distance between the two nodes on the circle, and $\langle k\rangle$ is the average degree. Finally, choosing $x(k) = (\gamma-1)k_0^{\gamma-1}k^{-\gamma}$, $k > k_0 \equiv (\gamma-2)\langle k\rangle/(\gamma-1)$ generates random networks with a power law distribution with exponent $\gamma > 2$. In most of the simulations, $\gamma = 2.7$ $\langle k\rangle = 3$ and $\alpha = 2$ are fixed.

Once the networks are built up, the traffic process is implemented in the following way. At each time step, $p = \Lambda N$ new packets are created with randomly chosen origins and destinations. For the sake of simplicity, packets are considered non-interacting so that no queues are used. The routing of information is modeled through even a shortest path delivery strategy or a greedy algorithm [88, 87]. In the latter, the second class of SF networks is used and a node i forwards a packet to node j in its neighborhood, which is the closest node (in the hidden metric space) to the final packet destination. Results are insensitive to the two routing protocols implemented.

To model the spreading dynamics we have implemented the aforementioned Susceptible-Infected-Susceptible model, in which each node can be in two possible states: healthy (S) or infected (I). Starting from an initial fraction of infected individuals $\rho_0 = I_0/N$, the infection spreads in the system as the nodes interact. A susceptible node has a probability β of becoming infected every time it interact with an infected neighbors. We also assume that infected nodes are recovered at a rate μ, which we fix to 1 for most of the simulations. After a transient time, we compute the average density of infected individuals, ρ, which is the prevalence of disease in the system. To account for link

concurrency, we consider that two nodes do not interact at all times t, but only when they exchange at least a packet. This situation is reminiscent of disease transmission on air transportation networks; if an infected individual did not travel between two cities, then regardless of whether or not those cities are connected by a direct flight, the epidemic will not spread from one place to the other. In this way, although a node can potentially interact with as many contacts as it has and as many times as packets it exchanges with its neighbors, the effective interactions are driven by a second dynamics (traffic). The more packets travel through a link, the more likely the disease will spread through it. On the other hand, once an interaction is at work, the epidemics spreads from infected to susceptible nodes with probability β. For example, if at time t node i is infected and a packet is travelling from node i to one of its neighbors node j, then at the next time step, node j will be infected with probability β. Therefore, susceptible and infected states are associated with the nodes, whereas the transport of packets is the mechanism responsible for the propagation of the disease at each time step.

9.1.1 Unbounded delivery rate

We firstly concentrate on an unbounded delivery rate scenario, in which every node can handle as much packets it receives. In this situation, congestion can not arise in the system. Fig. 9.1 shows the results for the stationary density of infected nodes ρ as a function of β and the traffic generation rate Λ for SF networks.

In this case the traffic level determines the value of both the epidemic incidence and the critical thresholds and it's important to notice the emergence of an epidemic threshold under low traffic conditions. This implies that for a fixed value of Λ, the epidemic dies out if the spreading rate is below a certain critical value $\beta_c(\Lambda)$. More intense packet flows yield lower epidemic thresholds. The reason for the dependence of the critical spreading rates on Λ is rooted in the effective topological paths induced by the flow of packets through the network. At low values of Λ, there are only a few packets traveling throughout the system, so the epidemic simply dies out because many nodes do not participate in the interaction via packets exchanges. As Λ grows, more paths appear between communicating nodes, thus spreading the infection to a larger portion of the network. Therefore, in traffic-driven epidemic processes the infection is constrained to propagate only through links that transmit a packet, and thus the number of attempts to transmit the infection depends on the flow conditions at a local level, namely, on the number of active communication channels

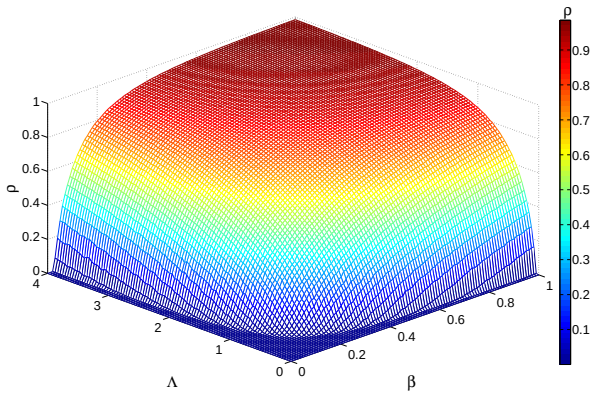

Figure 9.1: Dependence of epidemic incidence on traffic conditions for unbounded delivery rate. The density of infected nodes, ρ, is shown as a function of the spreading rate β and the intensity of flow Λ in SF networks. Flow conditions (controlled by Λ) determine both the prevalence level and the values of the epidemic thresholds. Increasing the number of packets traveling through the system has a malicious effect: the epidemic threshold decreases as the flow increases. Each curve is an average of 10^2 simulations starting from an initial density of infected nodes $\rho_0 = 0.05$. The network is made up of 10^3 nodes using the model in [88], results correspond to the greedy routing scheme. The remaining parameters are $\alpha = 2$, $\gamma = 2.6$ and $\langle k \rangle = 3$.

at each time step. As a consequence, the effective network that spreads the infection is no longer equivalent to the complete underlying topology. Instead, it is a map of the dynamical process associated with packet traffic flow. The conclusion is that the disease propagation process has two dynamical components: one intrinsic to the disease itself (β) and the other to the underlying traffic dynamics (the flow). To theorize about these effects we next formulate the analytical expression for the dependence of the epidemic threshold on the amount of traffic injected into the system, following a mean-field approach akin to the conventional analysis of the reaction driven case. Mathematically, the fraction of paths traversing a node given a certain routing protocol [191], the so-called algorithmic betweenness, b^k_{alg}, defines the flow pathways. Let us consider the evolution of the relative density, $\rho_k(t)$, of infected nodes with degree k. Following the heterogeneous mean-field approximation [163], the dynamical

rate equations for the SIS model are

$$\partial_t \rho_k(t) = -\mu \rho_k(t) + \beta \Lambda b_{alg}^k N \left[1 - \rho_k(t)\right] \Theta(t). \tag{9.1}$$

The first term in Eq. (9.1) is the recovery rate of infected individuals (we set henceforth $\mu = 1$). The second term takes into account the probability that a node with k links belongs to the susceptible class, $[1 - \rho_k(t)]$, and gets the infection via packets traveling from infected nodes. The latter process is proportional to the spreading probability β, the probability $\Theta(t)$ that a packet travels through a link pointing to an infected node and the number of *packets* received by a node of degree k. This, in turns, is proportional to the total number of packets in the system, $\sim \Lambda N$, and the algorithmic betweenness of the node, b_{alg}^k. Note that the difference with the standard epidemic spreading model is given by these factors, as now the number of contacts per unit time of a node is not proportional to its connectivity but to the number of packets that travel through it. Finally, $\Theta(t)$ takes the form

$$\Theta(t) = \frac{\sum_k b_{alg}^k P(k) \rho_k(t)}{\sum_k b_{alg}^k P(k)}. \tag{9.2}$$

Eq. (9.1) has been obtained assuming: (i) that the network is uncorrelated $P(k'|k) = k' P(k')/\langle k \rangle$, and (ii) that the algorithmic flow between the classes of nodes of degree k and k' factorizes $b_{alg}^{kk'} \sim b_{alg}^k b_{alg}^{k'}$. Although no uncorrelated networks exist, this approximation allows us to identify the governing parameters of the proposed dynamics. The second approximation is an upper bound to the actual value of the $b_{alg}^{kk'}$, whose mathematical expression is, in general, unknown. The validity of the theory even with these approximations is notable as confirmed by the numerical simulations.

By imposing stationarity $[\partial_t \rho_k(t) = 0]$, Eq. (9.1) yields

$$\rho_k = \frac{\beta \Lambda b_{alg}^k N \Theta}{1 + \beta \Lambda b_{alg}^k N \Theta}, \tag{9.3}$$

from which a self-consistent equation for Θ is obtained as

$$\Theta = \frac{1}{\sum_k b_{alg}^k P(k)} \sum_k \frac{(b_{alg}^k)^2 P(k) \beta \Lambda N \Theta}{1 + \beta \Lambda b_{alg}^k N \Theta}. \tag{9.4}$$

The value $\Theta = 0$ is always a solution. In order to have a non-zero solution, the condition

$$\frac{1}{\sum_k b_{alg}^k P(k)} \frac{d}{d\Theta} \left(\sum_k \frac{(b_{alg}^k)^2 P(k) \beta \Lambda N \Theta}{1 + \beta \Lambda b_{alg}^k N \Theta} \right) \Bigg|_{\Theta=0} > 1 \tag{9.5}$$

must be fulfilled, from which the epidemic threshold is obtained as

$$\beta_c = \frac{\langle b_{alg} \rangle}{\langle b_{alg}^2 \rangle} \frac{1}{\Lambda N}, \tag{9.6}$$

below which the epidemic dies out, and above which there is an endemic state. In Fig. 9.2 a comparison between the theoretical prediction and numerical observations is presented. Here, we have explicitly calculated the algorithmic betweenness for the greedy routing as it only coincides with the topological betweenness for shortest paths routing. The obtained curve separates two regions: an absorbing phase in which the epidemic disappears, and an active phase where the infection is endemic.

Equation (9.6) is notably simple but has profound implications: the epidemic threshold decreases with traffic and eventually vanishes in the limit of very large traffic flow in finite systems, in contrast to the expected result of a finite-size reminiscent threshold in the classical reactive–diffusive framework. Admittedly, this is a new feature with respect to previous results on epidemic spreading in SF networks. It is rooted in the increase of the effective epidemic spreading rate due to the flow of packets. This is a genuine effect of traffic-driven epidemic processes and generalizes the hypothesis put forward in the framework of a reaction-diffusion process [141] on SF networks. It implies that an epidemic will pervade the (finite) network whatever the spreading rate is if the load on it is high enough. Moreover, Eq. (9.6) reveals a new dependence. The critical threshold depends on the topological features of the graph, but at variance with the standard case, through the first two moments of the algorithmic betweenness distribution. As noted above, the algorithmic betweenness of a node is given by the number of packets traversing that node given a routing protocol. In other words, it has two components: a topological one which is given by the degree of the node and a dynamical component defined by the routing protocol.

Within our formulation, the classical result [163]

$$\beta_c = \frac{\langle k \rangle}{\langle k^2 \rangle}, \tag{9.7}$$

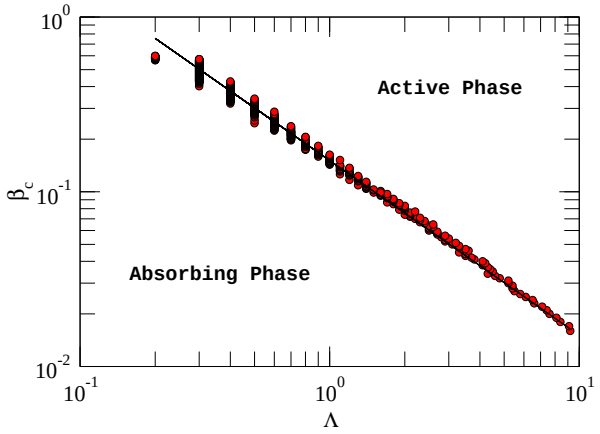

Figure 9.2: Comparison between numerical and theoretical critical points. Log-log plot of the critical thresholds, β_c, as a function of the rate at which packets are injected into the system, Λ. Two regions are differentiated: an active and an absorbing phase as indicated. The solid line corresponds to Eq. (9.6) with $\frac{\langle b_{alg}\rangle}{\langle b_{alg}^2\rangle}\frac{1}{N} = 0.154$. The agreement is remarkable even though Eq. (9.6) is derived using a MF approach. The underlying network, infection spreading mechanism and routing protocol are the same as in Fig. 9.1. Each curve is an average of 10^2 simulations. Remaining parameters are the same as in Fig. 9.1.

can be obtained for a particular protocol and traffic conditions, although we note that the microscopic dynamics of our model is different from the classical SIS. To see this, assume a random protocol. If packets of information are represented as w random walkers traveling in a network with average degree $\langle k\rangle$, then under the assumption that the packets are not interacting, it follows that the average number of walkers at a node i in the stationary regime (the algorithmic betweenness) is given by [198, 157] $b_{alg}^i = \frac{k_i}{N\langle k\rangle}w$. The effective critical value is then $(\beta\Lambda)_c = \langle k\rangle^2/(\langle k^2\rangle w)$, that recovers, when $\omega = \langle k\rangle$, the result in Eq. (9.7).

