Contact Networks of Mobile Agents and Spreading Dynamics

Von der Fakultät Mathematik und Physik der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

vorgelegt von

Marta C. González
Lagunillas, Venezuela

Hauptberichter:	Prof. Dr. Hans J. Herrmann
Mitberichter:	Prof. Dr. Günter Haag
Tag der mündlichen Prüfung:	3. July, 2006

Institut für Computerphysik der Universität Stuttgart 2006

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISBN 3-8325-1318-3

Logos Verlag Berlin
Comeniushof, Gubener Str. 47,
10243 Berlin
Tel.: +49 030 42 85 10 90
Fax.: +49 030 42 85 10 92
INTERNET: http://www.logos-verlag.de

Publications related to this thesis

I M.C. González, A.O. Sousa and H.J Herrmann, *Opinion Formation on a Deterministic Pseudo-fractal Network. Int. J. Mod. Phys. C*, **15**, 45-57 (2004).

II M.C. González and H.J Herrmann, *Scaling of the propagation of epidemics in a system of mobile agents*, *Physica A*, **340**, 741-748 (2004).

III M.C. González, H.J Herrmann and A.D. Araújo, *Cluster size distribution of infection in a system of mobile agents* , Proceedings of Medyfinol in La Serena, *Physica A*, **356**, 100-106 (2005).

IV P.G. Lind, M.C. González and H.J Herrmann, *Cycles and clustering in bipartite networks* , *Phys. Rev. E*, **72**, 056127 (2005).

V M.C. González, A.O. Sousa and H.J Herrmann, *Renormalizing Sznajd model on complex networks taking into account the effects of growth mechanisms*, *Eur. Phys. Jour. B*, **49**, 253-257 (2006).

VI M.C. González, P.G. Lind and H.J Herrmann, *Model of mobile agents for sexual interactions networks*, *Eur. Phys. Jour. B*, **49**, 371-376 (2006).

VII M.C. González, P.G. Lind and H.J Herrmann, *System of mobile agents to model social networks*, *Phys. Rev. Lett.*, **96**, 088702 (2006).

VIII M.C. González, P.G. Lind and H.J Herrmann, *Networks based on collisions among mobile agents*, submitted to Physica D (2006).

IX M.C. González, H.J. Herrmann, J. Kertész and T. Vicsek, *Schools: Friendship networks: Community structure versus ethnic preferences*, to be submitted to *Proc. Natl. Acad. Sci. USA*, (2006).

X V. Schwämmle, M. C. González, A. A. Moreira, H. J. Herrmann and J. S. Jr. Andrade, *The spread of opinions in a model with different topologies* (2006), to be submitted (2006).

Contents

Chapter 1

Zusammenfassung

Ein Netzwerk stellt ein Gebilde aus Knotenpunkten dar, die miteinander verbunden sein können. Obwohl Graphen vielfach zeichnerisch dargestellt werden, sind sie 'nur' mathematische Strukturen. Dies bedeutet vor allem, dass verschiedene Visualisierungen denselben Graphen darstellen können. Die Netzwerktheorie ist ein interdisziplinäres Forschungsgebiet, das mit Hilfe mathematischer und statistischer Methoden Netzwerke aus unterschiedlichen Bereichen untersucht. Beispiele für komplexe Netzwerke sind das Internet, soziale Netzwerke oder die dynamische Ausbreitung von Krankheiten (Epidemiologie) sowie die metaboliche und regulatorische Netzwerke. Insbesondere soziale Netzwerke bringen oft Phänomene hervor, die mit Hilfe statistischer Modellen beschrieben werden können.

Die steigende Verfügbarkeit von Computerkapazität erlaubt die Sammlung und Analyse von Daten in einem Maßstab, der heute viel größer ist als früher. Obwohl die genannten Beispiele für Netzwerke ganz unterschiedliche Funktionen und Eigenschaften haben können, teilen sie eine ähnliche Struktur. Diese Tatsache suggeriert ein allgemeines organisierendes Prinzip, das hinter den Einzelkomponenten und ihren Wechselwirkungen steckt. Statistische Physik ist als Teilgebiet der Physik geeignet um Netzwerktheorie zu betrachteten. *Komplexe Netzwerke* sind normalerweise nüztlich zur Darstellung komplexer Systeme. Ein *komplexes System* wird je nach Autor und Wissenschaftsgebiet unterschiedlich definiert. Eine allgemeine Definition ist folgende[1]: Ein komplexes *System ist ein System, das die Beschreibungen seines Gesamtverhaltens in einer beliebi-*

[1] http://de.wikipedia.org/wiki/Komplexität

gen Sprache erschwert, selbst wenn man vollständige Informationen über seine Einzelkomponenten und ihre Wechselwirkungen besitzt. Die gemessenen Strukturen von komplexen Netzwerken können weder durch stark verbundene Gittermodelle noch durch Zufallsgraphen modelliert werden.

Neuere Erkenntnisse auf diesem Forschungsgebiet [1, 2] erlauben es, Modelle, die das Entstehen von realen Netzwerken mit besonderen Eigenschaften erklären sollen, empirisch zu überprüfen. Ein Teil dieser Dissertation betrachtet anhand der Entwicklung entsprechender Modelle zum Beispiel die Ausbreitungsdynamik von Systemen, die *Meinungsbildung* oder *Infektionen*. Wir stellen Modelle von Kontaktnetzwerken vor, die nicht ursprünglich auf der Graphentheorie basieren, sondern aus Agenten bestehen, die aufeinander wirken und durch die Netzwerke erzeugen. Das Netzwerk ist dann *ein Resultat der Wechselwirkungen zwischen den Knoten*. Um unsere Modelle zu testen, werden wir verschiedene empirische *soziale Netzwerke* betrachten.

Die Forschungsschwerpunkte dieser Dissertation werden in Folgenden getrennt vorgestellt:

Modelle zur Meinungsbildung

Wir entwickelten ein Modell zur Meinungsbildung eines auf Beobachtungen begründeten selbstähnlichen Gitters, welches eine überraschend gute Übereinstimmung mit den Wahlergebnissen gewisser Staaten (Brasilien, Indien) lieferte. Darauf vervollständigten wir auch noch dieses Modell theoretisch durch einen Renormierungsgruppenansatz. Das selbstähnliche Gitter und ein Beispiel der Ergebnisse dieser Rechnungen sind in Fig. 1.1 dargestellt (siehe z.B. Artikel I und V). Alle vorherigen Resultate, die auf selbstähnlichen Gittern formuliert sind, wurden mit Ergebnissen stochastischer skalenfreier Netzwerke verglichen. Die gleiche Art von Resultaten ist auch mit einem allgemeineren Meinungsbildungsmodell formuliert worden (siehe z. B. Artikel X).

Ausbreitung von Epidemien auf Netzwerken mobiler Agenten

Wir konstruierten einen völlig neuen Typ von Netzwerken, welcher auf der Bewegung von Agenten beruht. Wir studierten auf diesen Netzwerken insbesondere die Ausbreitung von Epidemien und entdeckten eine Reihe erstaunlicher Eigenschaften. Zwei bekannte Resultate der Literatur verbinden sich durch unseres Modell: Ausbreitung von Epidemien durch Mean Field-Näherung (MF) und

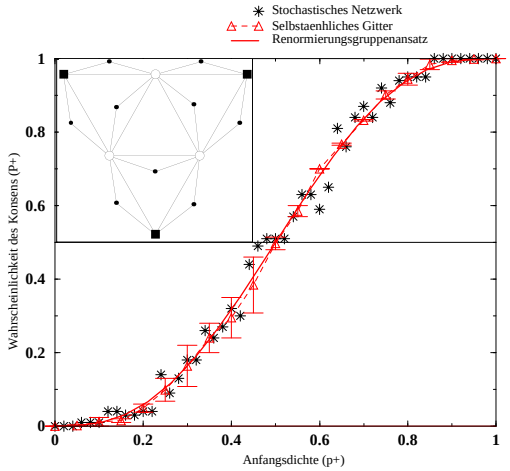

Figure 1.1: Vergleich: Renormierungsgruppenansatz (durchgezogene-Linie), Resultat der Simulation auf einem selbstähnlichen Gitter (Dreiecke mit Fehlerbalken) und auf einem stochastischen skalenfreien Netzwerk (Sterne). Inset: Beispiel eines selbstähnlichen Gitters aus drei Generationen.

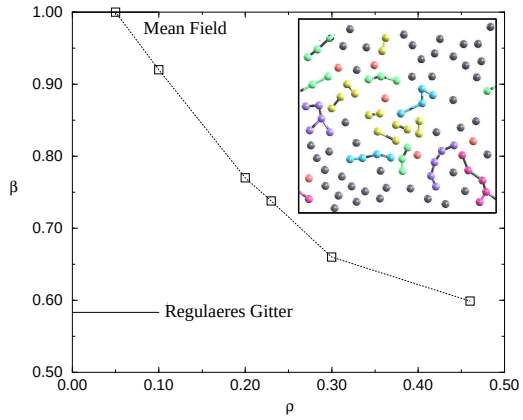

Figure 1.2: Numerischer Wert des kritischen Exponenten β der epidemischen Ausbreitung. Inset: Momentaufnahme des Clusters infizierter Agenten für ein System der Dichte $\rho = 0.20$

Modelle zur Ausbreitung von Epidemien auf regulären zweidimensionalen Gittern. Ein Beispiel des Modells und der Ergebnisse ist in Fig. 1.2 dargestellt (siehe auch Artikel II und III).

Modelle mobiler Agenten zur Bildung von sozialen Netzwerken

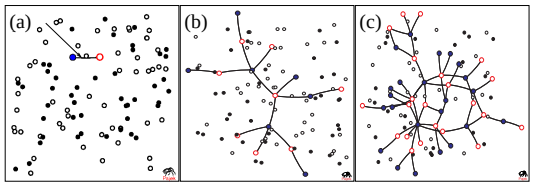

Figure 1.3: Momenteaufnahme des Systems von bewegten Agenten. Die Verbindungen zeigen Kollisionen. Gefüllte Knotenpunkte (Blau) und leere Knotenpunkte (Rot) stellen zwei verschiedene Arten von Knotenpunkte dar, z.B. verschiedene Geschlechter.

Wir haben das System von mobilen Agenten mit einer speziellen Kontaktregel und dem daraus entstehenden Kommunikationsnetz erweitert. Anschließend haben wir die Eigenschaften von diesen Netzwerken genauer untersucht und Ähnlichkeiten zu empirischen Daten sowohl bei sexuellen Kontakten als auch bei Freundschaften an Schulen gefunden. Der Arbeit zu sexuellen Kontakten hat erstmalig verschiedene topologische Eigenschaften von empirischen Netzwerken reproduziert (siehe Artikel IV und VI). Die Arbeit über Freundschaftnetzwerke wurde mit einer sehr großen Datenbank vergleichen und es entstanden sehr aufschlussreiche Querverbindungen (siehe Artikel VII und VIII). Beispiele der sexuale Netzwerke des Modells sind in Fig. 1.3 dargestellt.

Gemeinschaftsstruktur und die Rolle von Merkmalen in Freundschaftnetzwerken

Es wurde erfolgreich eine Methode zur Auswertung der Schulnetzwerke entwickelt (siehe Artikel IX). Es wurden *Netzwerke von Gemeinschaften* aus Freundschaftnetzwerken extrahiert, die nicht-triviale Struktur haben. Diese sind skalenfreie Netzwerke und zeigen zudem starke Korrelationen. In diesem Kontext, beschreiben Korrelationen ethnischen Gruppen. Unsere Methode ermöglicht die

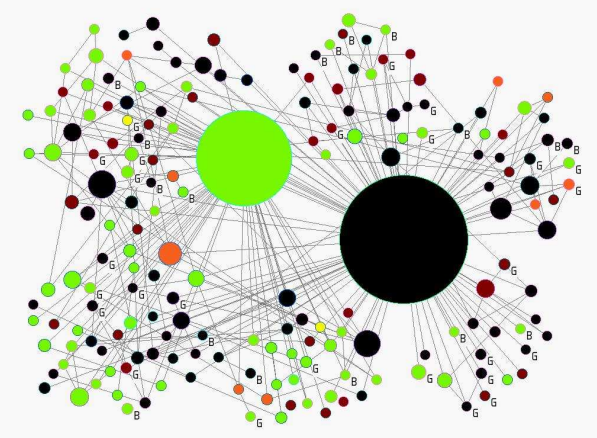

Figure 1.4: Netzwerk der Schulgemeinschaften. Die Farbe wird der ethnischen Mehrheit in der Gemeinschaft zugewiesen. Die Knotengrosse ist proportional zur Anzahl der Knoten in der Gemeinschaft.

Quantifizierung und den Vergleich dieser Segregation für verschiedene Schulen. Ein Netzwerk von Gemeinschaften einer Schule ist in Fig. 1.4 dargestellt.

Die hier vorgestellten Modelle bieten einen Ansatz zur die Untersuchung komplexer Netzwerke, die aus den Kontakten von mobilen Agenten entstehen. Die gefundenen Eigenschaften können auf die Anwesenheit von Gemeinschaften zurückgeführt werden. Der vorgestellte Ansatz präsentiert eine Verbindung zwischen der granularen Gastheorie und der Modellierung komplexer Netzwerke. Im quasi-stationären Fall ist es möglich, verschiedene Kollisionsregeln granularer Gase zu übernehmen. Eine mögliche Fragestellung für zukünftige Forschung, unter Verwendung der vorgestellten Methode, wäre: *Wie kann ein Netzwerk charakterisiert werden, das anhand von Kollisionsregeln gebildet wird, die aus dem Schwarmverhalten von Fischen, Vögeln oder Insekten hergeleitet wurden?*

Es ist wichtig zu betonen, daß das Ziel unserer Methode die Entwicklung eines möglichst einfachen Modells ist, welches sich auf den Aggregatzustands und nicht auf die Ebene der Komponenten konzentriert. In diesem Sinne ist unser Modell idealisiert und vielleicht ein anderer Ansatz gengenüber *Agentenbasierter Modelle*, die sich mit dem individuellen komplexen Verhalten - z.B. *Lernen und Anpassung* - beschäftigen. Die hier vorgestellten mobilen Agenten beinhalten Elemente der granularen Gastheorie um komplexe Netzwerke zu

modellieren.

Chapter 2

Introduction

A network is a set of entities referred as vertices or nodes pairwise connected among them through edges or links. Systems taking the form of networks are all around us. Examples include the Internet, the World Wide Web, social systems of acquaintances like friendship networks, infrastructure systems like the airport networks, biological systems like metabolic networks, among many other examples [1, 2]. The nodes represent the components of a given system, e.g., computer severs, web-pages, people, airports, molecules; and the links represent existing interactions among them, e.g., physical connections, virtual links, friendships, flights, chemical reactions.

The increasing availability and capacity of modern computers has allowed the collection and analysis of data on a scale far larger than previously possible. While the listed networks examples are totally different from each other in their function and attributes, they share a similar structure, suggesting general and common organizing principles beyond the specific details of the individual systems. In this context, a statistical physics approach has been exploited as a very convenient method because of its deep connection with graph theory and because of its ability to quantitatively characterize macroscopic phenomena in terms of the microscopic dynamics of the various systems. The concept "complex networks" refers to a network that is often the representation of a complex system [1]. So by studying the network one can study the underlying complex system itself. Complex networks contain much randomness but they contain certain structure

[1] A commonly accepted definition of *complex system* is a system consisting of many interacting units whose collective behavior cannot be explained from the behavior of the individual units alone.

that cannot be modeled as purely random graphs.

After some recent intensive research on the statistical properties of complex networks important open questions remain. One aspect of our research focuses on the development of appropriate models of spreading dynamics, such as diffusion models of opinion formation and spreading of infections. Particularly, we present models of contact that differ from the traditional static graph modeling that is at the core of classical graph theory: We incorporate underlying agent interactions that generate dynamical contact networks. The network itself is the result of the interactions among nodes.

Motivated by the close relations of the field with real world measurements, we collect and incorporate in the context of Statistical Physics literature new sets of empirical networks in order to test and validate the models here proposed. The contents of the chapters and the main achievements in each of them can be summarized as follows:

Modeling opinion dynamics on hierarchical networks

Stochasticity is a common feature of all network models that resemble real world topologies. That is, nodes are connected using probabilistic rules. This randomness present in the models makes it harder to gain a visual understanding in how all the nodes relate to each other. There are some models that successfully lead to realistic topologies and are constructed in a deterministic fashion [3, 4]. It is of major theoretical interest to solve the mechanisms of spreading dynamics in deterministic networks.

In Chapter 4 we adapt a recognized model of opinion formation and use it to simulate elections processes on a hierarchical network. We obtain vote distributions which resemble those observed in real elections in Brazil and India (published in paper I).

Further, we take advantage of the hierarchical structure of the deterministic network and provide analytical expressions for the case of two opinions. The obtained expressions reproduce very well the results of the simulations on random complex networks (as published in paper V).

In a third study in this area, we develop a continuum model for opinions. In contrast to Ising-like models where the opinion of an agent is represented by a integer number, in our proposed model an agent can carry n opinions at a time, represented by a vector with positive real components in an n-dimensional

space. We show that even in such a sophisticated model of opinion, the spreading mechanisms that lead to opinion formation are the same on either a deterministic or on a random complex network (as presented in paper X). The obtained results imply the possibility of using hierarchical networks for an analytic solution of the model that, as before, would predict well the dynamics on random complex networks.

Spreading dynamics in a system of mobile agents

Previous advances in modeling spreading dynamics consist in uncovering the influence of the network topology on the spreading dynamics. These studies which constitute an important advance in the understanding of spreading mechanisms, are at the same time, a static picture of the process.

In Chapter 5 we characterize the mechanisms of infection spreading in a system of mobile agents and show that compared to the static case, a new set of critical exponents governs the dynamics of infection (as published in paper II).

As a second study in the field, we calculate the size distributions of the spatio-temporal clusters of infected agents, finding the relations among the characteristic cluster size and the total number of infected agents (published in paper III).

Dynamical model of growing social networks

Motivated by the mentioned influence of the mobility on the spreading mechanisms, in Chapter 6 we develop a contact network of mobile agents that successfully resembles empirical networks of sexual contacts.

We present, for the first time in the literature, a model that is able to reproduce many topological features of the empirical sexual networks, testing it with respect to a null model given by the randomized version of the empirical networks (see paper VI).

For the appropriate structural analysis and comparison of the network model with the empirical data, it was necessary for us to develop a suitable analytical expression that measures the probability of occurrence of *cycles of even length* in networks. Our proposed expression appears to be particularly useful in the characterization of bipartite networks, like is the case of real data from heterosexual contact networks (published in paper IV).

Dynamical model of stationary social networks

One important application of complex networks are the friendship relations among people. Using a model of mobile agents we studied the friendships forming by collisions and found very convincing agreement with studies in American schools concerning 90118 pupils.

We present all the details of this model in Chapter 7 (as published in papers VII and VIII).

Community structure and and role of attributes in friendship networks

In Chapter 8 we propose a novel method of analysis that uncovers organizational aspects of the large data set of friendship networks.

Our study is able to give a new insight into the data concerning the detection of communities in the friendship networks. The role of weights and attributes are shown to be of particular importance in the analysis of the collected empirical data. We develop a method that is able to quantify racial segregation in the studied empirical networks (as presented in paper IX).

Additionally in Chapter 3 we present some basic concepts and the most promising advances already proposed in the field. We divide it in four sections about static measures of networks, examples and structure of social networks, proposed models of networks and spreading dynamics. Chapter 9 ends this thesis presenting the main conclusions and discussing the results of this study.

Chapter 3

Basics

3.1 Networks measures

In this section we present a set of measures that characterizes the topological structure of networks. The concepts presented here are basic tools for the appropriate evaluation of the developments presented along the rest of this work.

3.1.1 Definitions and notations

A network is a set of items, called nodes or vertices, with some connections between them called links or edges. Graph theory [5, 6] is the natural framework for the exact mathematical treatment of complex networks. A Graph $G = (\mathcal{N}, \mathcal{L})$ consists of two sets $\mathcal{N}$ and $\mathcal{L}$, such that $\mathcal{N} \neq \emptyset$ and $\mathcal{L}$ is the set of pairs of elements of $\mathcal{N}$. The number of elements of $\mathcal{N}$ and $\mathcal{L}$ are denoted by N and K, respectively. A node is usually referred by its index i in the set $\mathcal{N}$. Each of the links is defined by a couple of nodes (i, j), and is denoted as e_{ij}.

Single nodes joined by links is the simplest type of networks (Fig. 3.1a). For instance, there may be more than one type of node in a network, or more than one type of link (Figs. 3.1b-e). Nodes or links may have also different types of properties associated with them. Taking the example of a social network, the nodes may represent men or women, people of different races, age, or many other things. Links may represent friendship, but also could represent animosity, or a

professional acquaintance. They can carry weights, representing e.g. the time two people know each other. They can also be directed, pointing in only one direction; like a network representing telephone calls. Networks may be also partitioned in different ways. We will see here examples of bipartite networks: networks that contain nodes of two different types with links running only between unlike types of nodes. Along this work, we will find examples of each of the variations of the networks previously described.

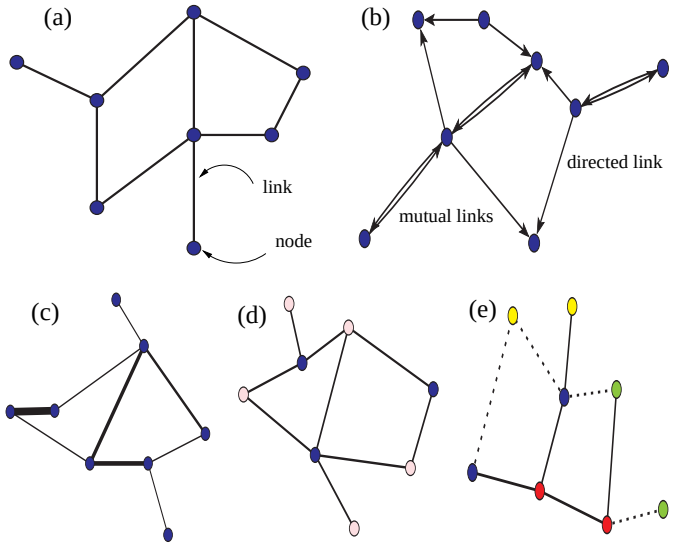

Figure 3.1: Examples of types of networks: (a) An undirected network with only a single type of node and a single type of link. (b) A directed network, in which each link has a direction. (c) A weighted network: Different link weights. (d) A bipartite network: Connections are present only between two type of nodes. (e) A network with various node and link types.

3.1.2 Node degree, degree distributions, and degree correlations

The *degree* k_i of a node i is the number of edges incident with the nodes [7]. If the network is directed, the degree of the nodes has two components: the number of outgoing links k_i^{out} (out degree) and the number of ingoing links k_i^{in} (in degree). The total degree is then defined as: $k_i = k_i^{out} + k_i^{in}$.

The basic topological characterization of a network is given by the *degree distribution* $P(k)$, defined as the fraction of nodes in a network having degree k. The nth-moment of $P(k)$ is defined as

$$\langle k^n \rangle = \sum_k k^n P(k). \tag{3.1}$$

The first moment $\langle k \rangle$ is the mean degree of the network. The second moment $\langle k^2 \rangle$ measures the fluctuations of the degree distribution. The divergence of $\langle k^2 \rangle$ in the limit of infinite network size, radically changes the behavior of dynamical process that take place over the network as it will be shown in section 3.4.1.

A large number of real world networks are *correlated* [9], in the sense that the probability of a node of degree k to be connected to another node of degree k', depends on k. In these cases $P(k)$ does not fully characterize the network structure and it is necessary to introduce the *conditional probability* $P(k'|k)$, being defined as the probability that a link from a node with degree k points to a node with degree k'. The direct evaluation of $P(k'|k)$ gives extremely noisy results for most of the real world networks because of their finite size. This problem can be overcome by defining the *average nearest neighbors degree* of nodes with degree k, as:

$$k_{nn}(k) = \sum_{k'} k' P(k'|k). \tag{3.2}$$

If there are no degree correlations, Eq. 3.2 gives $\langle k^2 \rangle / \langle k \rangle$, i.e. $k_{nn}(k)$ is independent of k. Correlated graphs are classified as *assortative* if $k_{nn}(k)$ is an increasing function of k, whereas they are referred as *disassortative* when $k_{nn}(k)$ is an decreasing function of k [10]. In other words, in assortative networks the nodes tend to connect to their degree peers, while in disassortative networks nodes with low degree are more likely connected with highly connected ones.

3.1.3 Shortest path length and betweenness

A measure of the typical distance between two nodes in a graph is given by the *average shortest path length*, defined as the mean shortest distance (d_{ij}) over all couples of nodes [7]:

$$\ell = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N}, i \neq j} d_{ij}. \tag{3.3}$$

Note that ℓ diverges if there are disconnected components in the graph, to avoid this divergence the summation is limited to the largest connected component.

The communication of two non-adjacent nodes, say j and k, depends on the nodes belonging to the paths connecting j and k. In consequence, a measure for the relevance of a given node can be obtained counting the number of geodesics going through it, which defines the so-called *node betweenness*. Together with the degree, the betweenness is one of the standard measures of the *node centrality*, originally introduced to quantify the importance of an individual in a social network [8]. The *betweenness* b_i of a node i, also referred as *load* [8], is defined as:

$$b_i = \sum_{j,k \in \mathcal{N}, j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \tag{3.4}$$

where n_{jk} is the number of shortest paths connecting j and k, while $n_{jk}(i)$ is the number of shortest paths connecting j and k and passing through i. Two commonly used algorithms to find shortest paths are the Dijkstra algorithm and the burning method [11, 12].

3.1.4 Clustering

Clustering is a measure for the probability that two neighbors of a node are connected between each other. In a network, high clustering means the presence of a high number of triangles. More precisely, c_i the *local clustering coefficient* of a node i, is a measure introduced by Watts and Strogatz in Ref. [13], and is defined as follows:

$$c_i = \frac{2e_i}{k_i(k_i - 1)}. \tag{3.5}$$

It is obtained by counting e_i the actual number of links connecting any two neighbors of i divided by $k_i(k_i - 1)/2$, the maximum possible number of links among all the neighbors of i. The clustering coefficient of the network is then given by the average of c_i over all the nodes in the network:

$$C = \langle c \rangle = \frac{1}{N} \sum_{i \in \mathcal{N}} c_i. \tag{3.6}$$

It is also useful to consider $c(k)$, which is the average local clustering c_i taken over the nodes with degree k, defined as [14]:

$$c(k) = \sum_{k',k''} P(k', k''|k) p_{k'k''}, \tag{3.7}$$

where the function $p_{k'k''}$ is the probability that nodes k' and k'' are connected. Note that $c(k)$ is a measure for three nodes correlations. Its recent calculation on real world networks has been of particular interest, because it is used to study the level of *hierarchy* and *modularity* in real complex networks [15].

3.1.5 Motifs

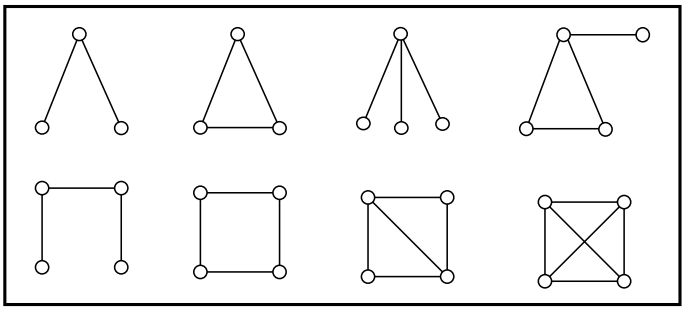

Figure 3.2: Examples of subgraphs analyzed for motifs detection. Shown are the two types of three-node and the six types of four-node non directed subgraphs.

A *motif* M is a *pattern* of interconnections occurring in a graph G *significantly* more often than in *randomized versions* of the graph [7]. *Randomized versions*, also called *null models*, are graphs generated with the same number of nodes, links and degree distribution as the original one, but where the links are re-wired at random. A *pattern* of interconnections is a n-node subgraph of G. As an example, the 2 possible types of 3-node and the 6 possible types of 4-node undirected subgraphs are illustrated in Fig. 3.2.

The statistical significance of M is then described by the Z-score, defined as [16, 17]:

$$Z_M = \frac{n_M - \langle n_M^{rand} \rangle}{\sigma_{n_M}^{rand}}, \tag{3.8}$$

where n_M is the number of times the subgraph M appears in G, and $\langle n_M^{rand} \rangle$ and $\sigma_{n_M}^{rand}$ are, respectively, the mean and standard deviation of the number of appearances in the randomized network ensemble.

The concept of *motifs* was originally introduced in Refs. [16, 17] by Alon and co-workers. They analyzed a broad spectrum of networks: from biochemistry, neurobiology, ecology to engineering. They found that specific motifs are associated to different types of networks. These measures have given a particular understanding on the **local** structure and functionality related to each type of network, and it represents an important complement to the **global** characterization of a network measured by the degree distribution.

In the last chapter of this thesis we show that in real world networks *motifs* can emerge not only looking at the topological patterns but also paying attention to the colors or *attributes* of the nodes. As remarked by Stauffer et. al. in his recent book on interdisciplinary Physics [49]: "*More than two thousand years ago Empedocles observed that some groups of people are like wine and water, mixing well, while others are like oil and water, mixing badly*". Based on the methods presented in this section, we quantify the level of ethnic or racial preferences in a large data base of schools friendship networks in U.S.A.

3.1.6 Community structures

The notion of community structure in networks and the first formal concepts of it have been proposed in the social sciences [8]. Given a graph G, a *community* is a subgraph G' whose nodes are tightly connected, i.e. a cohesive subgroup. Since the structural cohesion of the nodes of G' can be quantified in different ways, there are different formal definitions of community structures. Next, we illustrate two widely accepted concepts of community together with two different methods of detection of community structure recently proposed.

Betweenness algorithm

The concept of community structure is based on the frequency of links: *Communities* are group of nodes within which node-node connections are dense and between which connections are sparser. A schematic example of a network with such a community structure is shown in Fig. 3.3.

The method presented in this section was developed by Newman and coworkers [18]. It is based on identifying those links that are more likely to be *between* communities. Groups of communities are then identified removing those links from the network. The method is based on the concept of betweenness presented

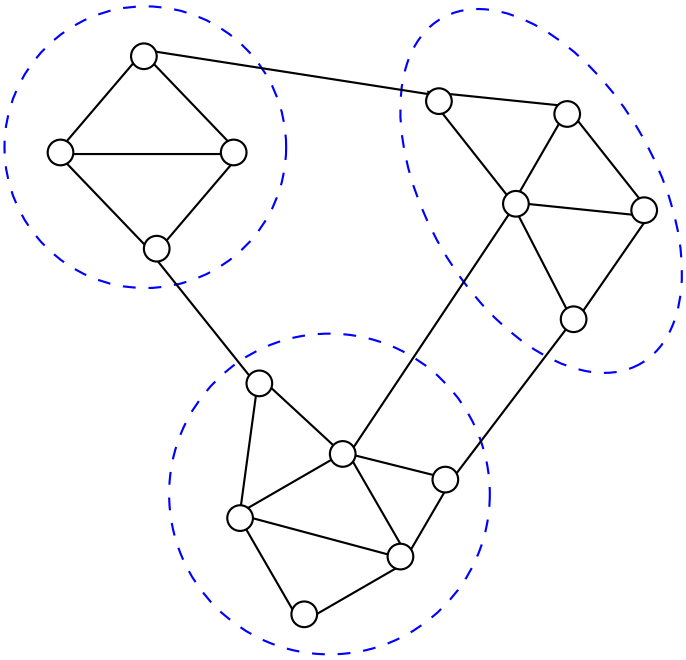

Figure 3.3: Schematic representation of a network with community structure. In this network there are three communities of densely connected nodes (enclosed by circles), with a much lower density of connections between them. By construction, the communities are also examples of 3-cliques of different sizes (see Sec. 3.1.6).

in section 3.1.3 generalized to the links. *Link betweenness* is the fraction of shortest paths between pairs of nodes that runs along a link. If there is more than one shortest path between a pair of vertices, each path is weighted by the inverse of the existing number of paths, such that the sum of all the paths is unity. Those links connecting communities will have high betweenness. By removing these links, groups are separated and so the community structure of the graph is revealed. The procedure can be halted at any point, and the resulting components in the network are taken to be the communities. The entire progression of the algorithm from top to bottom, can be represented in the form of a tree such as shown in Fig. 3.4.

In order to evaluate if the divisions are meaningful, at each level of the tree the

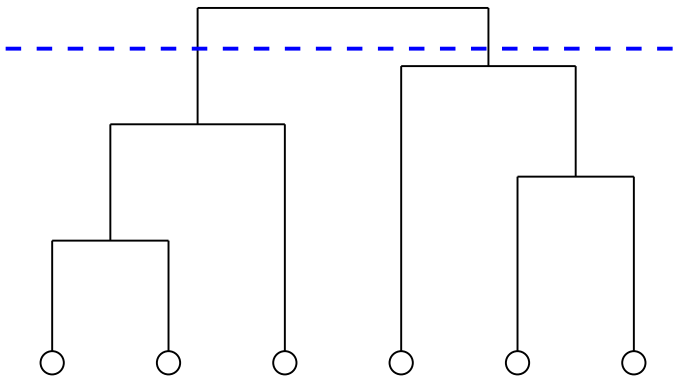

Figure 3.4: An example of a small hierarchical clustering tree. The circles at the bottom of the figure represent the nodes in the network. From top to bottom, the tree shows the sequential split of the network. The shape at the bottom denotes two communities in the graph, which corresponds to a peak in the calculated *modularity* at the level shown by the dotted line.

modularity m for this division is calculated, defined as [19]:

$$m = \sum_i (e_{ii} - a_i^2), \tag{3.9}$$

where e_{ij} denotes the fraction of ends of links in group i for which the other end of the link lies in group j, and $a_i = \sum_j e_{ij}$ is the fractions of all ends of links that lie in group i. Modularity measures the difference between the total fraction of edges that fall within -rather than between- groups (the first term) and the fraction one would expect if edges were placed at random (respecting node degree). Thus, high values of the modularity indicate divisions of the network in which more of the edges fall within groups than one would expect by chance. This indicates a strong significant split in the network and has been found to be a good indicator of functional network divisions in many cases [20]. The best division in the process is the one which gives the maximum modularity.

Clique percolation method

A stronger definition of community requires that all pairs of community members choose each other. Such a requirement leads to the definition of k-*clique*: a fully connected subgraph of k nodes. Weakening the requirement of adjacency by

the requirement of reachability, Vicsek and collaborations [21], introduced the concept of k-*clique community* as a union of all k-cliques that can be reached from each other through a series of *adjacent* k-cliques. *Adjacency* here means sharing $k-1$ nodes. As an example, the communities of Fig. 3.3 are examples of groups from adjacent 3-cliques or 3-*clique communities* of different sizes: 4, 5 and 6.

A particular feature of this method, in contrast to the other methods based on network divisions as e.g. the one presented previously, is that each node can be *member* of different communities. It allows the identification of interesting features in the *overlaps* or number of nodes belonging to different communities. Thus, it is possible to analyze *networks of communities*, where the nodes are the k-clique communities and the links are the members they share.