Results are robust for other network models and different routing algorithms. We have also made numerical simulations of the traffic-driven epidemic process on top of Barabási-Albert and random SF networks implementing a shortest paths delivery scheme. In this case, packets are diverted following the shortest path (in the actual topological space) from the packets' origins to their

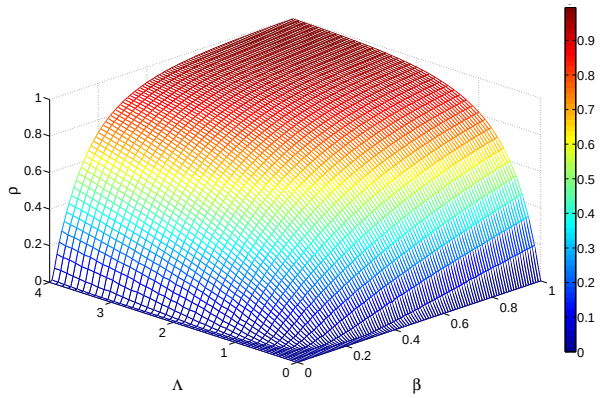

Figure 9.3: Density of infected nodes, ρ, as a function of traffic flow (determined by Λ) and the epidemic spreading rate β for random scale-free networks and a shortest paths routing scheme for packets delivery. Each point is the result of 10^2 averages over different networks and initial conditions. The exponent of the degree distribution of the network is $\gamma = 2.7$.

destinations. The rest of model parameters and rules for epidemic spreading remain the same. Figs. 9.3 and 9.4 show the results obtained for random SF networks generated via the configuration model and the Barabási-Albert model respectively. As can be seen, the phenomenology is the same for both types of networks: the epidemic threshold depends on the amount of traffic in the network such that the higher the flow is, the smaller the epidemic threshold separating the absorbing and active phases. On the other hand, for processes in which the delivery of packets follows a shortest path algorithm, Eq. (9.6) looks like

$$\beta_c = \frac{\langle b_{top} \rangle}{\langle b_{top}^2 \rangle} \frac{1}{\Lambda N}, \tag{9.8}$$

where b_{top} is the topological betweenness. To further confirm our findings on a realistic topology we run the model on top of the Air Transportation Network (ATN) [149]. The network composed by the direct flies between more the 3000 airports in the world, in which each node represents an airport and the links represents the direct connection between them. Although in the ATN links have weights accounting for the annual number of passengers voyaging on each connection, we considered the network as un-weighted and the shortest-

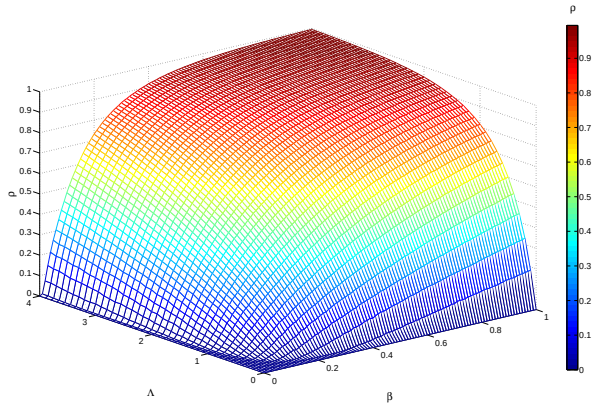

Figure 9.4: Density of infected nodes, ρ, as a function of traffic flow (determined by Λ) and the epidemic spreading rate β for BA scale-free networks and a shortest paths routing scheme for packets delivery. Each point is the result of 10^2 averages over different networks and initial conditions.

path routing protocol. Also in this case the results are confirmed as shown in Fig. 9.5. Fig. 9.6 also shows the agreement between the analytical prediction and the numerical simulations.

9.1.2 Bounded delivery rate

Equation (9.8) allows us to investigate also the equivalent scenario in the presence of congestion. Let us consider the same traffic process above but with nodes having queues that can store as many packets as needed but can deliver, on average, only a finite number of them at each time step. It is known that there is a critical value of Λ above which the system starts to congest [191]

$$\Lambda_c = \frac{(N-1)}{b^*_{alg}}. \tag{9.9}$$

Equation (9.9) gives the traffic threshold that defines the onset of congestion, which is governed by the node with maximum algorithmic betweenness b^*_{alg}. Substituting (9.9) in (9.6) we obtain a critical threshold for an epidemic spreading process bounded by congestion. Increasing the traffic above Λ_c will

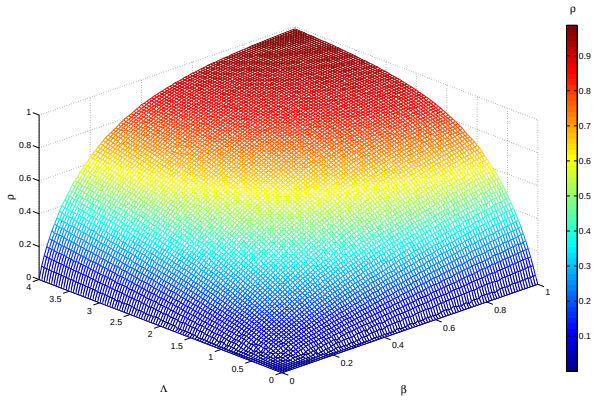

Figure 9.5: Density of infected nodes, ρ, as a function of traffic flow (determined by Λ) and the epidemic spreading rate β for the ATN (considered as unweighted) and a shortest paths routing scheme for packets delivery. Each point is the result of 10^2 averages over different networks and initial conditions.

gradually congest all the nodes in the network up to a limit in which the traffic is stationary and the lengths of queues grow without limit.

To illustrate this point, let us assume that the capacities for processing and delivering information are heterogeneously distributed [171, 194, 167] so that the larger the number of paths traversing a node, the larger its capability to deliver the packets. Specifically, each node i of the network delivers at each time step a maximum of $\lceil c_i = 1 + k_i^{\eta} \rceil$ packets, where η is a parameter of the model. In this case, the critical value of Λ in Eq.(9.9) is multiplied by the maximum delivery capacity [171]. Moreover, without loss of generality, we will explore the behavior of the model in random SF networks where the routing is implemented by shortest paths $b_{alg} = b_{top} \sim k^{\nu}$, being ν usually between 1.1 and 1.3 [165]. The previous assumption for the delivery capability thus allows to explore as a function of η the situations in which the delivery rate is smaller or larger than the arrival rate (defined by the algorithmic betweenness). Phenomenologically, these two scenarios correspond to the cases in which the traffic is in a free flow regime (if $\eta > \nu$) or when the network will congest (if $\eta < \nu$). We also note that the adopted approach is equivalent to assume a finite length for the queues at the nodes.

Fig. 9.7 shows the fraction of active packets on the network, as a function

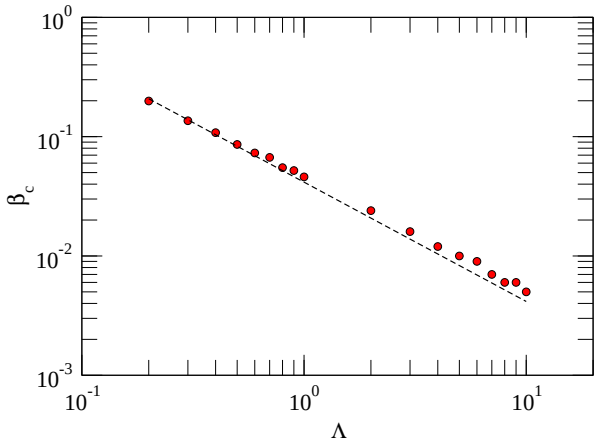

Figure 9.6: Comparison between numerical and theoretical critical points in the ATN. Log-log plot of the critical thresholds, β_c, as a function of the rate at which packets are injected into the system, Λ for the ATN. The dashed line corresponds to Eq. (9.6) with $\frac{\langle b_{alg}\rangle}{\langle b^2_{alg}\rangle}\frac{1}{N} = 0.041$. Despite existing degree correlations in the network, the agreement is remarkable. Each point is an average of 10^2 simulations.

of the spreading rate β and the rate at which packets are generated Λ for two different values of η using a shortest path delivery scheme on top of random SF networks. For $\eta = 0.8$, the epidemic incidence is significantly small for all values of the parameters Λ and β as compared with the results obtained when the rate of packets delivery is unbounded. On the contrary, when $\eta = 1.7$ the phase diagram is qualitatively the same as for the unbounded case, including the result that the epidemic incidence vanishes when Λ is large enough. A closer look at the dynamical evolution unveils an interesting, previously unreported, feature − when the rate at which packets are delivered is smaller than the rate at which they arrive, the average value of infected nodes saturates beyond a certain value of the traffic flow rate Λ. This effect is due to the emergence of traffic congestion. When the flow of packets into the system is such that nodes are not able to deliver at least as many packets as they receive, their queues start growing and packets pile up. This in turns implies that the spreading of the disease becomes less efficient, or in other words, the spreading process slows down. The consequence is that no matter whether more packets are

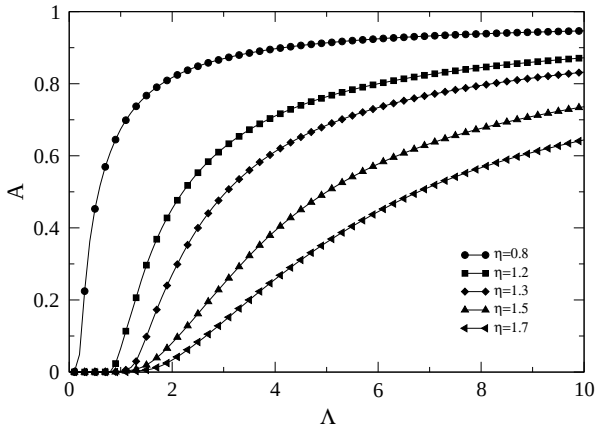

Figure 9.7: Fraction of active packets as a function of the traffic flow with bounded delivery rate. It represents the fraction of active packets A: packets still traveling in the network over the total amount of generated packets in a time period τ, as function of the traffic injected in the system Λ for different values of the delivery capacity η. The underlying network and the routing protocol are the same as in Fig. 9.3.

injected into the system, the average level of packets able to move from nodes to nodes throughout the network is roughly constant and so is the average level of infected individuals.

Fig. 9.8 illustrates the phenomenological picture described above. It shows the epidemic incidence ρ for a fixed value of $\beta = 0.15$ as a function of Λ for different values of η. The figure clearly evidences that congestion is the ultimate reason of the behavior described above. Therefore, the conclusion is that in systems where a traffic process with finite delivery capacity is coupled to the spreading of the disease the *epidemic incidence is bounded*. This is good news as most of the spreading processes in real-world networks involves different traffic flow conditions. Further evidence of this phenomenology is given in Fig. 9.9, where we have depicted the epidemic threshold as a function of Λ for two different values of η, less and greater than ν. When $\eta < \nu$ congestion arises, and the contrary holds for $\eta > \nu$ where the diagram is equivalent to that of unbounded traffic. The onset of congestion determines the value of β above which congestion starts. It is clearly visualized as the point beyond

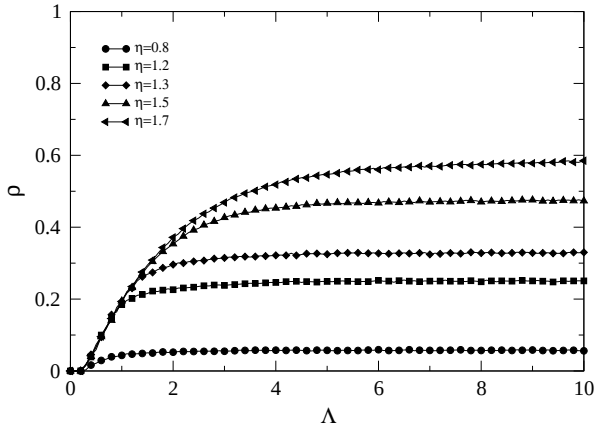

Figure 9.8: Epidemic incidence in traffic-driven epidemic processes with bounded delivery rate. The figure represents the average fraction of infected nodes ρ as a function of Λ for different delivery rates at fixed $\beta = 0.15$. When congestion arises, the curves depart from each other and the epidemic incidence saturates soon afterwards.

which the power law dependence in Eq. (9.6) breaks down. The plateau of β_c corresponds to the stationary situation of global congestion. A comparison for different values of η in the bounded delivery rate model is presented in Fig. 9.10.

9.2 Conclusions

In summary, we argued both analytically and numerically the conditions for the emergence of an epidemic outbreak in scale-free networks when disease contagion is driven by traffic or interaction flow. The study provides a more general theory of spreading processes in complex heterogeneous networks that includes the previous results as a particular case of diffusive spreading. Moreover, we have shown that the situation in which the epidemic threshold vanishes in finite scale-free networks is also plausible, thus providing an explanation to the long-standing question of why some viruses prevail in the system with a low incidence. Moreover, the new approach presented here provides a novel framework to address related problems. For instance, in the context of air-

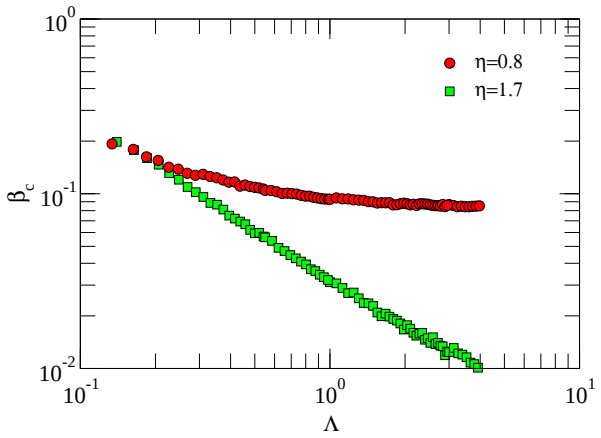

Figure 9.9: Epidemic thresholds as a function of Λ for two values of η. The onset of congestion (the arrow in the plot) marks the point, $\Lambda_c \approx 0.150$, at which the curve for $\eta = 0.8$ departs from Eq. (9.6), i.e., when the power law dependence breaks down. Soon afterwards congestion extends to the whole network leading to a bounded (from below) epidemic threshold. On the contrary, when the delivery rate is large enough (as in the case of $\eta = 1.7$), Eq. (9.6) holds for all values of Λ, thus resembling the unbounded delivery rate case.

transportation networks [140], a similar mechanism to the one reported here could explain the observed differences in the impact of a disease during a year [147]. One might even expect that, due to seasonal fluctuations in flows, the same disease could not provoke a system-wide outbreak if the flow were not high enough during the initial states of the disease contagion. Incorporating the non-diffusive character of the spreading process into current models has profound consequences for the way the system functions. Also the theory could help designing new immunization algorithms or robust protocols; one in particular being quarantining highly sensitive traffic nodes. On more general grounds, our conclusions point to the need of properly dealing with link concurrency. Further exploring this challenge will have far-reaching consequences for the study of dynamical processes on networks, and especially the relationship between structure and dynamics for networked systems. Ultimately, this paves the way towards a more complete theoretical framework of complex networks.