Although different numbers of k might be optimal for the local community structure around different nodes, there is a global criterion to fix its value for a given network: k is selected such that it gives the highest structure of communities possible from an analyzed network. In other words, in the related percolation phenomena [22] a giant component appears when the number of links are increased above some critical point. The smallest value of k is selected for which no giant community appears. In this way, it is possible to find as many communities as possible, without the negative effect of having a giant community that smears out the details of the community structure by merging many smaller communities.

This method is more accurate but consumes much more CPU time than the one described previously. The running time of the algorithm based on betweenness scales with system size as $O(N^3)$, much better than $O(expN)$ based on the clique percolation method. Recent algorithms based also on the concept of modularity have been proposed and scale as well as $O(n^2 \log n)$ [23, 24].

In chapters 7 and 8 we use this method of community detection to characterize the community groups that emerge from our proposed model of mobile agents and to compare it with the community structure observed in measured social networks.

3.1.7 Summary

In this section we presented the basic concepts and the measurement tools that have been developed in the last few years for the study and characterization of complex networks. To summarize, in order to compare network models with real world networks based on solid statistical grounds various levels of analysis must be fulfilled. These are: degree distribution, degree correlations, clustering and

community structure. Additionally, the identification of significant local patterns can be made through comparisons with appropriate null models of the network.

3.2 Networks in the real world

A recent boom of works from statistical physics to the study of networks has been largely driven by observations of the properties of *real world networks* and attempts to characterize and model them. There is data coming from different kinds of networks, from different branches of science: Social networks, information networks, technological networks and biological networks. In many studies they have been examined simultaneously and it is well established that, from the statistical point of view, they share common mathematical properties [1, 2, 13]. In the following, we describe this commonly observed mathematical structure; but we put particular emphasis on examples coming from social networks, which are the kind of systems particularly studied along this research.

3.2.1 Examples of social networks

The academic discipline with the longest history in the quantitative study of real world networks are the social sciences: Among the early works on the subject are the work of Jacob Moreno in the 1920's and 30's on friendship patterns within small groups [25]. In more recent years, studies of business communities [26] and of sexual patterns [27] have attracted particular attention. Reference [2] is a review directed to the statistical physics community and at the same time contains a rich recompilation of references from the social sciences.

We mention here only some of the most cited examples of social networks for which their statistical properties have been studied. There are examples from affiliation networks, e.g. the collaboration network of film actors, where two actors are connected if they have appeared together in a film [13, 28]; company directors networks [26], two directors are linked if they belong to the same board of directors; or the scientific Co-authorship networks in which individuals are linked if they have Co-authored one or more papers [29]. Another source of reliable data are communication records of certain kind, for example the network of telephone calls [30] and e-mails communications [31]. Still, the most used method of data recompilation for social networks, are the traditional questionnaires or direct interviews, that have been used, for example, to build networks of friendship and acquaintances [32].

In the next section we present some of the common topological structures measured in real world networks.

3.2.2 Basic structures observed

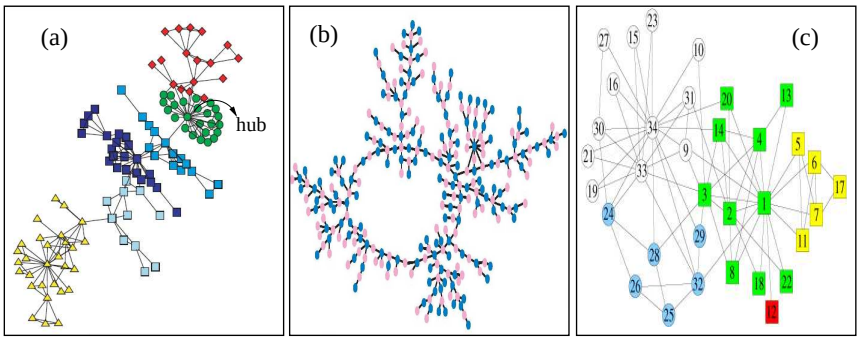

Figure 3.5: **(a)** An example of a small Co-authorship network among scientists at a private research institution (figure from Ref. [36]). (b) Network of sexual and romantic relations from adolescents in a high school (from Ref. [27]). (c) Network of friendships between individuals in the karate club study of Zachary [37](figure from [38]).

In d-dimensional lattices, the mean number of nodes one has to pass in order to reach an arbitrary chosen node, grows with the lattice size as $N^{1/d}$. In contrast, in most of the real networks, despite of the large size, there is a relatively short path between two nodes, that depends, at most, logarithmically on the network size; this property is known as *small-world effect*. It was first investigated by Milgram in the 1960s in a series of experiments to estimate the actual number of steps required for a letter passing person to person to reach an arbitrary target, it was reported to be in average only six steps [2].

The small diameter is a characteristic of networks with *randomly* linked nodes. As consequence of their randomness, the so called *random graphs*, have small diameter and small clustering coefficient (Eq. 3.6). But in contrast to random graphs, most real world networks have large values of the clustering coefficient. This paradoxical observation was addressed in the seminal work by Watts and Strogatz [13], who proposed the first model of *small-world networks*, having small diameter a high clustering.

Further studies of the statistical properties of diverse real world networks have

shown the existence of three classes of small-world networks, resumed by Amaral et. al. as follows [33]:

- Scale-free networks, characterized by a degree distribution that decays as a power law: $P(k) \sim k^{-\gamma}$.
- Broad-scale networks, characterized by a degree distribution that has a power law-regime followed by a sharp cutoff: $P(k) \sim k^{-\gamma} e^{-k/\kappa}$.
- Single-scale networks, characterized by degree distributions with a fast decaying tail: $P(k) \sim e^{-k/\kappa}$.

A natural question emerges "what are the reasons for such a rich range of possible structures?". A crucial advance in this explanation was given by Barabási and Albert [34] who proposed the first model for scale-free networks and showed that they emerge in the context of growing networks in which new nodes connect preferentially to the more highly connected nodes in the network, called *hubs*: "rich get richer".

In this context, there is also a possible explanation [33] for the appearance of a *power-law with cut-off* and the *single-scale* degree distributions. That is, that the preferential attachment can be hindered by two classes of factors, the *aging* of the nodes, i.e. with time the nodes stop making new links; and the *cost* associated to a limited capacity of the nodes for making links.

We illustrate with examples of social networks the three kinds of degree distributions mentioned above. An example of *scale-free* social network are the Co-authorship networks [29] for scientific contributions. This is clearly a growing network where the more prestigious researchers attract the larger number of new ones, e.g. students. In reality, the aging effect, of course must be present, but according to the observations, it seems to be low enough and allows the formation of a power-law degree distribution in most of the studied cases [1]. On the other hand, friendship networks are observed to have a strongly peaked or *single-scale* degree distribution. Here the explanation associated with *cost* applies well for this kind of network: There is a recurrent cost in terms of time and effort for maintaining a friendship, giving those limited resources, people can only maintain a certain number of friends, these networks are *single scale* [35]. Indeed, other relations that may imply little cost or 'one-time' cost to increase the degree, is the case of sexual contacts which are observed to be *broad-scale*.

[1] Among all the sample of measured Co-authorship networks there some that are *broad-scale*

Just for illustrative purposes, we display in Fig. 3.5 three small real networks of the kind mentioned before. Fig. 3.5a is from Ref. [36] and is an example of a small Co-authorship network at a private research institution. Nodes represent scientists and a line between two nodes indicates that they have Co-authored a paper. The shape of the nodes correspond roughly to different topics of research. This network has a *power-law* degree distribution. Fig. 3.5b represents the romantic and sexual network among adolescents in a high-school, it is from the research conducted by Moody and co-workers [27]. Its degree distribution has a *power-law* region with a cut-off. In contrast, Fig. 3.5c is a friendship network among members of a Karate Club, which later divided into two clubs, which are marked by two different symbols. This belongs to the group of early networks analyzed in social sciences, the research was conducted by Zachary [37], this small network is today a commonly used network to test methods of detection of community structure; the figure is reproduced from Ref. [38]. The degree distribution is peaked or *single scale*.

3.2.3 Some particularities

According to recent observations, social networks share characteristics which are fundamentally different from other types of networked systems [39]. It has been observed that the degrees of adjacent nodes are positively correlated in social networks but negatively correlated in most other networks: Social networks are assortative (see section 3.1.2). There is also some evidence that the clustering is far larger than expected by chance; while the level of clustering in non social networks does not differ from the clustering of its randomized versions; i.e. it depends mainly on the given degree distribution. Additionally social networks are divided into groups of communities. Some of them grow in time e.g. Co-authorship networks, and others remain roughly fixed in size, e.g. friendship networks at the work place.

The appropriate modeling of all these characteristics with a minimum number of parameters has been an open question. In one attempt, Newman and coworkers in Ref. [39] propose that community formation is the mechanism responsible for the presence of high clustering and of assortativity. In their model, they propose a priori for the distributions of group membership and the distribution of sizes of groups, together with a parameter p for link formation within a group and a generating function formalism, they produce networks which are assortative and with high clustering.

In Chapter 6 we present an alternative approach for growing social networks

based on mobile agents; the results compare favorably in terms of degree distribution and clustering distribution or three point degree correlations to real social networks and their randomized versions. With the same approach in its quasi-stationary version, i.e. non-growing networks, in Chapter 7 we are able to reproduce the observed degree correlation and community structure without labeling a priori the groups each node belongs to. Our model constitutes, until now, a successful approach to dynamical social networks with the least number of adjusting parameters.

3.3 Networks Models

This section focuses on the mathematical modeling of networks. We present here some of the more cited models of networks. Further details of the models can be found e.g. in Ref. [7].

3.3.1 Random graphs

The systematic study of random graphs was initiated by Erdös and Rényi (ER) in 1959. Starting with N disconnected nodes, *ER random graphs* are generated by connecting couples of randomly selected nodes with probability $0 < p < 1$. Graphs with K links will appear in the ensemble with probability $p^K(1-p)^{N(N-1)/2-K}$. In the limit of large $N \to \infty$, the graph has fixed average degree $\langle k \rangle = p(N-1)$.

The structural properties of ER random graphs vary as a function of p, they present a critical phase transition at $p_c = 1/N$, corresponding to a critical average degree $\langle k \rangle_c = 1$:

- If $p < p_c$ the graph has no component of size greater than $O(\ln N)$, and no component has more than one cycle.
- If $p = p_c$ emerges a largest component of size $O(N^{2/3})$.
- If $p > p_c$ the graph has a component of $O(N)$ with a number $O(N)$ of cycles, and no other component has more than $O(\ln N)$ and more than one cycle.

The transition is of second order and belongs to the universality class of mean field percolation. The degree distribution is well approximated by a Poissonian distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}. \tag{3.10}$$

ER random graphs are uncorrelated graphs, $P(k|k')$ and $k_{nn}(k)$ are independent of k. The average shortest-path varies as $\ell \sim \ln N / ln\langle k \rangle$ and the clustering coefficient is equal to $C = p = \langle k \rangle / N$ [7]. Hence ER random graphs have vanishing C in the limit of large system size.

3.3.2 Small world networks

The model presented by Watts and Strogatz (SW) is a method to construct graphs that have the small-world property and high clustering coefficient [13]. It is started with a ring of N nodes in which each node is symmetrically connected to its $2m$ nearest neighbors for a total of $K = mN$ edges. Then, for every node, each link that is connected to a neighbor is rewired to a randomly chosen node with probability p, and preserved with probability $1 - p$. Note that for $p = 0$ we have a regular lattice, while for $p = 1$ we have a random graph with the constraint of minimum degree $k_{min} = m$.

Increasing p slightly above zero, the shortest path length ℓ, decreases non linearly and depends logarithmically on the network size. Conversely, $C(p)$ decreases much slower, at most linearly with p. This leads to a region of values of p, where one has both small path length and high clustering, which is an observed property in real world networks. As for the degree distribution, when $p = 0$ it is a delta function centered in $2m$, while for $p = 1$ it is similar to that of a ER random graph.

3.3.3 Scale-free networks

The Bárabasi-Albert model (BA) is a model of growth inspired by the formation of the World Wide Web. The basic idea is that in the World Wide Web, sites with high degree acquire new links with higher rates than low degree nodes. Starting with $m + 1$ nodes all connected among them, at each time step, $t = 1, 2, ..., N - (m + 1)$, a new node j with m links is added to the network. The probability that node j connects to an existing node i is linearly proportional to

the actual degree of i:

$$P_{j\to i} = \frac{k_i}{\sum_l k_l} \tag{3.11}$$

Because every new node has m links the network at time t will have $N = m+1+t$ nodes and $K = m(t+1)+1$ links, corresponding to an average degree $\langle k\rangle = 2m$ for large times. In the limit $t \to \infty$, the model produces a degree distribution of the form $P(k) \sim k^{-\gamma}$, with $\gamma = 3$. The average shortest-path in the BA model increases as $\ell \sim \log N/\log(\log N)$ and the clustering coefficient vanishes with the system size as $C \sim N^{-0.75}$. The BA model has attracted an exceptional amount of attention in the literature, many authors have proposed modifications and generalizations to make the model a more realistic representation of the real world, resulting in a more flexible value of the exponent γ, or to reinforce the clustering coefficient and degree correlation.

3.4 Spreading Dynamics

In the previous sections we have focused on the static properties of contact networks. However, in real systems there is another important aspect to add to the study of contacts, that is flow or spreading dynamics through the contacts.

Spreading processes in networks are often modeled by cellular automata [40]. Each node of the network represents an agent that can be in only one of a finite numbers of states. Time is discrete, and at each time step the next step of each agent is computed as a function of its state and the state of its neighbors on the network. In particular, in section 3.4.1 we will discuss the SIS model of epidemic, or spreading without immunization and in section 3.4.2 models of opinion. These two classes of processes are radically different. We will focus on the discovered influence of the inclusion of network topologies in those standard models.

3.4.1 Epidemic spreading

In the SIS model each node exists in one of the two states, healthy or *susceptible* and *infected*. The infection spreads through nearest neighbors contact, an infected site passing the disease to its healthy neighbors at rate λ. Infected sites recover at rate $1-\lambda$ and are immediately susceptible to infection. The rates for a node with 4 neighbors are shown in Fig. 3.6. Since an agent must have an

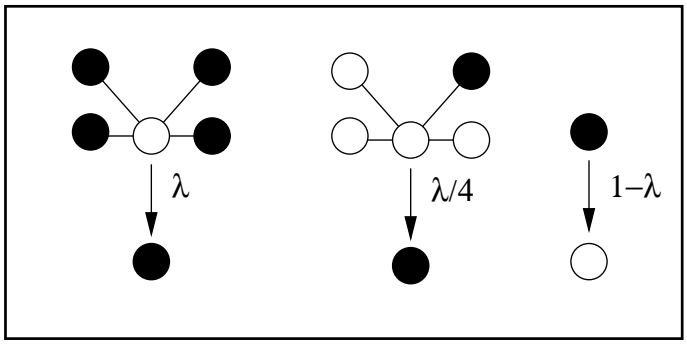

Figure 3.6: Example of SIS model for a chosen node with $k = 4$ neighbors. An infection event is chosen to occur with probability λ, and it occurs if the chosen node is susceptible (*white*) and if one randomly chosen neighbor is infected (*black*). With probability $1 - \lambda$, the node is cured, only if it is infected.

infected neighbor to become infected , the disease-free state is absorbing; persistence of the epidemic depends on the infection rate λ. The boundary between survival and extinction is marked by a critical point that characterizes transitions into an absorbing state.

Each step involves randomly choosing a process: *infection* with probability λ and *cure* with probability $1 - \lambda$. and choosing a node i. In a cure event a node, if infected, is cured. Infection proceeds only if i is susceptible and a randomly chosen node j is infected; if so, i is infected. Time is incremented by $\Delta t = 1/N$ after each step, successful or not , so that a unit time interval, on a network of N nodes, corresponds on average to one attempted event per site [41].

Following Ref. [42], we consider the relative density $\rho_k(t)$ of infected nodes with given degree k; i.e. the probability that a node with k links is infected. The mean field rate equation can be written as:

$$\delta_t \rho_k = -\rho_k(t) + \lambda k[1 - \rho_k(t)]\mathcal{P}(\rho(t)), \tag{3.12}$$

The creation term considers the probability that a node with k links is healthy $[1-\rho_k(t)]$ and gets the infection via a connected node. The probability of this last event is proportional to the probability of infection λ, the number of connections k and the probability that a given link points to an infected node $\mathcal{P}(\rho(t))$. For uncorrelated networks, its is given by:

$$\mathcal{P}(\lambda) = \frac{\sum_k kP(k)\rho_k(\lambda)}{\langle k \rangle} \tag{3.13}$$

Using the relations:

$$\rho = \int_m^\infty P(k)\rho_k dk. \tag{3.14}$$

For a network with scale-free distribution of the form $P(k) = (1+\gamma)m^{1+\gamma}k^{-2-\gamma}$ with $0 < \gamma < 1$ and first moment:

$$\langle k \rangle = \frac{1+\gamma}{\gamma} m. \tag{3.15}$$

The set of Eqs. [3.12 - 3.13] can be used to calculate $\mathcal{P}(\lambda)$ and then

$$\rho_k \sim \lambda^{1/(1-\gamma)}. \tag{3.16}$$

This means the total absence of any epidemic threshold and the associated critical behavior; i.e. $\lambda_c = 0$. The higher the node's degree, the smaller the epidemic threshold. Thus, in scale-free networks the unbounded fluctuations in the number of links emanating from each node ($\langle k^2 \rangle \rightarrow \infty$) eliminates the epidemic threshold.

Epidemic spreading on complex networks has recently attracted much attention in the literature. Various features have been investigated: different analytic formalisms, models with immunization, the incidence of degree correlations, strategies for halting the epidemics outbreak or fluctuating diseases (see details in Ref. [7]).