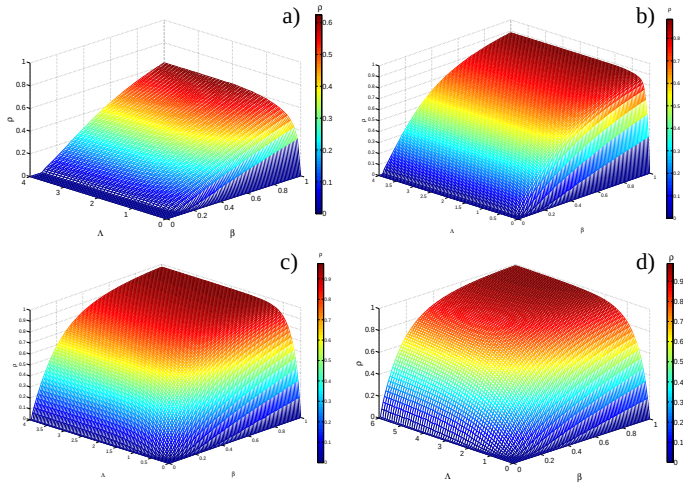

Figure 9.10: Comparison between different delivery capacity in the bounded delivery rate model. The plot represents density of infected nodes, ρ, as a function of traffic flow Λ and the epidemic spreading rate β for random scale-free networks and a shortest paths routing scheme with different values of the delivery capacity η: panel a) $\eta = 0.8$, b) $\eta = 1.0$, c) $\eta = 1.5$ and d) $\eta = 1.7$, for the random SF network and the shortest path delivery scheme.

Chapter 10

Impact of human behavior on epidemic outbreaks.

Recent disease outbreaks have highlighted the need to develop even more complex and sophisticated epidemiological models adapted to the way contagion spreads in our modern society [140, 153]. Just a few decades ago, some diseases would have produced only local outbreaks. However, the changes in human habits and the development of fast and world-wide transportation systems [140, 153, 149, 147, 211] have led to a situation in which a disease can spread in a few days producing a global threat to public health systems. Certainly, current theoretical and computational models have gained in complexity and have incorporated a number of important ingredients that make them more accurate in modeling pandemics [163, 155, 159, 14, 130, 144, 141, 142, 146, 158, 202]. For instance, we have been able to adapt models based on the assumption of random mixing of individuals within the population to include the interconnection patterns of the networks through which the disease disseminates in the system [1, 4]. The same happens with Agent Based computational Models (ABMs), which are increasingly more complex and able to track millions of virtual individuals, capturing their contacts and mobility patterns [200, 201, 206]. The expectancy is that these latter models would allow real time monitoring and prediction of contagion dynamics.

Latest diseases such as the severe acute respiratory syndrome (SARS) and the H1N1 influenza have shown that dynamical and topological properties of transportation networks are key factors to understand today's infections. For example, the presence of hubs dominate the uprise of epidemic processes in

scale-free (SF) networks [163, 155, 159, 14, 130, 146, 158]. Recently [158], it has been proposed that epidemic behavior also depends on the characteristics of transport processes that take place on top of the topology, in the form of traffic or interaction flow. In particular, the epidemic threshold turns out to be inversely proportional to the traffic flow in the system [158], such that the more intense the flow of individuals the lower the epidemic threshold. The latter mechanism suggests that limitations on the propagation of new infections through changes in mobility patterns might eventually shape the course of an epidemic.

Perhaps the most important of our current challenges in epidemic modeling is the inclusion of self-initiated (i.e., not due to external measures) human behavioral reactions in front of a disease [206, 207]. These kind of human responses to the presence of a disease include the increase of personal hygiene, voluntary vaccination and quarantine, changes in the frequency of contacts as well as irrational behaviors (fear, phobias, etc), all of which feed back to impact epidemic dynamics [207]. Although the latter reactions are quite common and well-documented, no systematic investigation has been carried out. Admittedly, the inclusion of human behavioral changes into theoretical and computational models is infrequent.

In this chapter we consider a metapopulation model that incorporates self-initiated changes in human behavior in response to an epidemic outbreak and study how these reactions influence the spread of infectious diseases. The model is general enough as to include a number of different responses. We find that whatever the response is, the global invasion threshold remains the same. However, the number of subpopulations affected by the outbreak does depend on the population behavior. If individuals retrain themselves from traveling, a lower impact is obtained. On the contrary, when a selfish behavior is adopted by bypassing hot epidemic spots, the outbreak reaches a larger fraction of the total population. Our results thus point out that predictions of disease progression should be made with care if information about behavioral adaptation is not incorporated (in particular, the worst scenario hypothesis should be revised).

10.1 The model

Our model belongs to the class of metapopulation systems widely used in epidemiology [209]. This modeling approach provides the theoretical and computational framework for describing the spread of diseases in spatially structured interacting subpopulations that are connected through a network. Each sub-

population consists of a number of individuals that are divided into several classes according to their dynamical state with respect to the modeled disease − for instance, susceptible, infected, removed, etc. The internal compartmental dynamics models the contagion dynamics by considering that people in the same subpopulation get in contact and may change their state according to their interactions and the infection dynamics. Finally, subpopulations also interact and exchange individuals due to mobility from one subpopulation to another. Figure 10.1 shows an schematic representation of the metapopulation system.

Metapopulation epidemic models with heterogeneous coupling patterns have been recently considered [142, 212]. Such framework includes the heterogeneities found in demographic and mobility patterns, hence, a subpopulation is further classified according to the number of subpopulations it is in contact with, i.e., its degree k. The results obtained for this kind of models show that local outbreaks depend on disease parameters. Furthermore, the global invasion threshold, that marks the point beyond which a local outbreak reaches other subpopulations and spreads throughout the metapopulation system, not only depends on the infection parameters, but also on the mobility rates of the individuals [142, 212]. This is a remarkable result as it allows a better understanding of current diseases and opens up new ways to deal with global threats to public health systems. For instance, one could study how different mobility patterns impact the spread of the disease, which will allow the evaluation of several strategies based on mobility reduction to contain a global pandemic.

10.1.1 Baseline Metapopulation System

Previous studies have considered a markovian setting for the movement of the individuals among subpopulations [142, 212]. Therefore, travelers do not have memory of their home and every unit time all of them have the same probability − as given by a diffusion rate λ_{ij} − to travel from subpopulation i to subpopulation j. While this approach has been widely used for very large populations, the most natural way to capture the spread of the infection due to the movements of individuals is to consider the non-markovian situation in which individuals are assigned a *home*, i.e., a subpopulation where they remain if they are not traveling. Moreover, every time individuals start a travel to other subpopulations, the latter should not necessarily be in their immediate (next-nearest) neighborhood. The most general scenario should include the

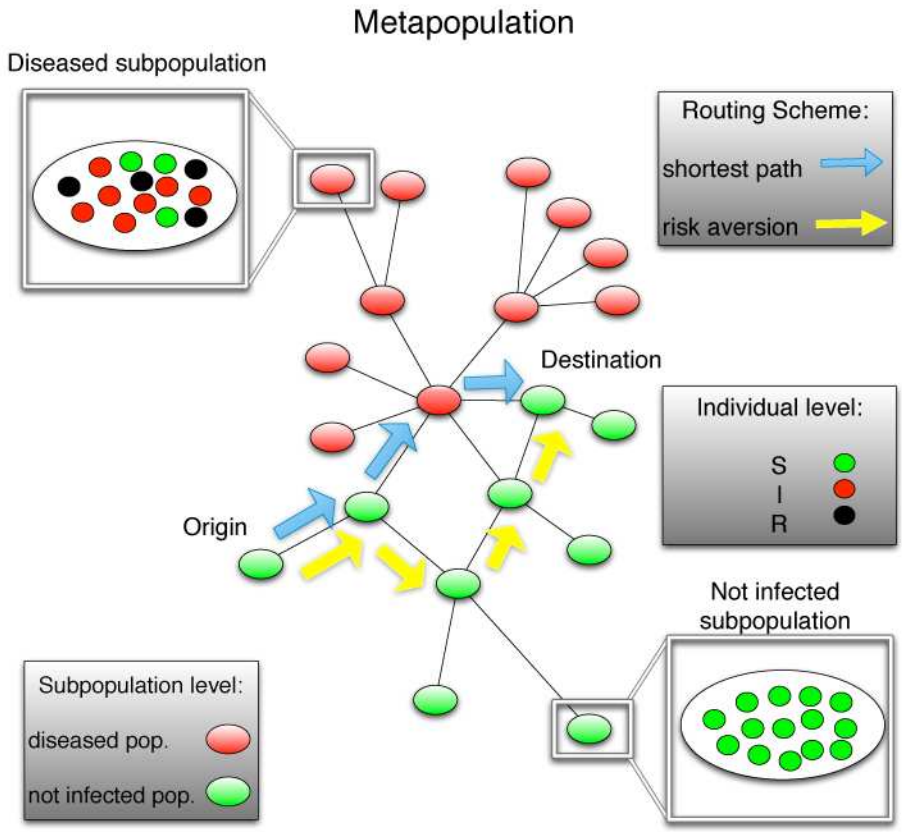

Figure 10.1: Schematic representation of the Metapopulation System. A population of individuals is divided in V subpopulations, which are connected with each other following a heterogeneous network. Within each subpopulation, individuals are classified according to their dynamical status as susceptible, infected and removed. In absence of behavioral changes (blue arrows), individuals move from a subpopulation to another at a rate λ following the shortest path connecting both subpopulations. We also illustrate one of the two mechanisms of risk aversion explored. When prevalence-based information is available, individuals might avoid traversing diseased subpopulations, at the cost of traveling along largest paths (yellow arrows).

possibility that individuals go through different subpopulations before arriving to their final destinations.

The present model considers a population of size N divided into V subpopulations. At the initial state, individuals are ascribed a *home*. Every time unit each of the N_i individuals ($N = \sum_i N_i$) of subpopulation i travels with probability $\lambda_{ij} = \lambda$. For simplicity, let us first consider the case in which destinations j are randomly chosen (other choices are discussed below). We note that this only applies to individuals at home, because when they are traveling, they move every unit time with probability one to the next subpopulation in their way to the final destination. Additionally, once the travelers arrive to their destinations, they go back home. As for the traveling routing, we consider as the most plausible setting that individuals travel along the shortest path connecting the origin and the destination. In each node a SIR dynamics takes place over a well mixed population of initial size $N_i(0) = w_i$, being w_i the strength of node i. When time goes on, $N_i(t)$ changes according to the number of individuals that has been received and has left the node. Within the nodes, one step of a SIR process takes place. The state of every individual inside a node i is changed according to the following probabilities: a susceptible individual becomes infected with probability $p^{(S\to I)} = 1 - (1 - \frac{\beta}{N_i})^{I_i}$, and an infected recovers with probability $p^{(I\to R)} = \mu$. Specifically, the exact number of individuals that changes its state is determined by a binomial distribution with the probability $p^{(S\to I)}$ (or $p^{(I\to R)}$) and the susceptible populations size $S_i(t)$ (or infected $I_i(t)$) as parameters. Note that in this scenario, R_0 only participates in the internal dynamics; individuals traveling through node i are involved in the SIR and thus can change their state while at node i.

The SIR model [209, 161] is characterized by the average number of infectious individuals produced by a single infected individual in a fully susceptible population, the so-called reproductive number $R_0 = \beta/\mu$. If $R_0 > 1$, an outbreak takes place [209, 161]. For the case of metapopulation models, the previous relation holds. However, if the mobility rate of the individuals of the originally infected subpopulation is not large enough as to ensure the seeding of other subpopulations, the outbreak could remain local. As noted before, a global invasion threshold have been shown to exist when individuals move among subpopulations following a markovian assumption [142, 212]. We first show that our metapopulation system without any human behavioral reaction also exhibits a global invasion threshold.