In chapter 5 of this research we investigate a novel aspect in this topic. The SIS model of infection described in this section is studied over a mobile population of agents, showing that a family of critical exponents is obtained as a function of the spatial correlations in the population.

3.4.2 Opinion formation

An interesting application concerning the structure of social networks is the modeling of the dynamics of opinion formation. Specific measurements that characterize the statistics behind the existence of different groups and affiliations within human populations, makes plausible the intents to model such aspect of human behavior. The idea behind this field is to find simple rules of interactions among the nodes or *agents*, each of which carries its own changing color or *opinion*, trying to reproduce the emergence of complex patterns observed in reality.

Such *opinions* can be defined by a finite number of integers as in the model proposed by Sznajd et. al. [43] or can even be represented by *real* numbers, having

a rich spectrum and opening the possibility for as many opinions as agents; like in the model proposed by Deffuant et. al. [44]. Here we summarize these two characteristic models of opinion formation. Others models as the Hegselmann and Krause [45], the voter model [46], the Galam's majority rule [47] and the Axelrod's model [48] will not be discussed.

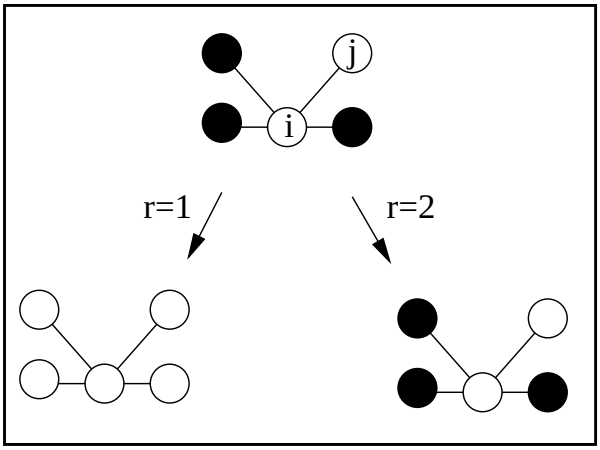

Figure 3.7: Example of the two different outputs generated by the Sznajd model on a cell with the same initial configuration and the rules "r-convince", with $r = 2$ and $r = 3$. Here, for the case $r = 2$ *consensus* is reached, i.e. an absorbing state where all the agents share the same opinion (*white*); for $r = 3$ the cell remains unchanged because there are not 3 neighboring nodes sharing the same opinion (*color*).

In the Sznajd model, the opinion o_i is an Ising like variable, assuming the value $+1$ and -1. At each time step r randomly selected neighboring agents transfer their opinion to all their neighbors if and only if they share the same opinion. Fig. 3.7 presents an example in a cell of 5 nodes and the two opinions presented as black and white. In the example, for the same initial configuration and the same pair of nodes (i, j), if the model is run with $r = 2$, "2 convince" consensus is reached, while for "3 convince" the cell remains unchanged.

In the Defuant model, o_i is a real number ($0 \leq o_i \leq 1$). At each time step, two randomly chosen interacting agents i and j check their opinions o_i and o_j. If the two opinions differ by more than a fixed threshold parameter ϵ ($0 < \epsilon < 1$), called the *confidence bound*, both opinions remain unchanged. If, instead $|o_i - o_j| < \epsilon$, then each opinion moves into the direction of the other by an

amount $\mu|o_i - o_j|$, with μ being a second tunable parameter ($0 < \mu \leq 1/2$).

Both dynamics have a natural absorbing state, in which all the agents share the same opinion or *consensus*. For a given initial configuration the dynamics is followed until the system reaches an equilibrium state characterized by the existence of one or several opinion groups, depending on the control parameters of the model.

As recently pointed out in Ref. [7] very few and not conclusive results exist for the consensus models on complex networks. We dedicate the first chapter of this thesis to investigate some aspects in this area. The greatest success of the Sznajd model is the possibility to reproduce numerically some of the empirical laws observed in political elections with a version of the model with many different opinions, each representing the preference of the node by a given candidate. In chapter 4, we demonstrate that this result is independent of the topology of the complex networks used; we reproduce previous observations of the model on a BA network, on a hierarchical network. This gives a powerful insight into the analytical treatment of the model, which allows us to propose an expression that coincides with the numerical simulations for the Sznajd model run on complex networks of growing and fixed size. Finally, in an attempt to generalize this observation to the case of more complex representations of opinions, we propose a model where each agent can have o-dimensional vector, which represents o different opinions and each component of the vector is like a Deffuant type of opinion, i.e. a real number $0 \leq o_i \leq 1$. We show that even for this kind of sophisticated model the transition to consensus does not depend on the details of the complex network used to model the interaction.

Chapter 4

Modeling opinion dynamics on hierarchical networks

In this chapter we study spreading dynamics on deterministic networks as a suitable theoretical approach to predict spreading dynamics on random scale-free networks. This chapter comprehends three related studies presented in three sections.

In the first section, the Sznajd model of socio-physics, that only a group of people sharing the same opinion can convince their neighbors, is applied on a scale-free network modeled by a deterministic graph. We study a model for elections based on the Sznajd model and the exponent obtained for the distribution of votes during the transient agrees with those obtained for real elections in Brazil and India. Our results are compared to those obtained using a Barabási-Albert scale-free network.

In the second section, we present a renormalization approach to solve analytically the Sznajd model of opinion formation on the same deterministic graph. For the case of two opinions, we present an expression for the probability of reaching consensus for a given opinion as a function of the initial fraction of agents with that opinion. The calculations reproduce the observed sharp transition of the model on a fixed network, as well as the smooth function for the model on a growing complex network.

In the last section of this chapter, we propose a more sophisticated model of

opinions where the agents interact with a non-linear rule, and each of them has n opinions, represented by the positive real components of an n-dimensional vector. We compare the behavior of this opinion model with the agents interacting through different networks, i.e. deterministic versus random structures.

4.1 Opinion Formation on a Hierarchical Network

The majority of networks used to generate a scale-free topology is stochastic, i.e. the nodes appear to be randomly connected to each other. These scale-free random networks have naturally a continuous degree distribution. But recently it has been shown that discrete degree distributions of some deterministic graphs asymptotically also exhibit a power law decay [50]. Furthermore, scale-free random networks are excellently modeled by such deterministic graphs [3, 4]. However the comparison between the behavior of stochastic and deterministic networks in the simulation of a particular model still remains open.

Opinions can either be made up by a person or taken over from another person. Sometimes some people try to force their opinions on others. In general, all people are free to form opinions. The mechanism of opinion formation is "normative", in the sense of what *ought to be*, opposed to a "positive" mechanism, which is based on observation *what is* [51]. Based on this facts, and with the necessary simplifying assumptions, socio-physics gave the opportunity to apply techniques of statistical physics to model opinion formation among people [47, 49, 52, 53].

One of the opinion formation models that has generated immediate interest in many authors on the field is the Sznajd model [43], which is based on the slogan "together we stand": Individuals are represented by the lattice nodes (one-dimensional in its first version), and each randomly selected pair of neighbors convinces all their neighbors of their opinions, if and only if the pair shares the same opinion; otherwise, the neighbors' opinion are not affected. It differs from other consensus models by dealing only with communication between neighbors, and the information flows outward as in rumor spreading: a site does not follow what the neighbors tell the site; some details of this model can be found in section 3.4.2.

Initially, two opinions ($+1$ and -1) are randomly distributed with probability p and $1 - p$ respectively over all the nodes of the lattice. The basic Sznajd model with random sequential updating always leads to a consensus (all sites have the same opinion and the whole system reaches a fixed point after a certain simulation time). A phase transition is often observed as a function of the initial con-

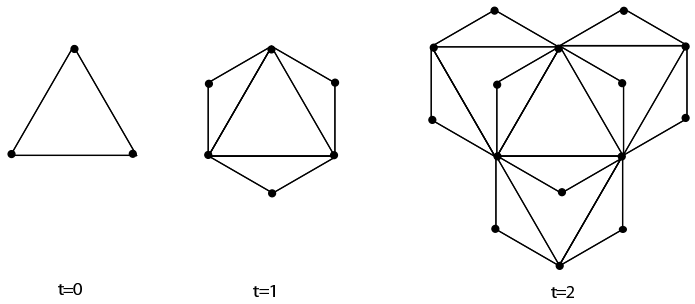

Figure 4.1: The three first generations of the scale-free pseudo-fractal graph. At each iteration step t, every edge generates an additional vertex, which is attached to the two vertices of this edge.

centration of opinion p. A generalization to many different opinions (instead of only ± 1) simulated on a Barabási-Albert network [34] reproduced quite well the results of the complex elections of city councillors in the state of Minas Gerais in Brazil in 1998 [54].

Next, we present the deterministic scale-free network to be used. Further, we simulate an election process, using the Sznajd model in this network. We compare our results with the same simulations carried out on a stochastic scale-free network (the Barabási-Albert network) and with state elections from Brazil and India.

4.1.1 Hierarchical network

The deterministic scale-free graph used in this work grows as follows: At each time step, every edge generates an additional vertex, which is attached to both end vertices of this edge. Initially, at $t = 0$, we have a triangle of edges connecting three vertices, at $t = 1$, the graph consists of 6 vertices connected by 9 edges, and so on (see Fig.4.1). The total number of vertices at iteration t is

$$N_t = \frac{3(3^t + 1)}{2} \tag{4.1}$$

This simple rule produces a complex growing network. Such a graph is called a *pseudo-fractal*. The resulting graph is not a fractal, as was already pointed out in Ref. [3]. The graph is surrounded by a long chain of edges and the resulting structure has not a fixed fractal dimension.

The presented hierarchical network has a discrete degree distribution. To relate the exponent of this discrete degree distribution to the standard γ exponent of a continuous degree distribution for random scale-free networks, we use a cumulative distribution $P_{cum}(k)$, which follows

$$P_{cum}(k) \sim k^{1-\gamma} \tag{4.2}$$

and is the probability that a node of the network hat at least k neighbors. It decreases as a power of k with exponent $\gamma = 1 + \ln 3/\ln 2$. In a similar way, the average clustering coefficient [3], which is the probability of existence of a link between two nodes when they are both neighbors of a same node, can be calculated for the infinite graph, $\bar{C} = 4/5$. One obtains a shortest-path-length distribution which tends to a Gaussian of width $\sim \sqrt{\ln N}$ centered at $\bar{l} \sim \ln N$ for large networks [3]. These properties concerning the degree distribution, the clustering coefficient and the mean length, are also present in a wide range of stochastic scale-free networks reported in the literature. They make our simple deterministic networks suitable to examine applications very often found on stochastic networks. In the next section we implement the Sznajd model on our scale-free pseudo-fractal.

4.1.2 Monte Carlo simulations of the Sznajd model

We let the fractal of Fig.4.1 grow and at each step assign sites with random opinion ± 1. At every step $t > 0$, we have the following process:

1. The network grows, i.e., 3^t new sites are added.
2. A random opinion (± 1) is set to each new node of the network, with probability p ($1 - p$) for opinion $+1$ (-1).
3. N_s Sznajd runs are performed. For each run, 3^t sites chosen randomly are analyzed and updated, i.e., one visits for the Sznajd model a number of sites equal to the number of sites added at that step to the network.

Three variations of the Sznajd model on the pseudo-fractal network have been investigated:

- **1 site convincing:** For each site i chosen, we change the opinion of all its neighbors to the site's opinion.

- **2 sites convincing:** For each site i chosen, we select randomly one of its neighbors. If this selected neighbor has the same opinion of the site i, then all their neighbors follow the pair's opinion. Otherwise, nothing is done.

- **3 sites convincing:** For each site i chosen, we select 2 of its neighbors at random. If all these three sites have the same opinion, they change the opinion of all their neighbors.

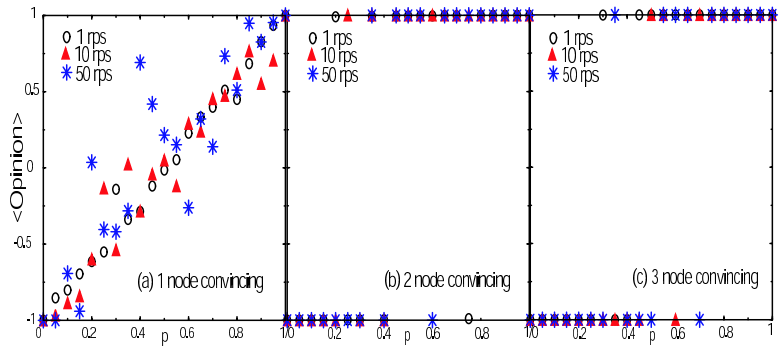

Figure 4.2: Sznajd model on a 29576 nodes pseudo-fractal network with $N_s = 1$, 10 and 50 runs per time step for (a) 1 node convincing, (b) 2 nodes convincing and (c) 3 nodes convincing.

Figure 4.2 shows the mean opinion for the three rules (1, 2 and 3 sites convincing) for different initial concentrations p and N_s Sznajd runs per step (rps): 1 rps ($N_s = 1$) corresponds to one realization of Sznajd per step, that is at each step, one chooses 3^t nodes of the network at random to simulate Sznajd. 10 rps ($N_s = 10$) and 50 rps ($N_s = 50$) correspond, respectively, to choose 10×3^t and 50×3^t nodes randomly per step. We let the network grow up to 29576 sites, that corresponds to 10 iterations. For 1 node convincing, the model cannot reach consensus, the results become more unpredictable, the larger the number of runs per step (Fig.4.2a). However, for two and three nodes convincing, full consensus is observed (Figs.4.2b and 4.2c). We clearly see that for rules 2 and 3 the order parameter (opinion) jumps drastically and shows strong hysteresis. Our results are very similar to the ones obtained on the Barabási-Albert network when the same rules are applied [56], except that the latter requires more Sznajd runs for the network to reach a fixed point, i.e., a full consensus.

Since after some time steps the rules 2 and 3 always lead to a consensus and the whole system reaches a fixed point, in Fig.4.3 we show the number of samples,

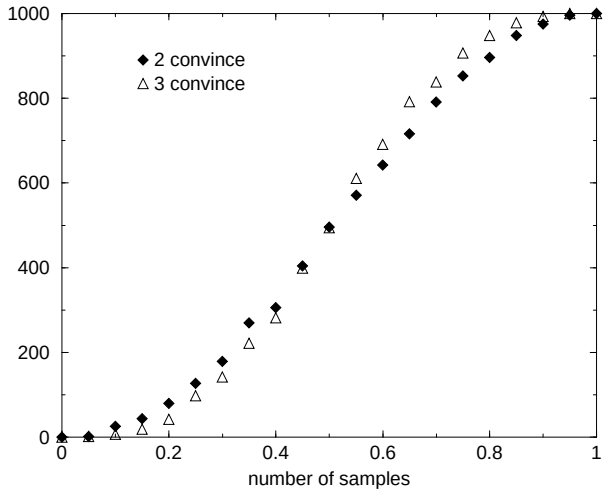

Figure 4.3: Monte Carlo simulation on a 29576 nodes pseudo-fractal network counting the number of samples, out of 1000, for which the fixed point all "up" is obtained when different values for the initial concentration p of nodes "up" are simulated. As we can see, the fraction of realizations with fixed point up depends on the probability p and on the rule implemented, for Sznajd models with 2 and 3 nodes convincing.

out of 1000, for which the fixed point all "up" (all with opinion $+1$) is obtained when different values for the initial concentration p of nodes "up" are simulated. As we can see, the fraction of realizations with fixed point "up" depends on the probability p and on the implemented rule. This is opposed to the results obtained with the same rules (2 or more nodes convincing) on a square lattice, where an abrupt change is observed for $p \geq 0.5$ [55]; but in agreement with the ones on a Barabási-Abert network [56] and on a 1-dimensional chain[43]. For 1 node convincing the system does not tend to a fixed point (consensus), while on square lattices it does [57, 58], although without the abrupt change of 2 sites convincing.

4.1.3 Simulation of elections with many candidates

We create a network of interacting nodes by using the pseudo-fractal network prescription as described before (Eq.4.1). In addition to our network, we also simulated the rule presented in [54] for modeling elections on a Barabási net-

work. We summarize the generation of the Barabási network as follows: At the beginning of the simulation we have 6 fully connected nodes. After that, we add more and more nodes and connect them with 5 of the present nodes chosen with a probability proportional to the number of nodes already connected to them. For comparison we generate both networks with the same number of nodes (voters), and the same number of candidates.

After generating a network, one starts with the "election process". N candidates are randomly distributed. The value n ($1 \leq n \leq N$) of a node on the network represents that this voter has preference for the candidate n. Initially, all sites start with value zero, meaning that there is no preference for any candidate, except for N sites that have the number of a candidate. Now, the electoral campaign starts (only voters with preference for a candidate can influence other voters, *à la Sznajd*). At each time step all the nodes are randomly visited: a random list of nodes assures that each node is reached exactly once. For each visit, we implement the following process:

- We choose a node i at random, if it has preference for a candidate, we choose among its connected nodes, a node j at random. Otherwise, we randomly select another node.
- If node j has the same candidate as node i, each node convinces all the nodes connected to it with probability:

$$p(k) = \frac{1}{n(k)^{1/\xi}} \tag{4.3}$$

  where $n(k)$ is the number of nodes connected to either i or j, and $\xi > 1$ a switching factor that is analyzed later and is a parameter of the model.
- If node j has no candidate, node i convinces it to accept its own candidate with the probability $p(i)$ of Eq. (4.3).
- If node j has a different candidate from node i, we choose another node i.

As in real elections, we do not wait for a kind of fixed point which here would correspond to all the nodes preferring the same candidate, but we count the votes at an intermediate time. We group in a histogram the *number of candidates* which received a certain number of *votes*. Because the bin size for the *votes* increases by a factor 2 for each consecutive bin we divide each point of the histogram by the bin size, for this reason we have numbers lower than one for the histogram of the number of candidates, this kind of "voting distribution" is used in the literature

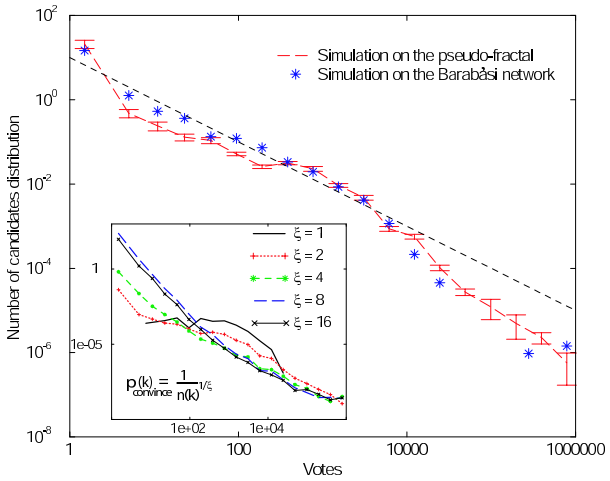

Figure 4.4: Plot of the voting distribution of an election process on the pseudo-fractal compared to the simulation on a Barabási network. Both networks have 797163 sites (voters) and 500 candidates. We average over 20 realizations for each network. The distribution for the pseudo-fractal is obtained after 15 iterations of the convincing process with $\xi = 2$ (Eq. 4.3) . The distribution for the Barabási network is obtained after 83 iterations. In the inset we show the results of the simulation on the pseudo-fractal for $\xi = 1$, 2, 4, 8 and 16, after 400, 20, 4, 2 and 1 iterations respectively. We see that the results depend on the rule chosen.

for analyzing similar results [54, 61]. In Fig.4.4 we see that the results of the voting distribution for the simulations on the pseudo-fractal and on the Barabási network agree very well, using $\xi = 2$ (Eq. 4.3) with the pseudo-fractal. In the inset we show the results of the simulation on the pseudo-fractal for $\xi = 1$, 2, 4, 8 and 16, after 400, 20, 4, 2 and 1 iterations, respectively. The results differ for each choice of ξ: the system reaches a fixed point more rapidly for larger parameter ξ, i.e, one needs less iterations in order to reach the same distribution's width. Using the same value used in the simulations on the Barabási network, $\xi = 1$, after nearly 2000 iterations, a fixed point (consensus) on the pseudo-fractal network could not be reached. A similar result is observed on the square lattice for the Sznajd model, where if convincing happens only with a certain probability, no complete consensus is found [55, 59].

For the simulation on the Barabási network we use $\xi = 1$, like reported in Ref. [54]; this choice means that on average each node convinces one other node at each process. As we show in Fig. 4.5 the degree distribution of the pseudo-

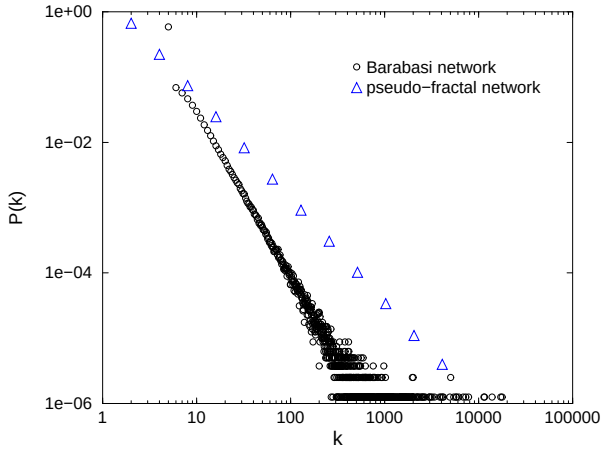

Figure 4.5: Comparison of the degree distributions of the Barabási network $P(k) \sim k^{-\gamma}$ ($\gamma = 2.9$), and of the pseudo-fractal network, $P(k) \sim k^{1-\gamma}$ ($\gamma = 1 + \ln 3/\ln 2$).

fractal is discrete and is given by $P(k) = k^{-\ln 3/\ln 2}$, while the degree exponent for the Barabási network presented in this work is $\gamma = 2.9$ [34]. Knowing $P(k)$ we can calculate the probability distribution to convince choosing one site at random, that is the probability of convincing Eq.4.3 multiplied by the degree distribution $P(k)$. For the convincing distribution of both networks to be similar, one has to chose Eq. 4.3 with $\xi = 2$ for the simulation on the pseudo-fractal. We obtain qualitatively the same results in both simulations, being 12 time faster on the pseudo-fractal than on the Barabási network, because it requires less memory space and computation time (see Fig.4.6).

In Fig. 4.7 we see that the shape of the voting distribution after a given number of iterations does not change with the size of the network. Changing the number of candidates doesn't alter the form of the distribution (see the inset of Fig. 4.7).

We consider the results from two Brazilian states (São Paulo and Minas Gerais) for the positions of local state deputies. In such elections the voters vote directly for the candidate and not for parties. Some elections occur with a high number of candidates, of the order of thousands, and with a number of voters of the order of millions. The official results for each state are available on the Internet [60].

In Fig.4.8 we see the results of a simulation on a 15-generations pseudo-fractal (21523362 nodes) and 1144 candidates, and compare it to the results of real elec-

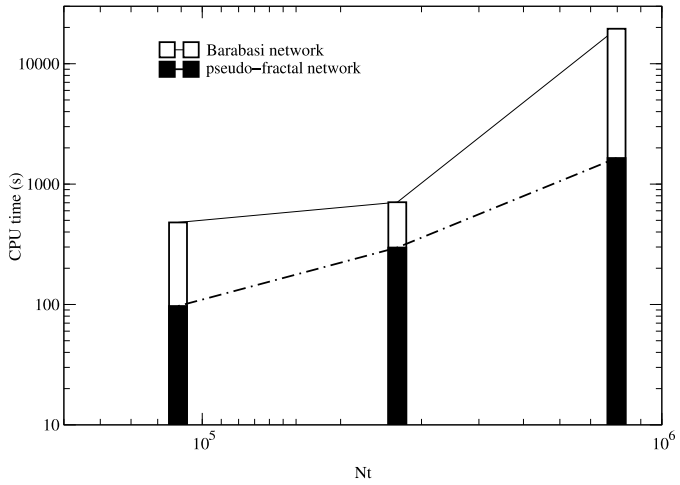

Figure 4.6: CPU time (seconds) on a PIV 2.4 GHz vs. size of the network for the pseudo-fractal network and for the Barabási network. We simulate an election process with 100 iterations of the convincing model, on networks of sites: 88575, 265722 and 797163 with 500 candidates. The computation time on the pseudo-fractal grows linearly with the size of the system, while for the stochastic network it grows exponentially.

tions. For the moment we do not take into account abstentions, or invalid votes in the simulation. The pseudo-fractal tends to consensus in a similar way as the real elections. After averaging over more than 100 realizations we see, that the deviations from a perfect line in the intermediate regions of votes, seen in the real elections, are not of statistical nature but seem due to the determinism of the network. The general behavior of the distribution of candidates of the simulation on the pseudo-fractal, however, agrees with the one observed in real results. The results of the distribution of candidates for the simulated and the real cases follow a hyperbolic law,

$$N(v) \propto 1/v, \tag{4.4}$$

for the number N of candidates having v votes, extending over two or three order of magnitude, with deviations only for small and large numbers of votes [61].

In addition we analyzed the behavior of the voting distribution for the elections

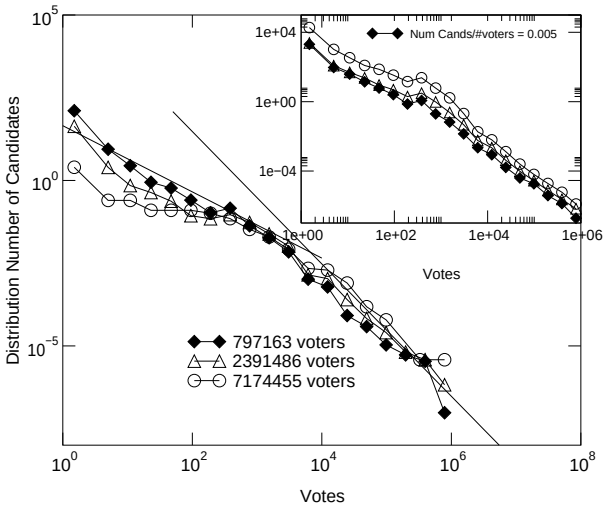

Figure 4.7: Test for size effects on the distribution of candidates for the simulation of elections on the pseudo-fractal with 12 generations (diamonds), 13 generations (triangles) and 14 generations (circles) with 2000 candidates each, after 15 iterations of the election process. In the inset we make the simulation with a number of candidates proportional to the number of sites of the network, keeping the relation $\#Cands/\#voters = 0.005$, we see that the results are size-independent. The solid lines are guides to the eye with slopes -1 and -2.

in India to the lower house of the Parliament (Lok Sabha). These elections are events involving political mobility and organizational mobility on an amazing scale. In the 1998 election to Lok Sabha there were 1269 candidates from 38 officially recognized national and state parties seeking election,1048 candidates from registered parties not recognized and 10635 independent candidates. A total number of 596185335 people voted. The Election Commission employed almost 400000 people to run the election. The official results of these elections are available on the Internet [62]. In contrast to the Brazilian elections, in this process the country is divided into 543 parliamentary constituencies, each of which return one representative to the Lok Sabha. That is, not all the voters elect among all the candidates of one state (like in Brazil), but there is one election process for each of the 543 representatives, that occurs in each parliamentary constituency. These parliamentary constituencies are selected by an independent Delimitation Commission, which creates constituencies which have roughly the same population, subject to geographical considerations and the boundary of the states and administrative areas.

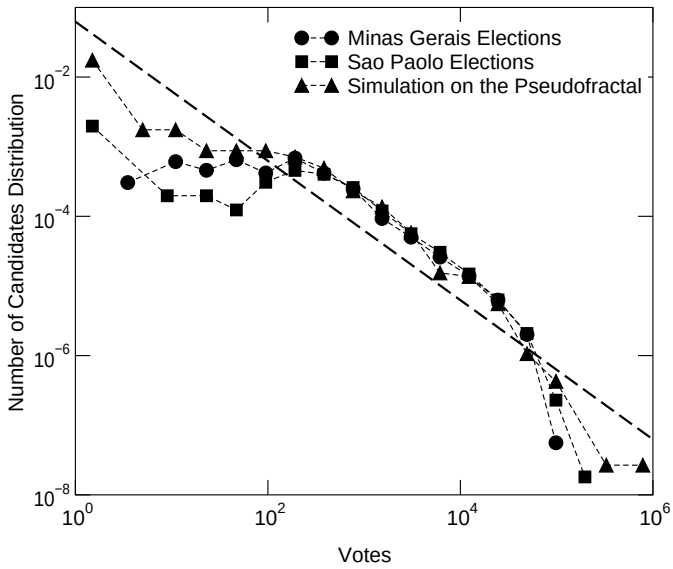

Figure 4.8: Result of the simulation of an election process after 20 iterations on a pseudo-fractal network of 21523362 nodes and 1144 candidates (triangles). Compared to the voting distribution for the state of Minas Gerais in 1998 (circles) (819 candidates, 11815183 voters) and the state of São Paulo (1998) (squares) (1260 candidates, 23321034 voters). Both axis are plotted on a logarithmic scale. The dashed straight line is a guide to the eye with slope -1. The bin size for the $votes$ increases by a factor 2 for each consecutive bin. The height of the distribution of the pseudo-fractal is multiplied by 10 to better see the comparison of the results.

At the end of the process one can analyze the voting distribution of each state of India. In Fig. 4.9 we observe the results for 5 states: Uttar Pradesh, Goa, Andhra Pradesh, Haryana and Kerala, with 85, 2, 42, 10 and 20 constituencies respectively. The voting distribution for each state is the superposition of different electoral processes for all the constituencies of the state. For each election in a constituency there is a voting result that corresponds to few candidates, between 5 and 10. The difference between the Indian and Brazilian results appears at this stage.

If one analyses the election process of each parliamentary constituency, even though one does not have a large number of candidates for the statistics, a dis-

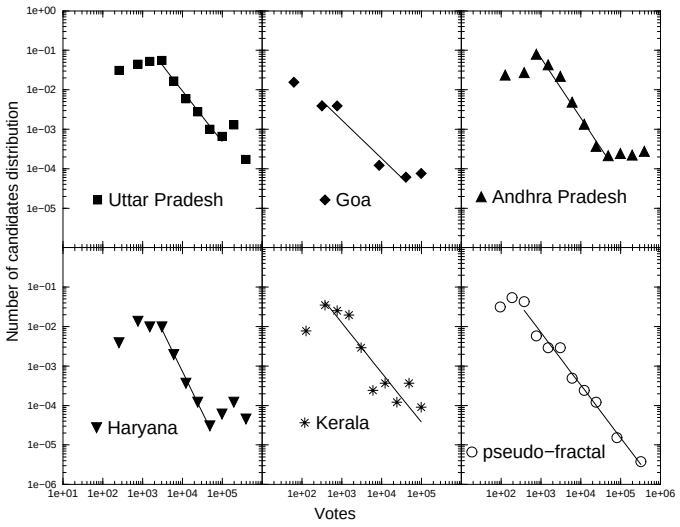

Figure 4.9: Voting distribution for state elections of India in 1998. For Uttar Pradesh (squares) with $55,015,804$ voters and 649 candidates. Goa (diamonds) with $532,766$ voters and 12 candidates. Andhra Pradesh (triangles up) with $31,829,338$ voters and 301 candidates. Haryana (triangles down) with $7516,884$ votes and 84 candidates and Kerala (stars) with $13,036,581$ voters and 108 candidates. We obtain results qualitatively comparable to this kind of processes on a pseudo-fractal with 12 generations and 80 candidates, after few iterations (5). The solid lines are square fits to the data, in the intermediate regions. The slopes are: -1.32 (Uttar Pradesh), -0.97 (Goa), -1.51 (Andhra Pradesh), -2.06 (Haryana), -1.26 (Kerala) and -1.32 (Pseudo-fractal)

tribution with slope ~ -1 is observed. However, the final results of the process for each Indian state provide a different profile far from the hyperbolic one ($1/v$ type of distribution) (see Fig.4.9), which was observed for many Brazilian states [54, 61]. In order to analyze the number of candidates N which received a certain fraction of votes v for the nationwide voting process, we have normalized the votes of each candidate by the total number of voters (Fig.4.10). As can be seen, the number of candidates N follows a power law $N(v) \propto v^{\alpha}$, with $\alpha = -1.3$ (for Brazil, $\alpha = -1$).

The differences in the Sznajd simulations on the pseudo-fractal network for the Indian and the Brazilian elections are mainly due to the number of candidates considered for each one. In the latter, the number of candidates is almost 0.005%

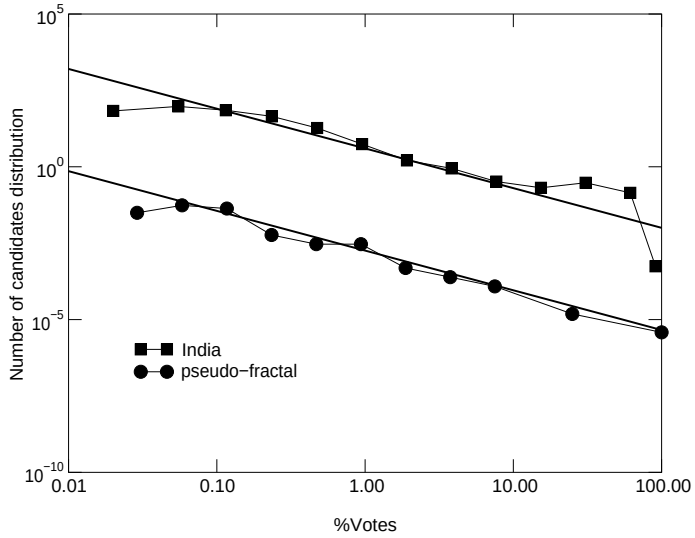

Figure 4.10: Voting distribution for India in 1998 (squares), compared to the simulation on a pseudo-fractal (circles) with 12 generations and 80 candidates, after few iterations (5). The solid lines are guide to the eye with slope -1.3

of the lattice nodes, while for the former it is 0.01%. The fixed point (consensus) is reached the faster, the larger the density of candidates. Because of that, the comparison of the real elections in India with our simulations were made taking account only 5 iterations of the Sznajd model on the pseudo-fractal network (20 iterations for Brazilian elections).

4.1.4 Summary

We studied the behavior of a deterministic scale-free network simulating a spreading of opinion model on it. We solved the Sznajd's model on a growing deterministic scale-free network, obtaining always a consensus of opinion after some runs of the model. The final opinion presents a smooth transition as a function of the initial concentration of opinions in the network for rules 2 and 3. The results coincide with the results reported for a stochastic scale free network [56]. We found that there is not scaling in the finite-size cut-off showing that the system is not critical (Fig.4.7). We simulated election processes on the network. The probability of convincing of the model has to be adapted ($\xi = 2$,

in Eq. 4.3) in order to avoid the differences in the values of the exponents γ of the degree distribution of our deterministic network and the γ exponent of a typical stochastic network (Fig. 4.4). We obtained the same results, reported for a stochastic scale-free network, but with computation times considerably lower (see Fig.4.6). We could use our model to reproduce qualitatively good complex electoral processes, such as the state elections for parliament in Brazil and India.