Let us consider a network $\mathbf{G}(V, M)$ with V nodes, M links and degree distribution $P(k)$. Each node i of the network is considered a subpopulation with

N_i individuals. We set the subpopulation size proportional to its degree. Next, assume a diffusion process in which each individual in a node i (origin) travel to another node j (destination) of the network with a probability λ. The destination j is selected with a probability proportional to the subpopulation size N_j. Once individuals engage in a travel, the shortest path is selected among all the possible paths connecting the origin and the destination of the travel. Besides, we also consider that individuals come back to their home subpopulations after they reach their destinations. A standard convenient representation of the system is provided by quantities defined in terms of the degree k:

$$N_k = \frac{1}{V_k} \sum_{i|k_i=k} N_i. \tag{10.1}$$

Let us consider that an individual of a subpopulation of degree k gets some infectious disease characterized by a reproductive number $R_0 > 1$. Let us define D_k^0 as the number of diseased subpopulations of degree k at generation 0. In the early stage, the number of diseased subpopulations is small, thus, we can study the evolution of this number using a tree-like approximation relating D_k^n with D_k^{n-1}. The average number of infected individuals in the class of degree k during the evolution of the epidemic is αN_k. The parameter α depends on the specific disease. Each infected individual stays in the infectious state for an average time μ^{-1}. Then the number of infected people circulating through the network after $n-1$ generations is:

$$\omega^{n-1} = \frac{\lambda\alpha}{\mu} \sum_{k'} D_{k'}^{n-1} N_{k'} \tag{10.2}$$

The number of infected individuals that will pass through a subpopulation of degree k will be a fraction of Eq. (10.2) proportional to the topological betweenness (in general, it is proportional to the algorithmic betweenness, but given that individuals are following the shortest path, it coincides with the topological one in our case). This measure is defined as:

$$b(i) = \sum_{\substack{j,l=1,n \\ i\neq j\neq l}} \frac{\mathcal{D}_{jl}(i)}{\mathcal{D}_{jl}}, \tag{10.3}$$

where $\mathcal{D}_{jl}$ is the total number of shortest paths from j to l and $\mathcal{D}_{jl}(i)$ is the number of shortest paths from j to l that goes through i. The latter quantity

also measures the centrality of a node assuming a diffusion scenario in which travelers go through the shortest paths. We can then write:

$$\gamma_k^{n-1} = \frac{b_k}{b_{tot}} \omega^{n-1}, \tag{10.4}$$

where b_{tot} is the sum of all the betweenness of the nodes. For the n^{th} generation we have:

$$D_k^n = V_k \left(1 - \frac{D_k^{n-1}}{V_k}\right) \left[1 - R_0^{-\gamma_k^{n-1}}\right], \tag{10.5}$$

where the first term on the right is the probability that the subpopulation is not already seeded by infected individuals and the last is the probability that the new seeded subpopulation will experience an outbreak. In the early time and for $R_0 \sim 1$ we can approximate the last expression considering:

$$\frac{D_k^{n-1}}{V_k} << 1, \tag{10.6}$$

and

$$1 - R_0^{-\gamma_k^{n-1}} \sim (R_0 - 1)\gamma_k^{n-1}, \tag{10.7}$$

obtaining:

$$D_k^n = (R_0 - 1)V_k \gamma_k^{n-1} = (R_0 - 1)\frac{\lambda\alpha}{\mu} V_k \frac{b_k}{b_{tot}} \sum_{k'} D_{k'}^{n-1} N_{k'}. \tag{10.8}$$

Considering at the equilibrium:

$$N_k = \frac{k}{\langle k \rangle} \bar{N}, \tag{10.9}$$

where $\bar{N} = \sum_k P(k) N_k$ is the average subpopulation size, we get:

$$D_k^n = (R_0 - 1)\frac{\lambda\alpha}{\mu} \bar{N} V_k \frac{b_k}{b_{tot}} \frac{1}{\langle k \rangle} \sum_{k'} D_{k'}^{n-1} k'. \tag{10.10}$$

Let us define now $\Theta^n = \sum_k D_k^n k$, then we have:

$$\Theta^n = (R_0 - 1)\frac{\lambda\alpha}{\mu} \bar{N} \frac{\Theta^{n-1}}{\langle k \rangle} \sum_k V_k k \frac{b_k}{b_{tot}}. \tag{10.11}$$

The last term needs can be further developed as:

$$\sum_k V_k k \frac{b_k}{b_{tot}} = \frac{V \sum_k P(k) k b_k}{V \sum_{k'} P(k') b_{k'}}. \tag{10.12}$$

Considering now $b_k \sim k^\eta$ one is left with:

$$\Theta^n = (R_0 - 1) \frac{\lambda \alpha}{\mu} \bar{N} \frac{1}{\langle k \rangle} \frac{\langle k^{1+\eta} \rangle}{\langle k^\eta \rangle} \Theta^{n-1}. \tag{10.13}$$

We finally get the global invasion threshold as:

$$R^* = (R_0 - 1) \frac{\lambda \alpha}{\mu} \bar{N} \frac{1}{\langle k \rangle} \frac{\langle k^{1+\eta} \rangle}{\langle k^\eta \rangle}. \tag{10.14}$$

We can write the threshold condition for the mobility rate:

$$\lambda \bar{N} \geq \frac{\langle k^\eta \rangle}{\langle k^{1+\eta} \rangle} \frac{\langle k \rangle \mu}{\alpha} (R_0 - 1)^{-1}. \tag{10.15}$$

These last two expressions are the crucial quantities, and give the conditions for a global outbreak. It is important to remind that in metapopulation networks the condition $R_0 > 1$ for each subpopulation is not enough to infer whether a finite number of subpopulations will be affected by the disease. In Fig. 10.4, we compare results from numerical simulations with the analytical prediction expressed in Eq.(10.15). The figure shows that the mean-field framework nicely recovers the value of the mobility threshold beyond which the movement of individuals from infected subpopulations to susceptible ones is large enough as to seed the latter and spread the epidemic to a system-wide scale.

Comparison with numerical results

To compare the analytical insights with the numerical results we choose as substrate an uncorrelated scale free network generated according to the configuration model with $\gamma = 2.5$ and $N = 3000$. First of all we tested the assumption made in Eq. (10.9), in which the number of individuals N_k at nodes of degree k, at the equilibrium, is proportional to k. To do so, we start the simulation with a population of $N_i \simeq 1000$ in each node, wait until the traffic equilibrium has been reached, and finally we collect the values of N_i. Fig. (10.2) shows the values of N_k as function of degree k, justifying our assumption. In order to

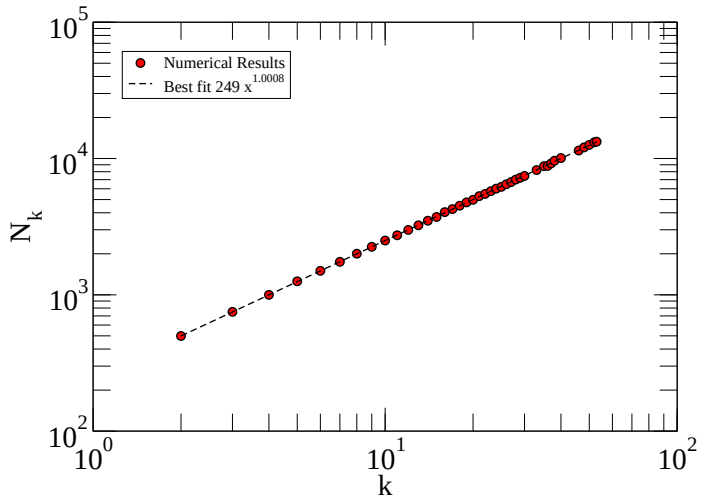

Figure 10.2: N_k as a function of the degree classes k and the best fitting curve, leading to a linear scaling.

calculate the critical mobility rate λ_c we use Eq. (10.15) and λ_c reads as:

$$\lambda_c = \frac{1}{\overline{N}} \frac{\langle k^{\eta} \rangle}{\langle k^{1+\eta} \rangle} \frac{\langle k \rangle \mu}{\alpha} (R_0 - 1)^{-1}. \tag{10.16}$$

Thus, we need to know the specific value of η in the chosen network and fix a value for R_0. To obtain an estimate for η we compute the value of the betweenness b_i for each node i and coarse grain by degree classes k.

$$b_k = \frac{1}{V_k} \sum_{i|k_i=k} b_i. \tag{10.17}$$

Note that to evaluate b_i of each node we decide to make a run of the simulation and register the number of packets that pass through a link over a very long period of time. In this way, the values of b_i are more precise and closer to the

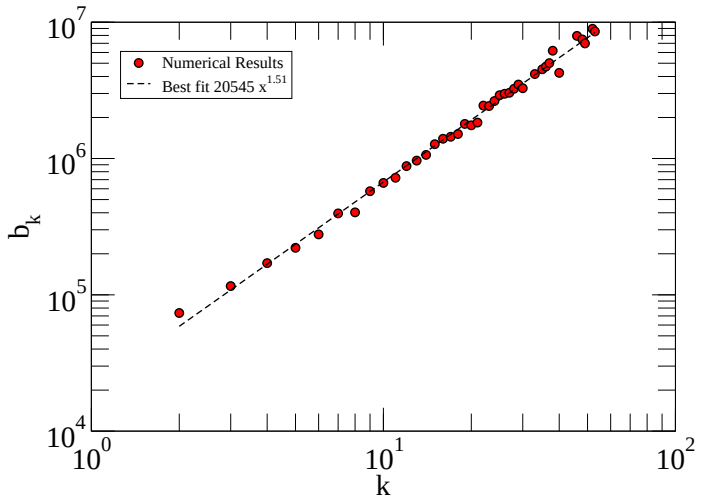

Figure 10.3: b_k as a function of the degree classes k and the best fitting curve, leading to a value of $\eta = 1.51$

actual dynamics. Fig. (10.3) shows b_k as function of the degree classes and the fit for η gives a value of $\eta = 1.51$.

Now we have all the ingredients to calculate the critical value λ_c and compare it with the numerical results.First we calculate the mean degree of the network $\langle k \rangle = 4.0$ and then the η^{th} moment of the degree distribution obtaining $\langle k^{\eta} \rangle = 10.93$ and finally the $(1+\eta)^{th}$ moment $\langle k^{1+\eta} \rangle = 158.60$, we also decide to fix $R_0 = 1.5$ and the value of $\bar{N} = 1000$, and substituting α in Eq. (10.16) this lead to:

$$\lambda_c = \frac{1}{\bar{N}} \frac{\langle k^{\eta} \rangle}{\langle k^{1+\eta} \rangle} \frac{\langle k \rangle \mu R_0^2}{2(R_0 - 1)^2} = \frac{1}{1000} \frac{10.93}{158} 4.0 \cdot 0.04 \cdot \frac{1.5^2}{2 \cdot 0.5^2} = 0.0000496. \tag{10.18}$$

In Fig. (10.4) we show the good agreement obtained when comparing the numerical simulations of the model and the analytical prediction for the global

invasion threshold.

10.1.2 Metapopulation System with Behavioral Changes

Next, we focus on our main goal in this work and explore several scenarios of self-initiated human behavioral changes and their impact on the spread of the disease. To this end, we consider that concurrent to the propagation of the infection, prevalence-based information about disease outbreaks triggers the spread of awareness, which promotes different behavioral changes. These reactions have been shown [207] to modify the disease state of the individuals [208, 213], the model parameters [214] and the contact structure [215, 203, 205, 204]. However, the influence of behavioral changes in metapopulation systems remains elusive. In our a system, human responses to the presence of a disease might have a direct impact on mobility and traveling habits, since avoiding infected spots can cause a rise in the use of alternative paths. A yet more drastic reaction to be included is that individuals retreat from traveling at all.

We consider that individuals react to prevalence-based information, and study two mechanisms of risk aversion. In the first mechanism we assume that people tend not to start a journey if the destination is known to be affected by a certain disease, thus the number of cancelled trips can be considered proportional to fraction of infected individuals at the destination. The second case (see Fig. 10.1) considers behavioral reactions that induce changes in traveling routes, a sort of epidemic-aware bypath routing. In the next sections we describe the different mechanisms we use to model behavioral changes.

10.1.3 Departure probability

During the H1N1 pandemic in 2009, especially in the early stage, a big drop in the number of travelers to (and within) Mexico was registered [210]. A first plausible mechanism to model behavioral changes is obtained changing the probability of departure according to the stage of the disease at a given destination. We can thus assume that individuals might decide to postpone their trips. Mathematically this behavior can be modeled as:

$$\lambda \to \lambda_{ij} = \lambda \left[1 - \frac{I_j(t)}{N_j(t)} \right]. \tag{10.19}$$

The expression above considers that the probability that each individual will travel is not anymore constant but a function of the epidemic incidence at destination. At the beginning of the spreading process, when the number of infected

individuals is small, the mobility is given by λ for all possible destinations j, since the second factor above is close or equal to 1. However, during the evolution of the disease, as soon the number of infected individuals increases, the departure probability starts to change from place to place and the mechanism becomes effective.

10.1.4 Rerouting

Let us assume that a traveler from subpopulation i has as destination subpopulation j, and that a node m is in the shortest path between its origin and destination. Let us also suppose that subpopulation m is experiencing a severe outbreak. The individual could decide to travel anyway but changes the route going through another, maybe longer but less risky alternative path. We modeled this kind of behavior by introducing a cost function:

$$c_m(t) = h\delta_m + (1-h)\frac{I_m(t)}{N_m(t)}, \tag{10.20}$$

where the parameter h is defined in the closed interval $[0,1]$ and δ_k can assume three values $[-1,0,1]$. -1 is associated to the shortest path (the individual will be one hop closer to its destination), 0 to a new path which does not change the current distance to the destination, and 1 otherwise (the individual will be one hop farther to its destination). At each time step, each individual that is traveling decides the next node to move to by minimizing the cost function, unless the next move leads to its destination.

It is worth noticing that the parameter h, although defined in the interval $[0,1]$ can take on only a small subset of meaningful values. If h is too small, the traveler essentially moves through the network following the landscape of epidemic incidence, as no information of its destination is taken into account when deciding where to move. This is a highly unrealistic situation that therefore sets a lower bound (> 0) to h. Similarly, one can easily show that h is also bounded from above. Although $h = 1$ mathematically corresponds to the limit of shortest path, this limit is obtained well before. Admittedly, one can show that in order for a traveler to go through a path one hop farther than the destination the following condition must be satisfied

$$\frac{I_-(t)}{N_-(t)} - \frac{I_+(t)}{N_+(t)} > \frac{2h}{1-h}, \tag{10.21}$$

where $\frac{I_-}{N_-}$ and $\frac{I_+}{N_+}$ are the densities of infected individuals at subpopulations one hop closer and one hop farther from the traveler's destination, respectively.

The same argument leads to the following condition with respect to the possibility of going through a path that does not change the current distance to destination:

$$\frac{I_-(t)}{N_-(t)} - \frac{I_=(t)}{N_=(t)} > \frac{h}{1-h}, \tag{10.22}$$

where $\frac{I_=}{N_=}$ is the density of infected individuals at a subpopulation which is at the same distance of the traveler's destination. We have checked that our results for the model including behavioral changes are qualitatively the same for different values of h in the interval $[0.05, 0.2]$ (beyond $h = 1/3$, no differences with respect to the shortest path results are obtained, which means that this limit has been reached at $h = 1/3$ as expected from Eq. (10.21)).

10.1.5 Mechanistic Numerical Simulations

We resort to extensive numerical simulations and study a variety of cases. Furthermore, motivated by the recent empirical evidence on the heterogeneous properties of transportation and mobility patterns, we carried out Monte Carlo simulations on top of the *Air Transportation Network* (ATN). The ATN is composed of passenger flights operating in the time period November 1, 2000, to October 31, 2001 compiled by OAG Worldwide (Downers Grove, IL) and analyzed previously in [149]. It consists of 3618 nodes (airports) and 14142 links. We use the weighted network in our analysis. Airports corresponding to a metropolitan area have been collapsed into one node in the original database. This network is highly heterogeneous in the number of connections between urban areas (subpopulations) as well as in the traffic w_{il} in terms of the number of passengers or available seats on a given direct route connecting two subpopulations i and l. Taking into account these traffic patterns, we assume that initially the number of individuals in subpopulation i is proportional to its strength $N_i = \sum_l w_{il}$. Moreover, as the travel flows are not homogeneous, we also consider that individuals choose their destinations proportionally to the strengths of each possible target subpopulation. In this way, we preserve the inter-city traffic patterns. Finally, it is possible to make the model more realistic by considering that infected individuals diffuse at a lower rate, λ^I, than susceptible subjects and that all travelers spend a time τ at their destinations before coming back home. Without loss of generality we assume that $\lambda^I = \lambda/2$ and that τ is drawn from a uniform distribution with mean equal to the traveled distance d_{il}.

The overall dynamics is determined by the way agents select which neighbor to visit next on their ways to their destinations, presuming that following

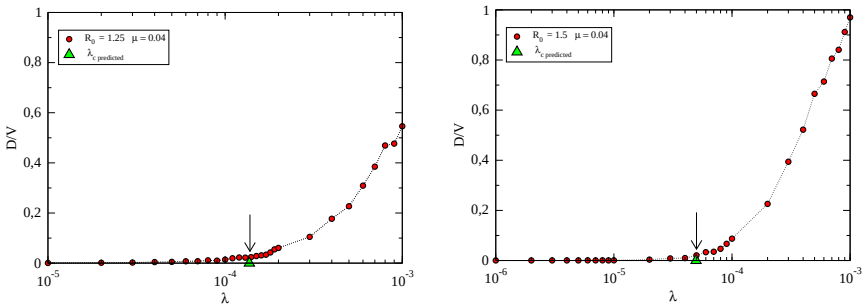

Figure 10.4: To compare the analytical insights with numerical results, we represent in the figure the number of diseased subpopulations D/V as a function of the mobility rate for two R_0 values. The analytical value (Eq.(10.15)) is indicated by the arrow. Full-circles are results from numerical simulations (the line is a guide to eyes). The value of α has been approximated by $\alpha = 2(R_0 - 1)/R_0^2$ [161]. The substrate topology is an uncorrelated scale-free network generated according to the uncorrelated configuration model [216] with $\gamma = 2.7$ and $V = 3000$ subpopulations. Other parameters are indicated in the figure.

the shortest path is the preferred solution in an infection-free scenario. At each subpopulation, traveling individuals are tracked in time and the evolution of the disease monitored. In addition, we study the invasion dynamics and measure the number of diseased subpopulations at time t, $D(t)$. All numerical results reported henceforth are averages taken over at least 100 realizations of the initial conditions and the stochastic dynamics. We next separately consider the aforementioned behavioral responses which comprehend a whole range of possibilities. Intuitively, risk-aversion strategies should not affect the invasion threshold. This is because near the global invasion threshold, the probability to find an infected subpopulation other than the one where the epidemic was originally seeded vanishes. Therefore, behavioral changes are expected to impact the course of the epidemic once the outbreak takes place and has reached a macroscopic fraction of the total number of subpopulations in the metapopulation system.

10.2 Discussion

In Fig. 10.5 we analyze the behavior of the density of diseased subpopulations D/V at the end of the global epidemic as a function of both the basic reproductive number R_0 and the traveling diffusion rate λ. The results reported correspond to three different situations. Symbols represent the final fraction of diseased subpopulations when no risk aversion mechanism is present. In this case, individuals diffuse from one subpopulation to another along the shortest path connecting their home with their destination. As specified above, infected subjects have a lower mobility rate than susceptible and removed individuals, and all of them spend a time τ at destination before engaging in the way back home. This constitutes our *null model*. Dotted lines correspond to the situation in which individuals react to the presence of the epidemic such that they travel with a probability that depends on the number of infected individuals at their destination. Finally, solid lines represent the second risk-aversion response, in which individuals seek a trade-off between the epidemic incidence along the path connecting the origin and destination of the travel and the length of the latter.

Figure 10.5 shows that in all cases analyzed the metapopulation system exhibits an invasion threshold which is independent of human behavioral changes. This threshold determines whether the flow of individuals between subpopulations is high enough in order to allow the infectious disease to invade a finite fraction of the whole metapopulation system. The figure also clearly evidences the radically distinct effects of the two risk aversion behavioral responses. When people have the option of deciding whether or not to engage in a travel, the fraction of diseased subpopulations at the end of the outbreak decreases with respect to the null case, thus pointing out that this kind of response is beneficial. The reason for the reduction in D/V is rooted in the effective reduction of the mobility rate of the individuals which lead to a smaller exposure of susceptible individuals, both while traveling and at home, to the disease. This is in line with the usual measures adopted externally by public health authorities, which are aimed at reducing the frequency of contacts between susceptible and infected individuals and therefore to contain the spreading of the disease by cutting down the number of susceptible-infected transitions.

A more striking and unexpected result is obtained for the second risk aversion scenario. As Fig. 10.5 shows, it turns out that when travelers decide to bypass the subpopulations with a high number of infected individuals, the disease penetration, as given by the number of subpopulations with local outbreaks, increases. The rationale behind this finding is that the increase in the

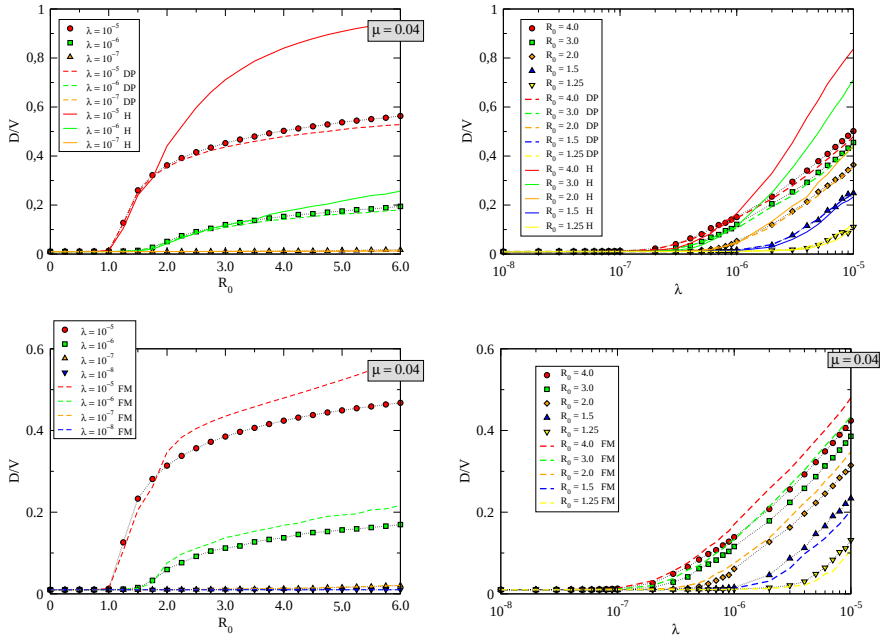

Figure 10.5: The figure compares the fraction of diseased subpopulations D/V when risk aversion mechanisms are active, with the situation in which such behavioral responses are not taken into account. Panels (a) and (b) show, respectively, the dependency of D/V with the basic reproductive number R_0 and the mobility rate λ for random scale-free networks. Symbols represent the results obtained when individuals do not react to the presence of the disease. The rest of results correspond to the two risk aversion responses: DP stands for the mechanism in which individuals decide whether or not to travel, while H corresponds to the value used for the cost function in the second case of behavioral change. Moreover, in panels (c) and (d) we compare the behavior of the fraction of diseased subpopulations D/V for the null and full versions of the metapopulation system on top of the worldwide air transportation network. The results confirms that the invasion threshold is independent of behavioral changes and that the latter have a significant impact on the invasion dynamics of the metapopulation. In particular, for fixed mobility rates, including behavioral changes could have a larger effect in D/V than increasing the value of R_0. Model parameters are those used in panels (a) and (b). See the main text for further details.

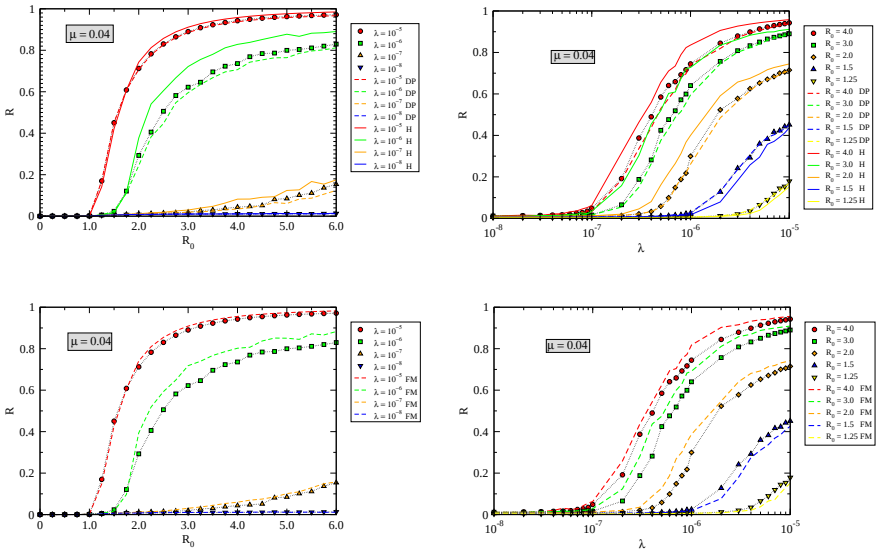

Figure 10.6: The figure compares the fraction of recovered individuals R when risk aversion mechanisms are active, with the situation in which such behavioral responses are not taken into account. Panels (a) and (b) show, respectively, the dependency of R with the basic reproductive number R_0 and the mobility rate λ for random scale-free networks. Symbols represent the results obtained when individuals do not react to the presence of the disease. The rest of results correspond to the two risk aversion responses: DP stands for the mechanism in which individuals decide whether or not to travel, while H corresponds to the value used for the cost function in the second case of behavioral change. Moreover, in panels (c) and (d) we compare the behavior of the fraction of recovered individuals R for the null and full versions of the metapopulation system on top of the worldwide air transportation network.

flow of individuals going through alternative paths brings in the infection to new subpopulations that would otherwise remain susceptible. This constitutes a very interesting finding, as one can thought of the whole process in terms of a social dilemma or a public goods game. Individuals adopt a sort of selfish behavior by avoiding highly infected spots, but as a consequence, the disease invades a larger fraction of the subpopulations in the metapopulation system. Thus, what is beneficial at the individual level, turns out to be detrimental for the wellbeing of the whole population. It is also worth stressing that the effects of this kind of behavioral changes in front of a disease non trivially depend on both R_0 and λ. As seen in the bottom panel of Fig. 10.5, for fixed and relative high mobility rates, this kind of risk aversion response may provoke that the number of diseased subpopulations for a disease with an associated $R_0^{'}$ be larger than the corresponding fraction of diseased subpopulations for another disease characterized by a different $R_0^{''}$ ($R_0^{''} > R_0^{'}$), but without this type of risk aversion effects.