4.2 Renormalizing Sznajd model on complex networks taking into account the effects of growth mechanisms

In this section we take advantage of the deterministic character of the scale-free network from the previous section and present an approach to solve analytically the Sznajd model on it.

On networks with fixed size, the results of the Sznajd model do not depend much on the spatial dimensionality and type of neighborhood selected (i.e., two nodes convince the others, three convince the others, etc.) [55, 63, 64]. In the case of q choices of opinion, the model has q homogeneous absorbing states, where all individuals choose the same opinion; in the context of opinion, one says the system reaches consensus. The case of two opinions ($q = 2$) has been the most studied, denoting opinions as Ising variables "up" or $+1$, and "down" or -1. In more than one dimension, the probability (P_{up}) of reaching consensus "all up" depends on the initial fraction p of individuals with opinion "up"; for $p > 0.5$, the probability of reaching "all up" as stationary state is close to one, while for $p < 0.5$ it is negligible, having a sharp transition at $p = 0.5$, which can be interpreted as a dynamical phase transition. Computer simulations in [65] indicate that the universality class associated to this dynamical phase transition is different from the universality class of the Ising model. The distribution of time needed to reach the stationary state is a peak followed by a fast decay [66].

Much less is known about the Sznajd model on growing networks. Interactions of groups of people in some circumstances can be thought of as a growing system, i.e., in a city with positive rate of immigration. In a first and simple approximation, it can be modeled by a growing scale-free network [1]. We showed in the previous section that if the Sznajd model is simulated during the growth of either a Barabási-Albert network or a hierarchical network [3], the system reaches consensus [56, 67]. But in contrast to the sharp transition observed for the networks

of fixed size, the probability that the system reaches "all up" for a growing complex networks is a smooth function of p the initial fraction of opinion "up". In addition, this function depends on the type of neighborhood selected.

In this section, we present a real space renormalization approach [68] to calculate the probability $P_{up}(p)$ of reaching consensus on opinion "up" as a function of p. Our results are for two common rules of neighborhood, namely "r-convince all their neighbors", with $r = 2$ and $r = 3$. We have obtained the two well-known results known for the model: a smooth function of p for the growing case and an expression which approximates the step function for fixed networks.

Next, we present the renormalization approach and the analytical expressions obtained, each case is compared with the results from the numerical simulations, previously reported in [67], as well as for the BA scale-free network.

4.2.1 Renormalization Approach

Our method can be very intuitive and is based on the method proposed by Galam to study bottom-up democratic voting by majority rule in a square lattice [47], where the predictions of the results on the lattice are based on the application of the majority rule to a basic cell of neighbors, called renormalization cell.

We find that given a neighborhood rule, it is enough to choose an appropriate generation of a hierarchical network for calculating $P_{up}(p)|_{r,g}$, which agrees with the numerical results of the model on growing networks. The subscript index r, g in $P_{up}(p)|_{r,g}$ denotes that the resulting function belongs to a chosen Sznajd rule (r) in a growing network (g). Subsequent self-iterations of $P_{up}(p)|_{r,g}$ result in a step function, i.e., $P^n_{up}(p)|_{r,g} = P_{up}(p)|_{r,f}$, where the subscript index f corresponds to the result obtained for a network of fixed size.

For a population fraction p with opinion "up", the general method is as follows:

- Given a neighborhood rule r, the chosen basic cell corresponds to the minimum generation t of the hierarchical network, such that $r > N_t$ (the r agents must have at least one agent to convince). We call this resulting number of nodes in the cell n_r.

- The probability of each possible configuration in a elementary cell is easily

calculated, such that

$$1 = P_{all}(p)|_r = \sum_{k=2}^{n_r} B_{n_r k} p^k (1-p)^{n_r - k}. \tag{4.5}$$

with the binomial coefficient $B_{n_r k}$:

$$B_{n_r k} = n_r!/[k!(n_r - k)!]. \tag{4.6}$$

- From all the configurations calculated above, we select the subset that gives "all up" when applying the selected Sznajd rule on the cell, the sum of all of them is $P_{up}(p)|_{r,g}$:

$$P_{up}(p)|_{r,g} = P_{all}(p)|_{r,up} \tag{4.7}$$

Next, we illustrate the result of the method with $r = 2$ and $r = 3$.

4.2.2 Case $r = 2$

Growing

For $r = 2$, the triangle of the generation $t = 0$ is the basic cell. Thus $n_r = 3$ and, for a given fraction p, all the possible configurations are:

$$1 = P_{all}(p)|_2 = p^3 + 2p^2(1-p) + 2p(1-p)^2 + (1-p)^3 \tag{4.8}$$

If we apply the selected Sznajd rule $r = 2$ over the triangle, only the configurations expressed in the first two terms of the sum give "all up". Therefore:

$$P_{up}(p)|_{2,g} = p^3 + 2p^2(1-p) = 3p^2 - 2p^3 \tag{4.9}$$

In Fig. 4.11, we can see the good agreement of Eq. 4.9 with the numerical results [67] for the Sznajd model on a growing hierarchical, as well as on the Barabási-Albert scale-free network [1].

Fixed

In order to recover the reported result on a fixed network, one makes renormalization iterations, which means simply self-composing the Eq. 4.9:

$$P_{up}(p)|_{2,f} = P_{up}^{n_i}(p)|_{2,g}, \tag{4.10}$$

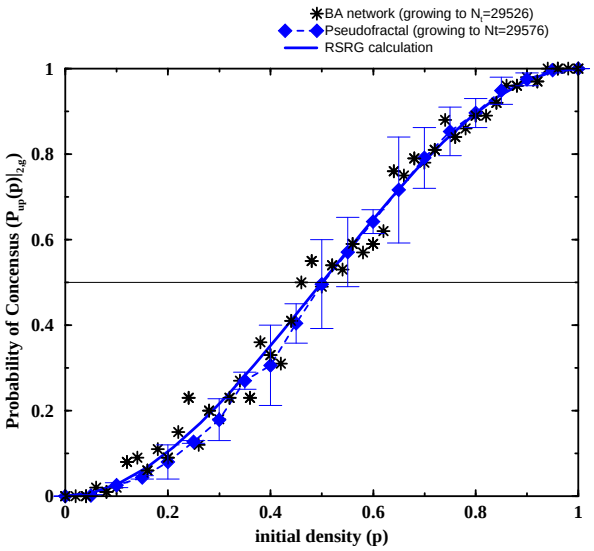

Figure 4.11: Comparison between the function presented in Eq. 4.9 (solid line) with Monte Carlo simulations on a growing hierarchical network (triangles with error-bars) and on a growing BA scale-free network (stars). In both networks, 29576 nodes are considered. We count the number of samples, out of 1000, for which the fixed point all "up" is obtained when different values for the initial concentration p of nodes "up" are simulated for rule $r = 2$.

and in the limit of a large number of iterations ($n_i - 1$), one recovers the step function observed numerically for the model on fixed networks. Note that the number of terms and the coefficients sizes increase very fast, as one can observe in the expression of only one composition:

$$P^2_{up}(p)|_{2,g} = 27\,p^4 - 36\,p^5 - 42\,p^6 + 108\,p^7 - 72\,p^8 + 16\,p^9, \qquad (4.11)$$

therefore, the multiple compositions presented in Fig. 4.11 are iterated using a computer. Figure 4.12 shows that the numerical simulations on large networks tend to the step function calculated from Eq. 4.11 with $n_i = 100$.

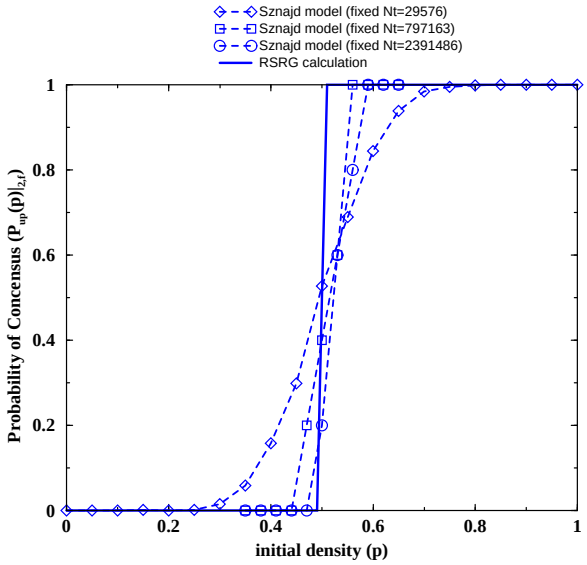

Figure 4.12: Eq. 4.10 with $n_i = 100$ (solid line) compared with simulations on a fixed hierarchical networks with $N_t = 29576$, 797163 and 2391486 nodes (dashed line with symbols). Other simulation conditions as presented in the caption of Fig. 4.11.

4.2.3 Case $r = 3$

Growing

The core of the method is the selection of the correct configurations after applying the Sznajd rule on it. As we will see for this rule, when the number of nodes in the renormalization cell is even, there are some symmetrical configurations which can have either "all up" or "all down" with the same probability. In this case only half of them are summed to P_{up}. For $r = 3$, the generation $t = 1$ is the basic cell. Thus $n_r = 6$ and, for a given fraction p, all the possible configurations are:

$$1 = P_{all}(p)|_3 = (1 + (1 - p))^6. \tag{4.12}$$

Note that the values of the binomial coefficient in the consecutive terms are: $1, 6, 15, 20, 15, 6, 1$. From the 20 configurations of the $4th$ term, there are 7 that give "all up"(shown in Fig. 4.15 at subsection 4.2.4), the corresponding 7 opposed cases which give "all down", and 6 symmetrical configurations shown in

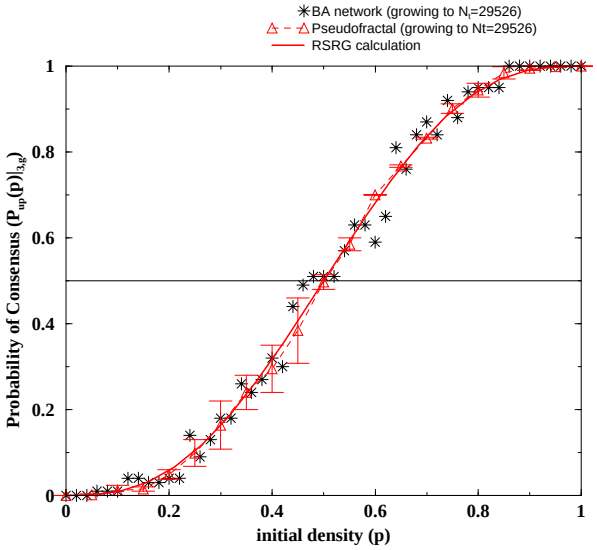

Figure 4.13: Eq. 4.13 (solid line) compared with the results from the simulations on a growing hierarchical network (triangles with error-bars) and on a growing BA scale-free network (stars) when $r = 3$. The other simulation conditions are the same of Fig. 4.11.

Fig. 4.16 (subsection 4.2.4) that can give either "all up" or "all down". Therefore, these group of configurations contribute with $7 + 0.5 \times 6$, and we have:

$$P_{up}(p)|_{3,g} = p^6 + 6\,p^5\,(1-p) + 15\,p^4\,(1-p)^2 + 10\,p^3\,(1-p)^3 \qquad (4.13)$$

In Fig. 4.13, we see that Eq. 4.13 agrees very well with the numerical results [67] for the Sznajd model on a growing network when the rule $r = 3$ is considered.

Fixed

The result of the composition for this case is far more complicated and only one self-composition of Eq. 4.13 ($n_i = 2$) already needs a computer, as shows the following expression:

$$P_{up}(p)|^2_{3,g} = -1249989\,p^{12} + 390897\,p^{11} - 158184\,p^{10} + 28561\,p^9 -$$

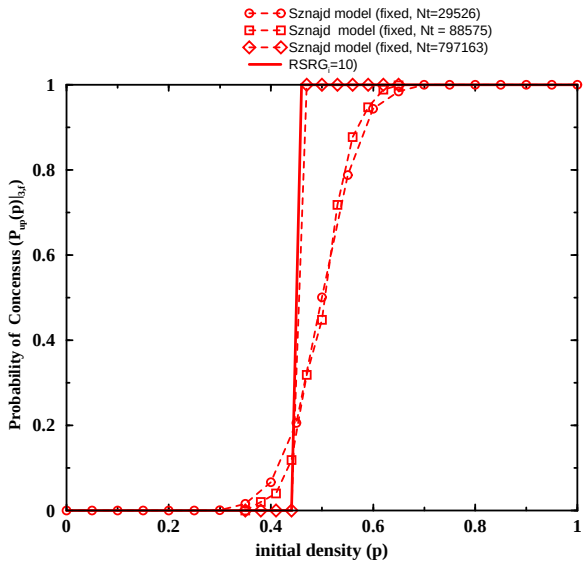

Figure 4.14: Eq. 4.13 self-composed in 9 steps (solid line) compared with the result from the simulations on a fixed hierarchical network networks with $r = 3$ (dashed lines with symbols). The other simulation conditions are the same as presented in the caption of Fig. 4.12.

$$643783179\,p^{18} + 270741222\,p^{17} - 100735317\,p^{16} + 41109081$$

$$p^{15} - 17504838\,p^{14} + 5585931\,p^{13} - 15244686567\,p^{24} +$$

$$11863411551\,p^{23} - 7642674243\,p^{22} + 4315583718\,p^{21} -$$

$$2347570026\,p^{20} + 1281132990\,p^{19} - 816731505\,p^{30} +$$

$$2281401855\,p^{29} - 5100164190\,p^{28} + 9199907505\,p^{27} -$$

$$13440029166\,p^{26} + 15908268375\,p^{25} - 2187\,p^{36} + 65610\,p^{35}$$

$$-925101\,p^{34} + 8148762\,p^{33} - 50268195\,p^{32} + 230706630\,p^{31}$$

In Fig. 4.14 we see the step function obtained with only 9 steps of composition compared with the numerical results on a fixed network of different sizes; as we see the results agree very well with the simulations of the model on large networks.

4.2.4 Additional Information: Configurations with the same fraction of nodes "up" and "down"

Here we present some of the possible configurations applying the Sznajd rule corresponding to the $r = 3$ on its appropriate renormalization pattern ($n_3 = 6$). In particular, we show the case of half of the nodes having opinion up, mentioned in Section 4.2.3,and represented by the fourth term in Eq. 4.13.

Figure 4.15 shows the 7 configurations that give as a result "all up", when applying the Sznajd rule, i.e., three consecutive nodes with opinion $+1$ convince all their neighbors. Note that interchanging $+$ and $-$, we have the 7 configurations for the opposed case of consensus "all down".

Figure 4.16 presents the 6 symmetrical configurations that have 3 consecutive nodes with $+1$, as well as 3 nodes with -1 giving consensus "all up" or "all down", respectively. Thus, these configurations contribute wit 0.5×6 to the probability of consensus "all up"(P_p), as showed in Section 4.2.3.

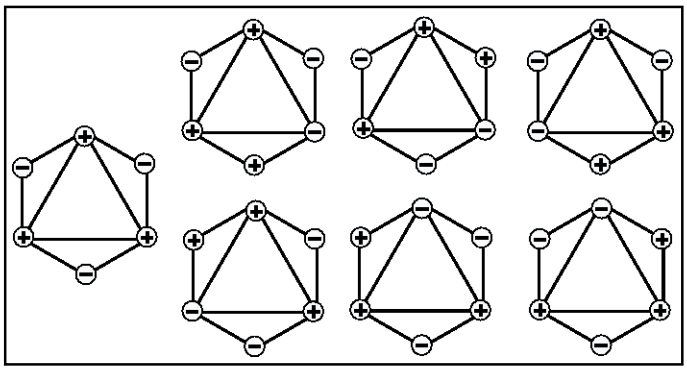

Figure 4.15: Configurations that generate consensus "all up" with $r = 3$ and the same fraction of opinions "up" and "down".

4.2.5 Summary

Based on opinion formation rules of the usual Sznajd model, we use a renormalization approach to give an expression for the probability of consensus into one opinion as a function of the initial fraction of this opinion.

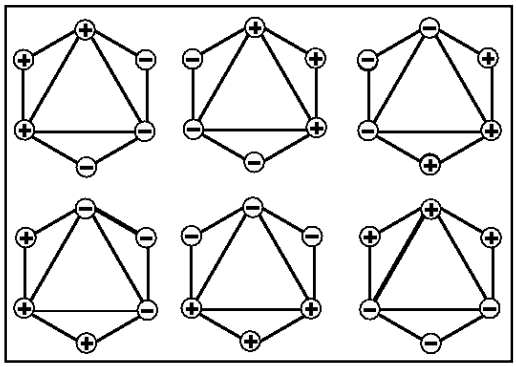

Figure 4.16: Configurations that generate either consensus "all up" or "all down" with $r = 3$.

We show that for a given Sznajd rule it is enough to solve exactly the model on an appropriate basic cell in order to find an expression for the smooth function, found numerically for the model on a growing network. Several self-compositions of the obtained function give the step function observed for the model on a network of fixed size. Further renormalization patterns should be tested, but in order to reproduce the results of the Sznajd model on growing SF networks, a SF hierarchical network must be chosen.

The proposed method could be, in principle, extended to other types of neighborhood and more interestingly to many choices of opinion ($q > 2$) which, as shown in section 4.1, can simulate elections processes [54, 66, 67], obtaining results consistent with some empirical observations[61].

4.3 A general model of opinion on different topologies

In this section, we present a general model of opinion, where the agents interact with a non-linear rule, and intends to be a general version of previous models where the opinion is represented by just an integer number.