We further represent in Fig. 10.6 the results obtained with the full version of the metapopulation system together with the null case for the fraction of recovered individuals R. For the fully developed metapopulation system, all mechanisms are active concurrently, so that the two risks aversion behaviors add in a way that individuals first decide whether or not to travel and if they engage in a travel, then they move following the second aversion dynamics. The results confirm that strongest effects are induced when alternative paths are used for movements between subpopulations. Here, we also confirm the existence of a global invasion threshold that does not depend on behavioral changes. Also, in Fig. 10.7 we present the infection trees (see caption for description) for the baseline and the fully developed model highlighting the differences not only in the number of infected subpopulations but also in the time and methods contagion occurs.

10.3 Conclusions

In summary, we have analyzed a metapopulation system where individuals move from origins to destinations that can be several subpopulations apart. The model incorporates several scenarios of self-initiated human behavioral responses in front of a disease. We first have shown that there is a global threshold associated to the subpopulation invasion dynamics. We have reported an analytical expression that gives the minimum number of individuals diffusing among subpopulations in order to have a global epidemic outbreak.

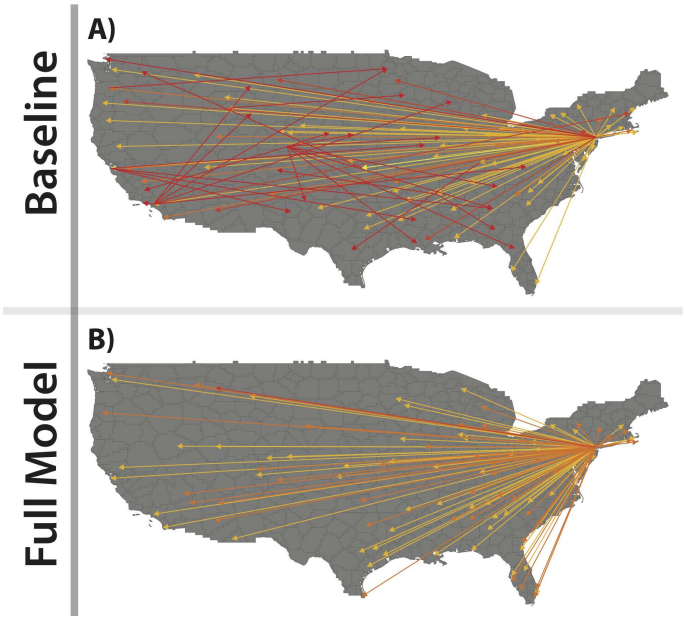

Figure 10.7: The picture depicts the infection tree (a directed tree in which a link between subpopulations i and j is drawn if an individual infected in i has caused an outbreak in j) for the baseline (A) and the fully developed model (B) on the US air transportation network. Colorscale from light yellow to dark red indicates the time at which contagion occurs.

Our main results concern the effects of self-initiated, prevalence-based behavioral responses and their impact on the mobility and traveling patterns and the number of diseased subpopulations at the end of the epidemic. Other reactions are expected to have a further impact in the evolution of a disease, but at a more local level.

We believe that future approaches should be addressed to improve our understanding of the spreading dynamics among subpopulations. After all, what differentiate today's global diseases is the fast time scale associated to the global pandemic. Our study may also be important in other contexts where population dynamics plays a role. We have shown that new migration routes could significantly modify the dynamics of a metapopulation system. Finally,

our findings suggest that human behavioral responses to the presence of a disease have a large impact not only at a local level, but more importantly, could shape the evolution of global pandemics. In this sense, behavioral adaptation should be included as a key factor in current epidemiological platforms. As our results show, some reactions could even lead to worst situations than previously thought.

Conclusions

Conclusions

As large networks are all around us and grow at increasing rates it becomes more and more important understand and control their functioning. To do so a comprehensive and multidisciplinary approach is often needed. This thesis tackles the problem of predicting the dynamical evolution of interconnected systems and enhances their capabilities from three different perspectives ranging from robotics to epidemiology.

First we concentrated on sensors and mobile networks tackling the problems of creating a distributed and robust topology for static sensors and the study of the evolution of cooperation when agents can move mimicking unmanned robots behavior. For the first problem we model a real environment as a 2D space in which nodes are distributed randomly and have a limited visibility radius. The problem consists of finding a pattern of connections that assures the whole network connectivity (when possible) and shows the optimal resilience to failures and targeted attacks. We have proposed a distributed algorithm that permits to create a random geometric graph with a prescribed degree distribution. Then we exploited recent results on percolation theory applied to complex networks topology to choose a multi-modal degree distribution that maximize both the percolation thresholds for failures and attacks. Numerical tests against a randomized ensemble of graphs, taken as a null model, showed that our recipe outperform the null model in several aspects of network resilience as the size of the giant component and the size and the number of connected components.

Then we placed on a more complex scenario characterized by interacting moving agents playing an evolutionary version of a social dilemma. In this case we concentrated on the evolution of cooperation as a convenient strategy with the aim of highlighting the conditions for the emergence of a collaborative behavior. We proposed a model in which agents move and interact with other agents in their visibility radius playing a round of prisoner's dilemma and collect

a payoff. An evolutionary step is then made assuring the selection of the best strategy. Numerical evidence showed that, although defection is favoured, there exists a consistent range of parameters space in which cooperation becomes advantageous. Such conditions are fulfilled when agents density is enough to produce small clusters in which cooperation can grow and players velocity is small enough to assure clusters stability for a reasonable amount of time. These results open many interesting questions that can be answered in the future. Is it possible to control agents movement via some simple evolutionary rules? Does cooperation allow agents to solve simple tasks collectively? Can we exploit such scenario to create next generation communication networks that relies on several mobile devices without a pre-constructed backbone?

In the second part we have faced a couple of interesting problems on top of communication networks: the study of fluctuations in mean flows in a traffic network and an optimization technique for congestion control. Firstly the problem of finding a relationship between the mean flow on a node and its standard deviation has been analyzed. Recent works on the topic have highlighted the presence of some scaling laws but their validity remains controversial. We presented a model based on random walks theory that incorporates most of the characteristics of real communication systems such as network structure and fluctuations in external systems arrivals. The solution of the model evidenced a direct relation between flows fluctuations and three factors, namely: the variations in the number of packets in the network, the degree of the nodes and the length of the time window in which measurements are performed. Analytical predictions have been confirmed via three different analysis: a simple numerical traffic model on top of a scale-free topology, real data analysis, and events simulations via a realistic traffic generator.

On chapter 7 we tackled an open question that received a lot of attention in the literature: finding an optimization technique that allows to handle high traffic rates on a network delaying the onset of the congestion. We approached the problem in a theoretical way proposing a minimal model for traffic simulation that incorporates a simple control strategy nowadays present in real communication networks: refusing to receive a packet with a certain probability to control the growth of nodes queues. We showed that some simple adaptive strategies can considerably slow down the onset of congestion, but modifying the nature of the transition between free flow regime and the congested state from a smooth to an abrupt one. As an additional element we also permit nodes a simple form of collaboration: consider in the adaptive process the congestion state of their neighbors. This latter ingredient has a profound impact on traffic dynamics leading to a delay on the onset of congestion and a

smooth transition between free flow and congested regimes. Another striking property of this local empathetic behavior is that by introducing it the system recovers the same performances that a global optimization strategy.

Although a huge quantity of works have been proposed on these two topics, many open questions remain unanswered. Can we produce traffic models that faithfully reproduce traffic flows on the Internet that are shaped by commercial and peering relationships rather than simple topological constraints? Is it possible to approach the traffic problem as a cooperative game between agents?

On the last part of the work we focused on spreading processes on networks. Our study started with a simple analytical formulation to model the spreading of a disease in a class of interaction rules. Then, we move on the modeling of a more complex scenario in which interactions are described by traffic flows. We conclude our analysis with a fully developed metapopulation system in which the impact of human responses to the epidemic spreading is considered. In the first study we proposed an alternative formulation to the classical heterogeneous mean field to study a SIS model on scale free networks. The model can be integrated numerically and we analytically derived its equivalence with the HMF by recovering the epidemic threshold. In addiction to the HMF this approach permits to predict the single nodes infection probability and can be applied to an entire class of interaction models ranging from contact process to a fully reactive scenario like the HMF. Numerical simulations confirm our findings. After that we have presented a model in which nodes interactions are driven by traffic flows in a coupled fashion. Packets are seen as quanta of interaction between individuals and as the way the disease can spread in the population. The analytical treatment of the model via an HMF approach showed that the epidemic threshold strictly depends on the traffic values of the system and, in case of a bounded delivery rate, it can assume a finite value also for very high traffic intensities. In the last chapter of this part a profound extension of the previous model has been presented. In this case a metapopulation framework has been adopted modeling subpopulations as nodes and incorporating human mobility. In a SIR scenario individuals can transport the disease from one supopulation to another. Numerical simulations confirmed the theoretical prediction for the global invasion threshold and two schemes for human responses to the epidemic diffusion have been analyzed. In the first individuals are allowed to cancel their journey with a probability proportional to the fraction of infected individuals at the destination. In the second we permit them to take a longer but less infected path to the destination. Numerical analysis showed an unexpected result. Although cancelling a journey can only slightly reduce the epidemic incidence a longer but safer path can have a dra-

matic outcome bringing the disease in places that otherwise will be untouched and raising consistently the epidemic incidence.

Despite of these recent advances in epidemic modeling on complex networks a complete knowledge of epidemic dynamics is still far from our view. Models able to cope with individuals differences in infection rates and efficient policies to block the spread of a disease are needed. Also in modeling interactions more complex approaches that incorporate temporal patterns are of crucial importance.

Concluding, in this thesis the characteristics of several dynamical models on top of large and complex topologies have been presented and some control techniques have been tested. Today's world is complex and it's impossible to think at our lives without the connections that we all share. But such a complexity can be managed and exploited to bring us to a better life. In this work we have tried to shed light on some complex behaviors that emerge in technological networks and I hope that this contribution as little as it is could be useful to someone else in the future.

Other Research Activities

Simultaneously with the research for the development of this thesis, other topics in different areas have been dealt with:

- *Synchronization of non-identical oscillators in a power micro-generation scenario.* Future scenarios of electrical distribution networks with many, geographically dispersed power micro-generators from renewable sources impose to cope with the problem of the functional stability of such networks. Does the network topology play a role in ensuring a viable synchronization stability ? We present a modified version of the Kuramoto model, used to mimic a network of power generators, has been applied to estimate the synchronization behavior of large networks with different topologies (random, scale-free) in homogeneous (i.e. all nodes and links with equal properties) configurations. The network stability upon link(s) removal (in term of resistance to synchronization loss) has been also analyzed.

- *De-Novo Assembly of Quasispecies via Next-Generation Sequencing.* The problem of reconstruct a viral quasispecies from a Next Generation Sequencer (NGS) consists in reproducing the original genome from the samples obtained via a shotgun sequencing machine. In this field we applied some concepts from complex networks theory for the de-novo assembly (i.e. without the need of any reference genome) of NGS samples of a quasispecies, in particular viral quasispecies differentiating within a host, and eventually quasispecies showing evidence of recombination. The proposed method relies on a statistical procedure to construct an overlap graph and then on an algorithm based on random walks to reconstruct the different variants of a quasispecies from a NGS sample. The overall graph can be also analyzed via classical networks analysis tools to obtain insights on

quasispecies features as multiple variants with variable mutation rates, recombinants, and repeated regions.

- *Effect of Topology on Diversity of Spatially-Structured Evolutionary Algorithms* We study of the effects of networks topology to spatially-structured evolutionary algorithms' dynamics. We focused our study on takeover time and diversity of the solutions, applying the algorithm on a multimodal optimization problem. Using as algorithms' underlying structure different networks models we studied the relationship between algorithm's dynamic, i.e. takeover time, first hitting time and number of distinct optima found at convergence, and networks' features like clustering coefficient and characteristic path length. A comparison with a panmictic evolutionary algorithm is made to study the effects of the introduction of a structure in the mating dynamics of the algorithm, resulting in an enhancement of diversity and containing the takeover time and first hitting time overhead.

Publications

Internetional Journals

(i) Meloni S., Gómez-Gardeñes J., *Local empathy provides global minimization of congestion in communication networks*, Physical Review E 82 (2010) 056105.

(ii) Gómez S., Arenas A., Borge-Holthoefer J., Meloni S., Moreno Y., *Discrete-time Markov chain approach to contact-based disease spreading in complex networks*, Europhysics Letters, 89 (2010) 38009.

(iii) Meloni S., Arenas A. and Moreno Y., *Traffic-driven epidemic spreading in finite-size scale-free networks*, Proc. Natl. Acad. Sci. U.S.A., 106 (2009) 16897.

(iv) Meloni S., Buscarino A., Fortuna L., Frasca M., Gómez-Gardeñes J., Latora V., Moreno Y., *Effects of mobility in a population of prisoners dilemma players*, Physical Review E, 79 (2009) 067101.

(v) Meloni S., Gómez-Gardeñes J., Latora V., Moreno Y., *Scaling Breakdown in Flow Fluctuations on Complex Networks*, Physical Review Letters, 100 (2008) 208701.

(vi) Rosato V., Meloni S., Issacharoff L., Tiriticco F., *Is the topology of the internet network really fit to its function?*, Physica A 387 (2008) 16891704.

(vii) Rosato V., Issacharoff L., Tiriticco F., Meloni S., De Porcellinis S., Setola R., *Modelling interdependent infrastructures using interacting dynamical models*, International Journal of Critical Infrastructures, 4 (2008) 63-79.

Books Chapters

(i) Meloni S., Arenas A., Gómez S., Borge-Holthoefer J., Moreno Y., *Modeling Epidemic Spreading in Complex Networks: Concurrency and Traffic*, in Handbook of Optimization in Complex Networks, (2010) Springer-Verlag Berlin Heidelberg.

(ii) Rosato V., Simonsen I., Meloni S., Issacharoff L., Peters K., Von Festenberg N., Helbing D., *A Complex Systems View of Critical Infrastructures* in Managing complexity: insights, concepts, applications, (2008) Springer-Verlag Berlin Heidelberg.

(iii) Gasparri A., Meloni S., Panzieri S., *Growing Fully Distributed Robust Topologies in a Sensor Network*, in Modelling, Estimation and Control of Networked Complex Systems, 2009 Springer-Verlag Berlin Heidelberg.

Proceedings of International Conferences

(i) Rosato V., Bologna S., Meloni S., Pascucci E., *Interdependency effects measured on complex interdependent networks*, in Compeng 2010: Complexity in Engineering (2010) IEEE.

(ii) Azzini A., De Felice M., Meloni S., Tettamanzi A.G.B., *Soft Computing Techniques for Internet Backbone Traffic Anomaly Detection*, in proceedings of 2009 Evo* Conference, EvoWorkshops, LNCS 5484 (2009) 99104, Springer-Verlag Berlin Heidelberg.

Bibliography

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[2] R. Albert and A.-L. bási, Rev. Mod. Phys. **74**, 47 (2002).

[3] Newman, M. E. J. The structure and function of complex networks. *SIAM Review* 45:167-256. (2003)

[4] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, Rev Mod. Phys. **80**, 1275 (2008).

[5] B. Bollobás, *Random Graphs*, Academic Press, London (1985).

[6] D.B. West, *Introduction to Graph Theory*, Prentice-Hall, Englewood Cliffs, NJ (1995).

[7] F. Harary, *Graph Theory*, Perseus, Cambridge, MA (1995).

[8] D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998).

[9] M. Marchiori, V. Latora, Physica A **285**, 539 (2000).

[10] S. Wasserman, K. Faust, *Social Networks Analysis*, Cambridge University Press, Cambridge (1994).

[11] M.E.J. Newman, M. Girvan, Phys. Rev. E **69**, 026113 (2004).

[12] M. Boguñá, R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).

[13] M. Boguñá, R. Pastor-Satorras, A. Vespignani, Lect. Notes Phys. **625**, 127 (2003).

[14] M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[15] M.E.J. Newman, Phys. Rev. E **67**, 026126 (2003).

[16] M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA **99**, 78217826 (2002).

[17] R. Solomonoff, A. Rapoport, B. Math. Biophys. **13**, 107-117 (1951).

[18] P. Erdös, A. Rényi, Publicationes Mathematicae **6**, 290297 (1959).

[19] B. Bollobás, Discrete Math. **33**, 1 (1981).

[20] A. Barrat, M. Weigt, Eur. Phys. J. B **13**, 547 (2000).

[21] M. Boguñá, F. Papadopoulos, D. Krioukov, Nat. Commun. **1**, 62 (2010).

[22] W. Aiello, F. Chung, L. Lu, Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, ACM, New York, (2000).

[23] G. Bianconi, A. L. Barabási, Europhys. Lett., **54**(4), 436442 (2001).

[24] A. Barabasi and Z. Oltvai, Nature Reviews Genetics **5**, 101 (2004).

[25] A.-L. Barabási, R. Albert, H. Jeong, Physica A **272** (1999).

[26] P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85** (2000).

[27] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. **85** (2000).

[28] J. Davidsen, H. Ebel, S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).

[29] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, (1944).

[30] A. Rapoport, A.M. Chammah, *Prisoners Dilemma*, University of Michigan Press, Ann Arbor, (1965).

[31] M.A. Nowak, R.M. May, Nature **359**, 826 (1992).

[32] B.J. Kim, A. Trusina, P. Holme, P. Minnhagen, J.S. Chung, M.Y. Choi, Phys. Rev. E **66**, 021907 (2002).

[33] H. Ebel, S. Bornholdt, Phys. Rev. E **66**, 056118 (2002).

[34] E. Lieberman, C. Hauert, M.A. Nowak, Nature **433**, 312 (2005).

[35] P. Holme, A. Trusina, B.J. Kim, P. Minnhagen, Phys. Rev. E **68**, 030901 (2003).

[36] P. Holme, B.J. Kim , Phys. Rev. E **65**, 026107 (2002).

[37] E.A. Bender, E.R. Canfield, J. Combin. Theory A **24**, 296 (1978).

[38] M. Molloy, B. Reed, Random Struct. Algorithm **6**, 161 (1995).

[39] M. Molloy, B. Reed, Combin. Probab. Comput. **7**, 295 (1998).

[40] M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E **64**, 26118 (2001).

[41] J. Dall, M. Christensen, Phys. Rev. E **66**, 016121 (2002)

[42] A. S. Tanenbaum, *Computer Networks*, 4th Edition, Prentice Hall PTR, (2003)

[43] S. Valverde, R.V. Sole, Eur. Phys. J. B **38**, 245 (2004).

[44] B. Tadić , S. Thurner, Physica A **332**, 566 (2004).

[45] B. Tadić , S. Thurner, G.J. Rodgers, Phys. Rev. E **69**, 036102 (2004).

[46] M. Takayasu, H. Takayasu, T. Sato, Physica A **233**, 824 (1996).

[47] P. Holme, Adv. Complex Systems **6**, 163 (2003).

[48] R. Guimerá, A. Arenas, A. Dìaz-Guilera, F. Giralt, Phys. Rev. E **66**, 026704 (2002).

[49] A. Arenas, A. Dìaz-Guilera, R. Guimerá, Phys. Rev. Lett. **86**, 3196 (2001).

[50] R. Guimerá, A. Arenas, A. Dìaz-Guilera, Physica A **299**, 247 (2001).

[51] R. Guimerá, A. Dìaz-Guilera, F. Vega-Redondo, A. Cabrales, A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).

[52] P. Echenique, J. Gomez-Gardenes, Y. Moreno, Phys. Rev. E **70**, 056105 (2004).

[53] P. Echenique, J. Gomez-Gardenes, Y. Moreno, Europhys. Lett. **71**, 325 (2005).

[54] P. Grassberger, Math. Biosci. **63**, 157 (1983).

[55] N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, Hafner Press, New York, (1975).

[56] O. Diekmann, J. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley, New York (2000).

[57] Lloyd, A. L. , May, R. M. , J. Theor. Biol. 179, **1-11**, 8 (1996).

[58] Grenfell, B. T. , Bolker, B. M. , Ecol. Lett. **1**, 6370 (1998).

[59] Bascompte, J., Solé , R.V., *Modeling Spatiotemporal Dynamics in Ecology*, Springer, New York, (1998).

[60] Hanski, I., Gilpin, M.E., *Metapopulation Biology: Ecology, Genetics, and Evolution*, Academic Press, San Diego, (1997).

[61] Levins, R., Bull. Entomol. Soc. Am. **15**, 237240 (1969).

[62] D. M. Blough, M. Leoncini, G. Resta, P. Santi, IEEE Trans. Mob. Comput., **5(9)** , 1267-1282 (2006).

[63] S. A. Borbash, E. H. Jennings, Distributed topology control algorithm for multihop wireless networks, In *in Proc. 2002 World Congress on Computational Intelligence (WCCI 2002)*, 355-360 (2002).

[64] A. Cerpa, J. Elson, M. Hamilton, J. Zhao, D. Estrin, L. Girod, Habitat monitoring: application driver for wireless communications technology. *SIGCOMM LA '01: Workshop on Data communication in Latin America and the Caribbean*, 20-41, ACM New York, NY, USA, (2001).

[65] R. Cohen, D. Ben Avraham, S. Havlin, Phys. Rev. E, **66**, 036113 (2002).

[66] R. Cohen, K. Erez, D. Ben Avraham, S. Havlin, Phys. Rev. Lett., **85** , 4626-4628 (2000).

[67] O. Dousse, P. Thiran, M. Hasler, Connectivity in ad-hoc and hybrid networks, In *Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE INFOCOM 2002.*, 1079-1088 (2002).

[68] P. Gupta and P. R. Kumar, Critical power for asymptotic connectivity, In *Proceedings of the 37th IEEE Conference on Decision Control*, 1106-1110 (1998).

[69] M. T. Hajiaghayi, N. Immorlica, V. S. Mirrokni, IEEE/ACM Trans. Netw., **15**(6), 1345-1358 (2007).

[70] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, M. Turon, Health monitoring of civil infrastructures using wireless sensor networks, In *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*, 254-263, ACM, New York, NY, USA, (2007).

[71] V. Lesser, M. Atighetchi, B. Benyo, B. Horling, A. Raja, R. Vincent, T. Wagner, X. Ping, S. XQ Zhang, The intelligent home testbed. *Proceedings of the Autonomy Control Software Workshop (Autonomous Agent Workshop)*, (1999).

[72] X.-Y. Li, P.-J. Wan, Y. Wang, and C.-W. Yi, Fault tolerant deployment and topology control in wireless networks, In *MobiHoc '03: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, 117-128, ACM, New York, NY, USA, (2003).

[73] J. Liu, B. Li, Distributed topology control in wireless sensor networks with asymmetric links, In *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, **3**, 1257-1262 (2003).

[74] Vipin Mehta, M. El Zarki, Wirel. Netw., **10**(4), 401-412 (2004).

[75] M. Molloy, B. Reed, Combin. Probab. Comput, **7**, 295-305 (1998).

2(5), 1-8 (2007).

[76] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. H. Growdon, M. Welsh, P. Bonato, Analysis of feature space for monitoring persons with parkinson's disease with application to a wireless wearable sensor system, In *In Proceedings of the 29th IEEE EMBS Annual International Conference*, Lyon, France, (2007).

[77] U. Nandini Raghavan, S. R. T. Kumara, Int. J. Sen. Netw., **2**(3/4), 201-210 (2007).

[78] V. Shnayder, B. R. Chen, K Lorincz, T. R. F. Fulford Jones, M. Welsh, Sensor networks for medical care, In *SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems*, 314-314, ACM, New York, NY, USA, (2005).

[79] M. B. Srivastava, R. R. Muntz, M. Potkonjak, Mobile Computing and Networking, 132-138 (2001).

[80] D. Stauffer, A. Aharony. *Introduction to percolation theory*, CRC Press, (1998).

[81] T. Tanizawa, G. Paul, S. Havlin, H. E. Stanley, Phys. Rev. E, **74**(1), 016125 (2006).

[82] B. Thallner, H. Moser, U Schmid, Topology control for fault-tolerant communication in wireless ad hoc networks. *Wireless Networks*, (2008).

[83] G. Werner-Allen, K. Lorincz, M. Welsh, O. Marcillo, J. Johnson, M. Ruiz, J. Lees, IEEE Internet Computing, **10**(2), 18-25 (2006).

networks. In *4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07.*, 304-313, (2007).

[84] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, (Cambridge University Press, Cambridge, 2004).

[85] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[86] M.A. Serrano, M. Boguñá, and A. Díaz-Guilera, Phys. Rev. Lett. **94**, 038701 (2005).

[87] M. Boguñá, D. Krioukov, Phys. Rev. Lett. **102**, 058701 (2009).

[88] M. Boguñá, D. Krioukov, and K.C. Claffy, Nature Phys. **5**, 74 (2009).

[89] C. Caretta Cartozo and P. De Los RiBaraos, Phys. Rev. Lett. **102**, 238703 (2009).

[90] M. Argollo de Menezes, and A. L. Barabási, Phys. Rev. Lett. **92**, 028701 (2004).

[91] J. Duch, and A. Arenas, Phys. Rev. Lett. **96**, 218702 (2006).
A. Vázquez, Phys. Rev. E **71**, 035102(R) (2005).
Eur. Phys. J. B **49**, 259 (2006).

[92] T. Ohira, and R. Sawatari, Phys. Rev. E **58**, 193 (1998).

[93] R. V. Solé, and S. Valverde, Physica A **289**, 595 (2001).

[94] S. Valverde and R. V. Solé, Physica A **312**, 636 (2002).

[95] R. Guimerá, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).

[96] R. Guimerá, A. Arenas, A. Díaz-Guilera, and F. Giralt, Phys. Rev. E **66**, 026704 (2002).

[97] B. Tadic and G.J. Rodgers, Adv. Complex Systems **5**, 445 (2002).

[98] Z. Toroczkai, and K. E. Bassler, Nature **428**, 716 (2004).