Every agent i is characterized by its own opinion vector $a_n^{(i)}$ of $n = 1, .., O$ opinions. Each element of this vector corresponds to a different opinion about the same topic. For instance, agent j is 20% communist and 80% capitalist (given

$O = 2$): $a_1^{(j)} = 0.2, a_2^{(j)} = 0.8$. Each time step every agent actualizes its opinion vector by comparing its values to the ones of k other agents. These other agents are chosen by the topology of the graph, and the agent updates its opinion vector due to the following rule,

$$\hat{a}_n^{(i)}(t) = \sum_{l=1}^{k} a_n^{(i)}(t) a_n^{(l)}(t) + kg(t), \tag{4.14}$$

where $g(t)$ presents a stochastic variable, distributed uniformly in the interval $[0, \eta]$. This stochasticity can be interpreted to be due to misunderstandings among the agents, the spread of wrong information, or other perturbing actions. The interaction term in this model is of second order, and thus favors strong opinions on one topic. The factor k avoids that agents which have more connections feel less noise. In order to guarantee that the sum of opinions is equal to one, the vector is normalized afterwards,

$$a_n^{(i)}(t+1) = \frac{\hat{a}_n^{(i)}(t)}{\sum_{m=1}^{O} \hat{a}_m^{(i)}(t)}. \tag{4.15}$$

The main parameter of this model is given by the maximal noise η which we will call from now on the control parameter. Its role corresponds to the one of a temperature in physical systems. Other free parameters of the system are the number of agents N, the number of opinions O of an agent and the number of agents k to interact with per time step, which depends on the topology of the interaction network.

4.3.1 Influence of the interacting topology

We compare the behavior of the presented opinion model if the agents interact with their k nearest neighbors on different networks topologies. We study two different kinds of scale-free networks; i.e. networks with a power law degree distribution $k^{-\alpha}$.

Those are the Barabási-Albert network (BA) [34] and the Apollonian network [4]. Those networks have considerable topological differences, that can be expressed in terms of their clustering coefficient C. The BA network has a clustering coefficient, C, which depends on the network size as $N^{-0.75}$. It is independent of the degree of the nodes. In contrast, the Apollonian network has hierarchical structure with C depending on the degree of the node as a power

law of the degree and its average value is high ($C \approx 0.8$) and independent of the network size N. Both types of scale-free networks, with and without hierarchical structure have shown to be good models for rather different kinds of social interaction networks, from social collaboration networks [15] to networks of sexual contacts [69].

Further, we show that despite of the structural differences of these networks, the formation of consensus depends mainly on the noise and is independent of the specific topology of the scale-free network studied in the case of two opinions. We compare the behavior of the dynamics of the two complex networks mentioned above, and analyze also the results on a regular network with $k = 6$. The chosen regular network is generated from a chain ($k = 2$), adding interactions up to the third nearest neighbors. Simulations with mean field (MF) interactions are also simulated. By MF, we mean that at each step an agent interacts with k agents chosen randomly among all the others. We observe that the transition to consensus as a function of noise for the two scale-free networks, seems to belong to the same type of transition as MF.

At each time step pairwise interactions run over all the agents. The results reveal that there are two different absorbing states the system can reach. At small values of the control parameter (maximum noise η) one opinion completely dominates the system, o_{max}. For a noise η larger than a critical value η_c, each opinion remains with the same frequency, $1/O$. The order parameter D is the frequency of the agents which have an opinion vector with the same dominant opinion, being itself dominant in the system. To be more precise: for each agent we search its strongest opinion and then count, for each opinion, "n" the number of agents with this opinion as their dominant one. The largest value n_D, and so the most dominant one of the system, determines $D = n_D/N$. $\langle D \rangle$ means, that we average D over many time steps. This order parameter is normalized, so that is unity if all agents have the same opinion to be the dominant one, a state we call the consensus state. The value $1/O$ corresponds to an uniform distribution of opinions. A transition occurs between consensus and uniform distribution, when $\langle D \rangle$ goes from 1 to $1/2$ in the case of two opinions (left side of Fig. 4.17).

In figure 4.17 we show $\langle D \rangle$ vs. η for the model on the networks BA (triangles), and Apollonian (solid line), compared to the result of MF connections (plus signs) and the regular network (circles). The results of the figure are the average over 20 realizations on systems of $N = 124$ agents and 2 opinions. The differences in the type of transition from consensus $D = 1$ to $D = 1/2$ is due to differences in the dynamics of the model depending on the topology.

We illustrate this fact, from Fig. 4.17b to Fig. 4.17d, plotting $a_{o_{max}}$ vs. time, for

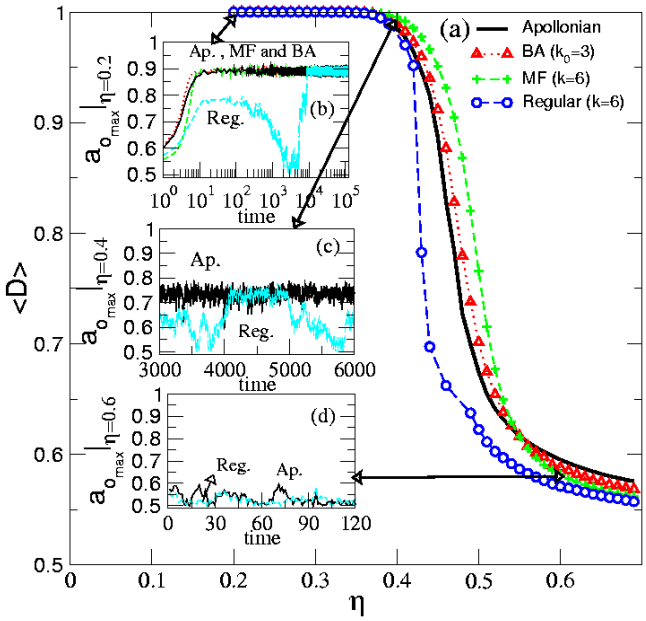

Figure 4.17: (a)Influence of the topology of networks on the transition to consensus ($D = 1$) as a function of noise (η). The transition on two different scale-free networks, the Apollonian (solid line) and the Barabási-Albert network (triangles) is similar to the one observed in mean field interactions (plus signs), and differs from the transition on a regular lattice (circles). In the three insets we plot the value of the dominant opinion, $a_{o_{max}}$ vs. time. (b) Comparison of the behavior of $a_{o_{max}}(t)$ on the four networks: Apollonian (solid line), BA (dotted line), MF (dashed-line) and Regular (long dashed line) for a fixed noise ($\eta = 0.2$). One observes that for this noise, which is below the critical noise, in the regular network the emergence of consensus takes longer in scale-free and MF interactions, which have similar behavior (three upper curves). (c) Near but below the transition, for $\eta = 0.4$, we compare the response of the regular and the Apollonian network. It is observed that for the former there is an intermittency among consensus $D = 1$ and $a_{o_{max}} = 0.79$ and not consensus $D = 0.5$ and $a_{o_{max}} = 0.5$. This behavior is not observed in the complex networks. (d) Above the transition ($\eta = 0.6$) the consensus is broken and the dynamics of the opinion $a_{o_{max}}$ vs. time behaves similar by in regular and complex networks. All simulation runs are with systems of 124 agents.

different values of η with the corresponding arrows marking the value of $\langle D \rangle$. For $\eta = 0.2$, which is below the critical noise, for each type of interaction the system reaches consensus $D = 1$, with a dominant opinion $a_{o_{max}} = 0.89$. For the scale-free networks and MF the consensus is reached considerably faster than for the regular lattice as we see in Fig. 4.17b. This effect is due to the metastable states of competing regions with the same opinion which takes longer to reach consensus on regular structures.

On the regular lattice the system presents an intermittency near but below the transition point ($\eta_c = 0.45$). We observe this intermittency of the dynamics in Fig. 4.17c, comparing the value of the dominant opinion $a_{o_{max}}$ vs. $time$, for the Apollonian and the regular network with $\eta = 0.4$.

Above the critical noise, there is no consensus and the fraction of agents with dominant opinion is $\langle D \rangle \approx 1/2$. The response of the system is similar for scale-free and regular networks, as is shown in Fig. 4.17c with $\eta = 0.6$.

In order to illustrate the behavior of the dynamics as function of the noise, we display each of the nodes i of the Apollonian network in a plane, and represent the value of $a^{(i)}_{o_{max}}$ in the color scale presented in Fig.4.18. For $\eta = 0.2$ in Fig. 4.18a we see a snapshot before reaching stationary state of consensus, which is observed in Fig 4.18b. In Fig. 4.18a there are two regions in space with $a \approx 1$. Finally the region with a larger fraction of agents dominates and consensus ($D = 1$) can be reached with $a_{o_{max}} = 0.89$ (Fig. 4.18b). In Fig. 4.18c for a larger noise of $\eta = 0.4$ still below the transition, the value of the dominant opinion is lower, $a_{o_{max}} = 0.70$, and almost consensus is observed with a high fraction of agents, $\langle D \rangle = 0.98$, sharing the same opinion. In contrast, for $\eta = 0.6$, $a_{o_{max}} = 0.55$, is shared only by approximately half of the agents.

4.3.2 Summary

The response of the system to reach consensus originates in the model dynamics as opposed to the particular features of the network. An important characteristic of the transition to consensus is the dimension associated to the space of agent interactions. The dynamical response of the opinion model for both scale-free networks is similar to the one observed for MF interactions and each of these networks represents long range interactions. In contrast, differences are reported with a regular lattice, which has spatial dimension two, i.e. nearest neighbors interactions.

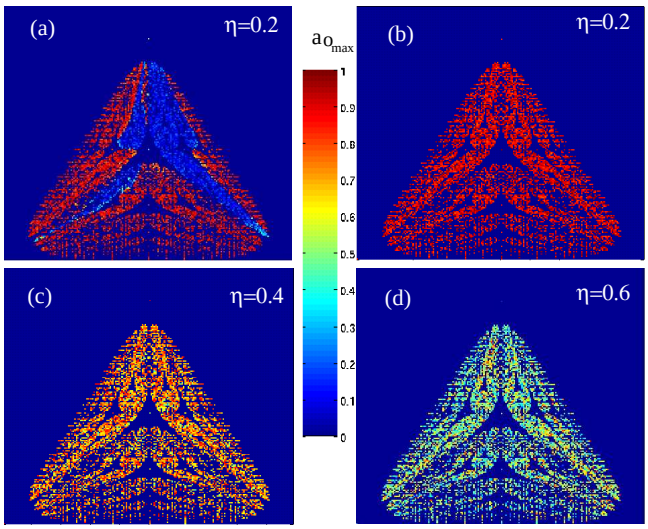

Figure 4.18: Snapshots of the value of the dominant opinion $a_{o_{max}}$ for different values of η on the Apollonian network with 9844 nodes and 2 opinions. (a) With $\eta = 0.2$, before reaching consensus there are two competing regions with different dominant opinions $a \approx 1$. Note that only the value of one opinion is shown and the darker nodes have the other opinion as dominant. (b) Snapshot for the same noise as in (a) when consensus ($D = 1$) with $a_{o_{max}} = 0.89$ is reached. (c) Below the critical noise with $\eta = 0.4$, the great majority of the nodes ($D = 0.98$) have the same opinion $a_{o_{max}} = 0.75$. (d) Above the critical noise with $\eta = 0.6$, no consensus is observed, $a_{o_{max}} = 0.62$ and is shared by half of the population $\langle D \rangle = 0.5$.

As was previously observed for the Sznajd model of opinion formation, for the general model that we present here, the response of the system in terms of opinion formation is qualitatively the same for a deterministic scale free network, as for a random scale free network. This implies a clear advantage for an analytical treatment of this model in a similar way as was done for the Sznajd model in section 4.2 [70].

Chapter 5

Spreading dynamics in a system of mobile agents

In this chapter we investigate the influence of agent mobility in spreading of infections. For a two-dimensional system of agents modeled by molecular dynamics, we simulate epidemics spreading. Our resulting network model is time-evolving. In section 5.1, we study the transitions to spreading as function of density, temperature and infection time. In addition, we analyze the epidemic threshold associated to a power-law distribution of time of infection.

Clusters of infected individuals are defined from data from health laboratories, but this quantity has not been defined and characterized for epidemics models on statistical physics. In section 5.2, we show that all the moments of the cluster size distribution at the critical rate of infection are characterized by only one exponent, which is the same exponent that determines the behavior of the total number of infected agents. No giant cluster survives independent on the magnitude of the rate of infection.

5.1 Scaling of the propagation of epidemics

The statistical spreading of infections, information or damage, involves non-equilibrium phenomena. Fluctuations and spatial correlations play an important role and are often exactly not solvable. Usually these processes are studied on a

lattice that can be regular [41, 71], hierarchical or small-world [42]. But in most cases the population in question is mobile. Therefore, in this work we study a system of particles moving according to a simple dynamics. We simulate on it a known contact process [72], described in terms of a 'SIS' model of infection, or infection without immunization, i.e. particles are either healthy or infected, and are susceptible to re-infection after healing, thus the name of the model (SIS: susceptible-infected-susceptible).
We characterize the transition to spreading of the epidemic dynamics, and obtain a continuous range of critical exponents changing the density of the system. The observed behavior results to be the 'SIS' analogous of a model of 'stirred percolation' [73–75] which was used to describe epidemic dynamics with immunization ('SIR').

5.1.1 Model

The simulations are carried out on a square shaped cell of linear size L with periodic boundary conditions. $L = \sqrt{\rho/N}$, is given by the number of particles (N) and the density (ρ). The particles are represented by 'soft-disks' of radius r_0 moving continuously on the plane. The interaction between two particles at positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ corresponds to a Lennard-Jones potential truncated at its minimum,

$$u(\boldsymbol{r}_i, \boldsymbol{r}_j) = U_0 \left[\left(\frac{2r_0}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \right)^{12} - 2 \left(\frac{2r_0}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \right)^{6} \right] + U_0, \;\; |\boldsymbol{r}_i - \boldsymbol{r}_j| \leq 2r_0, \tag{5.1}$$

reduced units are used in which U_0, r_0 , k_B (the Boltzmann constant) and m (the particle mass) are all unity.
Along this work, the particles are considered to be 'agents', to model the 'SIS' process described above. In a simple version one can assume that at each time two agents (i, j) interact or 'collide' (that is, if $|\boldsymbol{r}_i - \boldsymbol{r}_j| \leq 2r_0$), the infection propagates from an infected agent to a susceptible one. In most of our simulations we used a simple initial condition: (i) at time $t = 0$, N agents are distributed regularly in the cell, (ii) have the same absolute velocity v with randomly distributed directions, (iii) a central agent is infected and the rest are susceptible. At each time step the positions $\{\mathbf{r_i}\}$, velocities $\{\mathbf{v_i}\}$ and the infection state $\{\sigma_i\}$ of the system is updated. We use a molecular dynamics (MD) scheme of cell subdivision with the leapfrog integration method [76]. Once an agent is infected, it heals and becomes susceptible again after a fixed number of time steps, that is the 'time of the infection' (Δt_{inf}), and is a free parameter of the model.

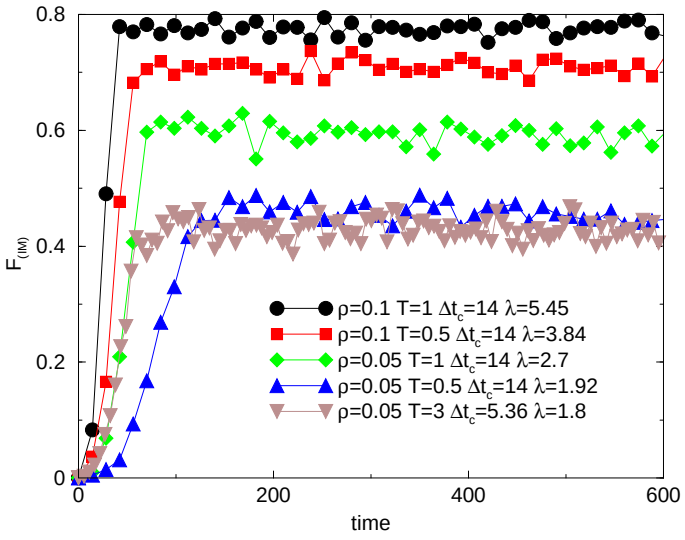

Figure 5.1: Starting with one infected agent, the figures shows how the infection spreads with time and after a certain period of time ('transient'), the fraction of infected agents (F_{IM}) fluctuates around a mean value. The fraction of infected agents in this 'quasi-equilibrium' state, increases with the infection rate (λ).

5.1.2 Scaling

The dynamics of an epidemics is described in terms of the infection rate (λ). That is defined as the number of agents one agent infects before healing. Therefore, in this model

$$\lambda \equiv \Delta t_{inf}/\tau_{coll}, \tag{5.2}$$

where τ_{coll} is the mean time between two collisions, and depends on the mean velocity of agents ($\langle v \rangle$). Neglecting the interaction potential with respect to the kinetic energy, one has:

$$\langle v \rangle = \sqrt{\frac{k_B T \pi}{m}}. \tag{5.3}$$

Thus, the mean number of collisions ($\langle n_{coll} \rangle$) of one agent during a period of time t, is given by the area within which it interacts ($2r_0\langle v \rangle$t), multiplied by the density,

$$\langle n_{coll} \rangle = \rho \; 2r_0 \langle v \rangle t. \tag{5.4}$$

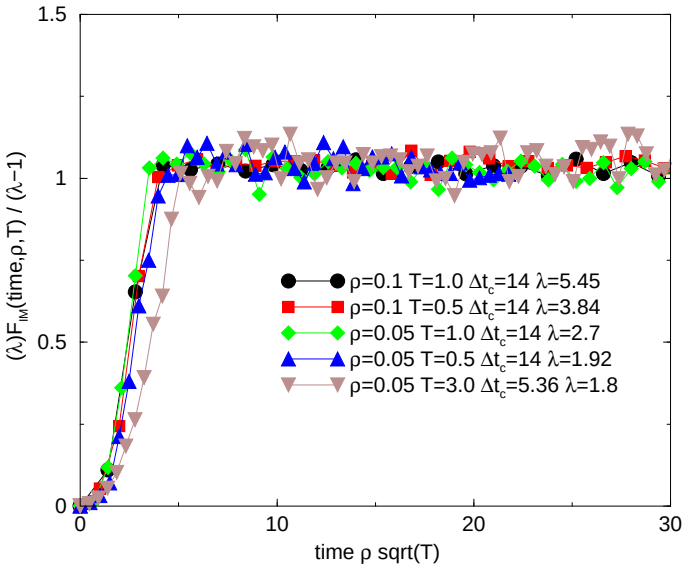

Figure 5.2: The plot shows the collapse of data of Fig. 5.1. The vertical axis was divided by the mean field approximation of $F_{IM}(\lambda)$ and the horizontal axis is divided by the collision time (τ_{coll}).