[99] B. Kujawski, J.G. Rodgers, and B. Tadic, Lecture Notes in Computer Science **3993**, 1024 (2006).

[100] B. Tadic, G.J. Rodgers and S. Thurner, Int. J. Bif. and Chaos, **17**, 2363 (2007).

[101] S. Sreenivasan, R. Cohen, and E. Lopez, Phys. Rev. E **75**, 036105 (2007).

[102] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).

[103] V. Rosato, L. Issacharoff, S. Meloni, D. Caligiore, and F. Tiriticco, Physica A **387**, 1689 (2008).

[104] S. Scellato, L. Fortuna, M. Frasca, J. Gómez-Gardeñes, and V. Latora, Eur. Phys. J. B **73**, 303 (2010).

[105] B. Danila, Y. Yu, J.A. Marsh, and K.E. Bassler, Phys. Rev. E **74**, 046106 (2006).

[106] B. Danila, Y.D. Sun, and K.E. Bassler, Phys. Rev. E **80**, 066116 (2009).

[107] R. Yang, W.X. Wang, Y.C. Lai, and G. Chen, Phys. Rev. E **79**, 026112 (2009).

[108] W.X. Wang, B.H. Wang, C.Y. Yin, Y.B. Xie, and T. Zhou, Phys. Rev. E. **73**, 026111 (2006).

[109] J. Gómez-Gardeñes and V. Latora, Phys. Rev. E **78**, 065102(R) (2008).

[110] K.I. Goh, B. Khang, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).

[111] Z.X. Wu, G. Peng, W.M. Wong, and K.H. Yeung, J. Stat. Mech. P11002 (2008).

[112] G. Yan, T. Zhou, B. Hu, Z.Q. Fu, and B.H. Wang, Phys. Rev. E **73**, 046108 (2006).

[113] B. Danila, Y. Yu, S. Earl, J.A. Marsh, Z. Toroczkai, and K.G. Bassler, Phys. Rev. E **74**, 046114 (2006).

[114] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E **70**, 056105 (2004).

[115] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, EPL **71**, 325 (2005).

[116] X. Ling, M.-B. Hu, R. Jiang, and Q.-S. Wu, Phys. Rev. E. **81**, 016113 (2010).

[117] D. De Martino, L. Dall'Asta, G. Bianconi, and M. Marsili, Phys. Rev. E **79**, 015101(R) (2009).

[118] D. De Martino, L. Dall'Asta, G. Bianconi, and M. Marsili, J. Stat. Mech. P080232 (2009).

[119] W. Huang, and T.W.S. Chow, Chaos **19**, 043124 (2009).

[120] W.-X. Wang, Z.-X. Wu, R. Jiang, G. Chen, and Y.-Ch. Lai, Chaos **19**, 33106 (2009).

[121] K. Kim, B. Kahng, and D. Kim, EPL **86**, 58002 (2009).

[122] G. Petri, H.J. Jensen, and J.W. Polak, EPL **88**, 20010 (2009).

[123] S. Floyd, V. Jacobson, IEEE/ACM Trans. Netw. **1(4)**, 397413 (1993).

[124] G. Bianconi, Phys. Lett. A **303**, 166 (2002).

[125] G. Bianconi, EPL **81**, 28005 (2008).

[126] B. Guerra, and J. Gómez-Gardeñes, Phys. Rev. E **82**, 035101(R) (2010).

[127] L. A. N. Amaral, A. Scala, M. Barthélemy, H. E. Stanley, Proc. Nat. Acad. Sci. USA **97**, 11149-11152 (2000).

[128] R. M. Anderson, R. M. May, *Infectious diseases of humans: Dynamics and Control*, Oxford University Press, Oxford (1992).

[129] A. L. Barabási, R. Albert, Science **286**:509-512 (1999).

[130] M.Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. **92**:178701 (2004).

[131] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA **101**:3747-3752 (2004).

[132] M. Boguña, C. Castellano, R. Pastor-Satorras, Phys. Rev. E **79**:036110 (2009).

[133] Caldarelli, G. *Scale-Free Networks*. Oxford University Press, Oxford (2007).

[134] C. Castellano, R. Pastor-Satorras, Phys. Rev. Lett. **96**:038701 (2006).

[135] C. Castellano, R. Pastor-Satorras, Phys. Rev. Lett. **98**:029802 (2007).

[136] C. Castellano, R. Pastor-Satorras, Phys. Rev. Lett. **100**:148701 (2008).

[137] M. Catanzaro, M. Boguña, R. Pastor-Satorras, Phys. Rev. E **71**:056104 (2005).

[138] Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., Faloutsos, C. ACM Trans. Inf. Syst. Secur. **10**(4):13 (2008).

[139] F. Chung, L. Lu, V. Vu, Proc. Natl. Acad. Sci. USA **100**:6313-6318 (2003).

[140] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, BMC Medicine **5**:34 (2007).

[141] V. Colizza, R. Pastor-Satorras, A. Vespignani, Nature Physics **3**:276-282 (2007).

[142] V. Colizza, A. Vespignani, Journal of Theoretical Biology **251**:450-467 (2008).

[143] D.J. Daley, J. Gani, *Epidemic Modelling*, Cambridge University Press, Cambridge (1999).

[144] S. Eubank, H. Guclu, V.S. Anil-Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, N. Wang, Nature **429**:180-184 (2004).

[145] Gallos, L. K., Argyrakis, P. Phys. Rev. Lett. **92**:138301 (2004).

[146] J.G. Gardeñes, V. Latora, Y. Moreno, E. Profumo, Proc. Nat. Acad. Sci. USA **105**:1399-1404 (2008).

[147] Grais, R. F., Ellis, J. H., Kress, A., Glass, G. E. Modeling the spread of annual influenza epidemics in the U.S.: The potential role of air travel. *Health Care Management Science* 7:127-134. (2004)

[148] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, Y. Moreno, Europhys. Lett. **89**:38009 (2010).

[149] R. Guimerà, S. Mossa, A. Turtschi, L. A. N. Amaral, Proc. Natl. Acad. Sci. USA **102**:7794-7799 (2005).

[150] M. Ha, H. Hong, H. Park, Phys. Rev. Lett. **98**:029801 (2007).

[151] J.-D. J. Han, N. Bertin, T. Hao et al, Nature **430**: 88-93 (2004).

[152] H. W. Hethcote, SIAM Review **42**:599-653 (2000).

[153] L. Hufnagel, D. Brockmann, T. Geisel, Proc. Natl. Acad. Sci. USA **101**:1512415129 (2004).

[154] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, Y. Aberg, Nature **411**:907-908 (2001).

[155] A. L. LLoyd, R. M. May, Science **292**:1316-1317 (2001).

[156] J. Marro, R. Dickman, *Nonequilibrium phase transitions in lattice models*. Cambridge University Press, Cambridge (1999).

[157] S. Meloni, J.G. Gardeñes, V. Latora, Y. Moreno, Phys. Rev. Lett. **100**:208701 (2008).

[158] S. Meloni, A. Arenas, Y. Moreno, Proc. Natl. Acad. Sci. USA **106**:16897-16902 (2009).

[159] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B **26**:521-529 (2002).

[160] J. D. Murray *Mathematical Biology*. Springer-Verlag, Germany, Berlin (2002).

[161] J. D. Murray, *Mathematical Biology*. Springer-Verlag, 3rd Edition (2007).

[162] M. E. J., Newman, S. Forrest, J. Balthrop, Phys. Rev. E **66**:035101 (2002).

[163] R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. **86**:3200-3203 (2001).

[164] R. Pastor-Satorras, A. Vespignani, Phys. Rev. E **63**:066117 (2001).

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

[166] J. G. Restrepo, E. Ott, B. R. Hunt, Phys. Rev. E **76**:056119 (2007).

[167] V. Rosato, S. Meloni, L. Issacharoff, F. Tiriticco, Physica A **387**:1689-1704 (2008).

[168] M. A. Serrano, D. Krioukov, M. Boguñá, Phys. Rev. Lett. **100**:078701 (2008).

[169] J. Shao, S. V. Buldyrev, L. A. Braunstein, s. Havlin, H. E. Stanley, Phys. Rev. E **80**:036105 (2009).

[170] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press, Oxford (1987).

[171] L. Zhao, Y.-C. Lai, K. Park, N. Ye, Phys. Rev. E **71**:026125 (2005).

[172] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, Cambridge, Massachusetts, and London, England, (2006).

[173] G. Szabó, and G. Fáth, *Phys. Rep.* **446**, 97 (2007).

[174] F.C. Santos and J. M. Pacheco, Phys. Rev. Lett. **95**, 098104 (2005).

[175] J. Gómez-Gardeñes, M. Campillo, L. M. Floría and Y. Moreno, Phys. Rev. Lett **98**, 108103 (2007).

[176] J. Poncela, J. Gómez-Gardeñes, L. M. Floría, and Y. Moreno, New J. Phys. **9**, 184 (2007).

[177] J. Vukov, G. Szabó, and A. Szolnoki, Phys. Rev. E **77**, 026109 (2008).

[178] V. M. Eguíluz, M. G. Zimmermann, C. J. Cela-Conde, and M. San Miguel, Am. J. Soc. **110**, 977 (2005).

[179] R. Jiménez, H. Lugo, J. A. Cuesta, and A. Sánchez, J. Theor. Biol. **250**, 475 (2008).

[180] F. C. Santos, J. M. Pacheco, T. Lenaerts, PLoS Comput. Biol. **2**(10), e140 (2006).

[181] J. Poncela, J. Gómez-Gardeñes, L. M. Floría, A. Sánchez and Y. Moreno, PLoS ONE **3**, e2449 (2008).

[182] V. Srivastava, J. Neel, A. B. MacKenzie, *et al.*, IEEE Communications **7**, 46 (2005).

[183] M. H. Vainstein, A. T. C. Silva, and J. J. Arenzon, J. Theor. Biol. **244**, 722 (2007).

[184] D. Helbing and W. Yu, Adv. Comp. Syst. **11**, 641 (2008).

[185] D. Helbing and W. Yu, Proc. Nat. Acad. Sci. USA **106**, 3680 (2009).

[186] A. Pusch, S. Weber, and M. Porto, Phys. Rev. E **77**, 036120 (2008).

[187] S. Assenza, J. Gómez-Gardeñes, and V. Latora, Phys. Rev. E **78**, 017101 (2008).

[188] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press, Cambridge, UK, 2004).

[189] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[190] F. Wang, Y. Moreno, and Y. Sun, Phys. Rev. E **73**, 036123 (2006).

[191] R. Guimera, A. Diaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).

[192] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E **70**, 056105 71 (2004).

[193] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Europhy. Lett. **71**, 325 (2005).

[194] S. Sreenivasan, R. Cohen, E. Lopez, *et al.*, Phys. Rev. E **75**, 036105 (2007).

[195] M. Argollo de Menezes and A. L. Barabási, Phys. Rev. Lett. **92**, 028701 (2004).

[196] M. Argollo de Menezes and A. L. Barabási, Phys. Rev. Lett. **93**, 068701 (2004).

[197] J. Duch and A. Arenas, Phys. Rev. Lett. **96**, 218702 (2006).

[198] J.D. Noh and H. Rieger, Phys. Rev. Lett **92**, 118701 (2004).

[199] J. Gómez-Gardeñes and V. Latora, Phys. Rev. E **78**, 065102(R) (2008).

[200] N.M Ferguson, Nature **466**:733 (2007).

[201] D. Brockmann, L. Hufnagel, T. Geisel, Nature, **439**, 462-465 (2006).

[202] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J.J. Ramasco, A. Vespignani, Proc. Natl Acad. Sci. **106**:21484 (2009).

[203] M.C.J. Bootsma , N.M. Ferguson, Proc. Natl Acad. Sci. **104**, 7588-7593 (2007).

[204] H. Markel, H.B. Lipman, J.A. Navarro, A. Sloan, J.R. Michalsen, A.M. Stern, M.S. Cetron, JAMA **298**, 6 (2007).

[205] W. Goffman, V.A. Newill, Nature **4955**, 225-228 (1964).

[206] J.M.Epstein, J. Parker, D. Cummings, R.A. Hammond, PloS ONE **3**, e3955 (2008).

[207] S. Funk, M. Salath, V. A. A. Jansen, J. R. Soc. Interface, in press (doi:10.1098/rsif.2010.0142) (2010).

[208] S. Funk, E. Gilad, C. Watkins, V.A.A. Jansen, Proc. Nat. Acad. Sci. USA **106**:6872-6877 (2009).

[209] M. J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, Princeton, 1st Edition, (2007).

[210] Mexican Authorities, *Boletín Mensual de Estadística Operacional. Secretaría de comunicaciones y transportes (2009).*

[211] V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, Proc. Natl. Acad. Sci, **103**, 2015-2020 (2006).

[212] V. Colizza, A. Vespignani, Phys. Rev. Lett. **99**: 148701 (2007).

[213] K. T. D. Eames, J. R. Soc. Interface **6**, 811814 (2009).

[214] F. C. Coelho, C. T. Codeo, PLoS Comput. Biol. **5**, 1000425 (2009).

[215] T. Gross, C. J. D. Lima, B. Blasius, Phys. Rev. Lett. **96**:208701 (2006).

[216] M. Catanzaro, M. Boguña, R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).