The pre-factor in Eq. (5.4) is the collision frequency, and its inverse, gives the mean time between collisions,

$$\tau_{coll} = \frac{1}{\rho\ 2r_0}\sqrt{\frac{m}{\pi T k_B}} \tag{5.5}$$

In Fig. 5.1 we see the fraction of infected agents ($F_{IM}(t)$ vs. $time$) for different values of temperature (T), density (ρ) and time of infection (Δt_{inf}). In each case we start with one infected agent and after a transient, the system fluctuates around a value F_{IM}, which depends on λ. This mean value can be calculated using a mean field approach,

$$\frac{\partial F_{IM}(t)}{\partial t} = -F_{IM}(t) + \lambda F_{IM}\left[1 - F_{IM}\right] \tag{5.6}$$

The first term of the r.h.s is the fraction of agents that heals and the second term, is the fraction of agents that becomes infected. After the transient, $\partial F_{IM}(t)/\partial t \sim 0$. Thus

$$F_{IM}(\lambda) = \begin{cases} 0 & \text{if } \lambda \leq \lambda_c \\ 1 - 1/\lambda & \text{if } \lambda > \lambda_c \end{cases} \tag{5.7}$$

where $\lambda_c = 1$, is known as the critical rate of infection. In Fig. 5.2 we see the collapse of the realizations of Fig. 5.1, with the expression obtained in Eq. (5.7). The dynamics of the system is characterized by the infection rate λ, which contains all the free parameters of the system.

5.1.3 Transition to spreading

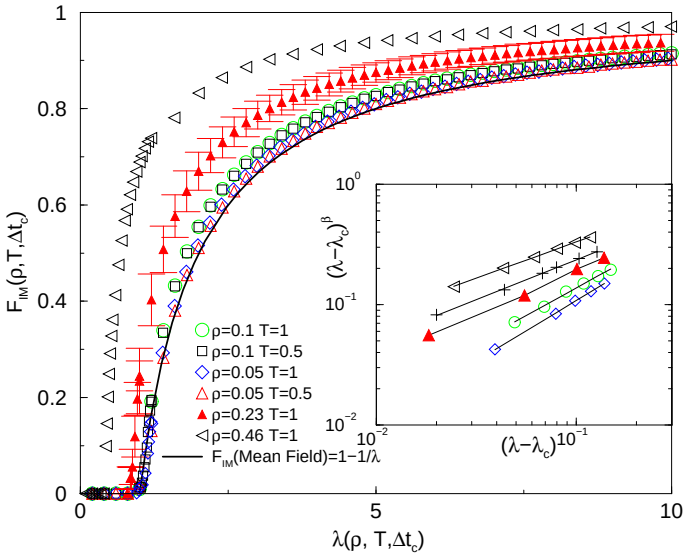

Figure 5.3: Fraction of infected nodes F_{IM} as function of λ for different conditions of T and ρ. The data agree with the MF approximation (Eq. (5.7)) only for low densities (< 0.1). The inset shows the $\log - \log$ plot of F_{IM} vs. $\lambda - \lambda_c$ with $T = 1$ and increasing densities: $\rho = 0.05$, $\lambda_c = 1.05$ (diamonds), $\rho = 0.1$, $\lambda_c = 1.06$ (circles), $\rho = 0.23$, $\lambda_c = 0.862$ (filled triangles), $\rho = 0.30$, $\lambda_c = 0.73$ (crosses), and $\rho = 0.46$, $\lambda_c = 1.05$ (rotated triangles). The circles are linear fits of the data and have respectively slopes: $\beta = 1.0$, $\beta = 0.92$, $\beta = 0.738$, $\beta = 0.66$, and $\beta = 0.599$.

We analyze the transition to spreading of the epidemic. In systems of 1254 agents after the transient F_{IM} is averaged, over 500 initial conditions at a given λ. In Fig. 5.3 we see that the results agree with the mean field approximation for densities lower than 0.1. Increasing the density, changes the shape of

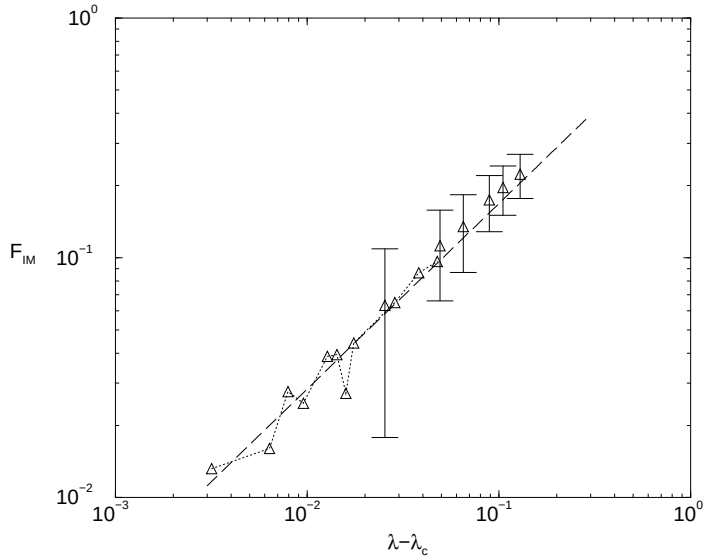

Figure 5.4: $Log - log$ plot of the fraction of infected nodes F_{IM} as function of $\lambda - \lambda_c$ for ($\rho = 0.2$, $T = 1$), with $\lambda_c = 0.935$. The dashed line is a fit to the form $F_{IM} \sim (\lambda - \lambda_c)^{\beta}$, with exponent $\beta = 0.773$.

the transition curve. Near the critical point the F_{IM} follows a power law, $F_{IM} \sim (\lambda - \lambda_c)^{\beta}$. The inset of Fig. 5.3 in a double-logarithmic plot, demonstrates this power law behavior, with straight lines of slope β. As can be seen, the value of β depends on the density of the system. Smaller values of β indicate more significant changes of $F_{IM}(\lambda)$ near the transition. Even for low densities (i.e. $\rho \in [0.1, 0.2]$), where one expects MF to be valid, the critical exponents change with the density.
In order to confirm this observation, we study the transition in detail. Fig. 5.4 shows the $\log - \log$ of F_{IM} vs. $(\lambda - \lambda_c)$, averaging over 1000 realizations in systems with $\rho = 0.2$, $T = 1$, and $\lambda_c = 0.935$. We see that the data fit the expression $\sim (\lambda - \lambda_c)^{\beta}$, with exponent $\beta = 0.773$.

Other critical exponents, are obtained, if one averages the survival probability of infection $P(t)$, the number of infected agents $n(t)$ and the square distance of spreading $R^2(t)$ and plots them against time. At the critical point, they are expected to display asymptotic power laws [77],

$$P(t) \sim t^{-\delta}, \qquad n(t) \sim t^{\eta}, \qquad R^2(t) \sim t^{z}. \tag{5.8}$$

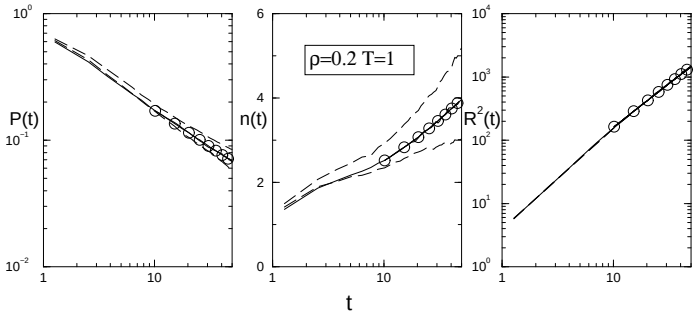

Figure 5.5: Evolution of the survival probability of infection $P(t)$, the mean number of infected agents $n(t)$ and the mean square distance of spreading $R^2(t)$ in time. Each graph contains three curves near criticality. With $\rho = 0.2$, from bottom to top: $\lambda = 0.93$, 0.94 and 0.95. The circles are linear fits to calculate the critical exponents

	MF	$\rho = 0.1$	$\rho = 0.2$	Contact Process(2D)[41]
λ_c	1	1.0(8)	0.9(4)	1.6488(1)
β	1	0.9(2)	0.7(7)	0.583(4)
δ	1	0.5(9)	0.5(3)	0.4505(10)
η	0	0.1(5)	0.2(5)	0.2295(10)
z	1	1.3(0)	1.2(7)	1.1325(10)

Table 5.1: Critical rate of spreading and exponents for SIS model on moving agents with $\rho = 0.1$ and 0.2, contact process in two dimensions, and estimates obtained by mean field MF.

The relations in Eq. (5.8) apply at long times, and require that the infection does not reach the boundaries of the system. Results for the three quantities $P(t)$, $n(t)$ and $R^2(t)$, averaged over $\sim 10^3$ realizations, with systems of $\sim 10^4$ agents, and fixed temperature ($T = 1$), are shown in Fig. 5.5 for $\rho = 0.2$.

The values of the critical exponents are reported in table 5.1. The known hyperscaling relation of the dynamics $4\delta + 2\eta = dz$, where d is the dimension of the system, is recovered within the range of numerical errors. Increasing the density of the system we find a continuous change in the critical exponents of the epidemic dynamics, they go from MF values to the exponents for contact process on a two dimensional lattice [41]. Numerical estimates of the critical exponent β vs. ρ are shown in Fig. 5.6.

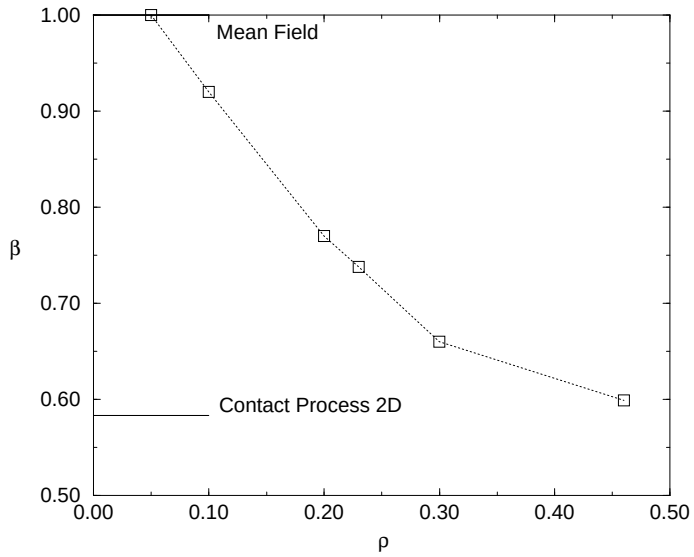

Figure 5.6: Numerical estimates for the critical exponent β, for systems with different density.

The dependence of the critical exponents of the dynamics on the density of agents (ρ), is analogous to the dependence on the 'flammable' fraction of space (ϕ), observed in 'forest fire' dynamics, or epidemics dynamics with immunization [73, 75], described through 'stirred percolation' models. 'Stirred percolation' consists in random walkers transmitting charge at collisions and was proposed by de Gennes [74] to describe the behavior of conductivity in binary mixtures [78].

5.1.4 Power-law distribution of infection times

Considering the same value of infection times (Δt_{inf}) for each agent corresponds to situations with homogeneous connectivity. In order to extent the model to some real world situations where the number of contacts varies greatly from one agent to another, we assign Δt_{inf} to each agent following a power-law distribution,

$$P(\Delta t_{inf}) = (\gamma - 1)\Delta t_{inf}^{-\gamma} \qquad \Delta t_{inf} \geq 1. \tag{5.9}$$

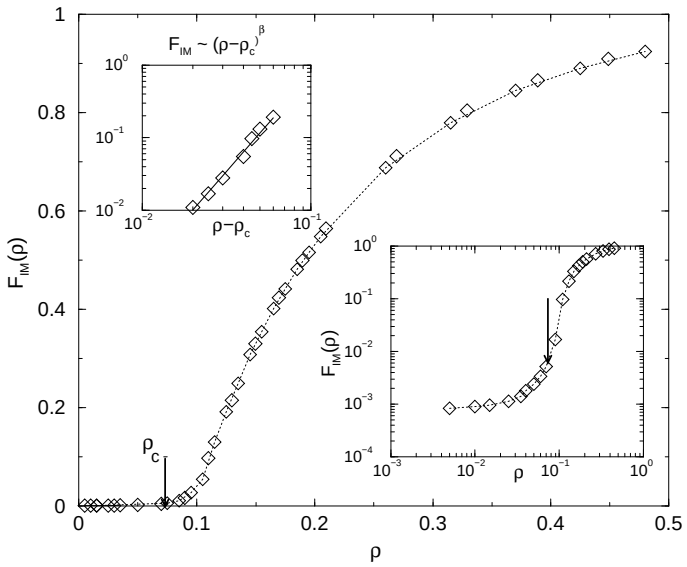

Figure 5.7: Fraction of infected nodes vs. density, with power-law distribution of infection times (Eq. 5.9) for $\gamma = 2.4$. $F_{IM} \sim (\rho - \rho_c)^{\beta}$ is shown in the upper-left-corner inset with $\rho_c = 0.065$ and $\beta = 2.6(5)$. The bottom-right-corner inset is the main plot in $\log - \log$ scale to see better the tail of the spreading curve, for $\rho < \rho_c$.

As a result, for $2 < \gamma \leq 3$, the epidemic threshold tends to zero, like has been observed for SF networks [42] (see Fig. 5.7).
However, the shape of the spreading curve has a point of inflection (ρ_c) above which, the infection is much larger ($F_{IM} \sim (\rho - \rho_c)^{\beta}$, $\rho_c = 0.06(5)$, $\beta = 2.65$, for $\gamma = 2.4$).
We see that the more 'connected' agents are responsible for the absence of an epidemic threshold (tail of the curve). The infection would spread only if the *mean* rate of spreading is larger than unity, that is

$$\lambda_c \equiv \frac{\langle \Delta t_{inf} \rangle}{\tau_{coll}} = \frac{\gamma - 1}{\gamma - 2}\ \rho\ 2r_0 \sqrt{\frac{\pi T k_B}{m}} > 1. \tag{5.10}$$

$\lambda_c = 1$ gives the inflection point ρ_c, which for $\gamma = 2.4$ is $\rho_c = 0.0718$ (according to Eq. 5.10, for $r_0 = 2^{1/6}$ and $T = 1$), and agrees with the numerical value $\rho_c = 0.06(5)$, reported in Fig. 5.7 averaging over about 10^3 realizations with about 10^3 agents each.

5.1.5 Summary

Novel effects are observed studying the SIS model of infection on a system of mobile agents. A continuous range of critical exponents is observed as function of the density of agents, recovering mean field predictions for lower densities, two-dimensional exponents of contact process, increasing the density. Introducing a power-law distribution of infection times, the epidemic threshold vanishes due to the more 'infecting agents'; but still there is a *critical* rate of spreading, which depends on the exponent of the distribution and the mean time of collision among the agents.

5.2 Cluster size distribution of infection

In a small and highly urbanized nation like Singapore dengue outbreaks or epidemics are identified as "clusters". A dengue **cluster** or focus of transmission is defined as at least two confirmed cases, with no recent travel history, that are located within 200 m of each other (taken as the flight range of the *Aedes aegypti*) and whose dates of the onset of symptoms are within three weeks of each other [79]. Most of the efforts have been directed towards 'SIS' [42, 72, 77] models of infection, which was presented in the previous section. Analytical and numerical expressions describe the dynamics of the SIS model in terms of the rate of spreading λ, the evolution of the survival probability of the infection $P(t)$, the mean number of infected agents $n(t)$ and the mean square distance of spreading $R^2(t)$ in time, which are quantities difficult to compare with real data of epidemics. This chapter suggests comparison with public health data, analyzing a scaling function for **clusters numbers** on a SIS model of infection.

An important ingredient of our work is the mobility of agents, contrasted to most of the models of epidemics where the population is modeled by static networks [42, 72, 77]. As we showed in the previous chapter [80] the critical exponents, which depend on the density of the system, i.e spatial correlations and mobility of the agents play an important role. Here we use the model of mobile agents to define clusters of infections and analyze its dependency on the rate of infection λ and on the mobility of the agents.

We find that the network of clusters of infections remains disconnected and no matter how large the rate of infection, no giant cluster is formed. We show that

in the transition to spreading, the moments of the cluster size distribution are described by an exponent β, which is the exponent that characterizes the fraction of infected mass $F_{IM} = N_{Inf}/N$, defined as the ratio of the number of infected agents (N_{Inf}) and the total amount of population (N). Thus the number of clusters depends on λ, and mobility and spatial correlation of the agents influence this dependency.

5.2.1 Method

Our objective is to characterize the behavior of the clusters of infected individuals. When agent j infects agent i a link is created among them, the link lasts until one of them heals, meanwhile each of them continues making links with other susceptible agents through the same rule.

A cluster is thus defined as a group of infected agents connected by links. Note that in contrast to percolation, where clusters are given by occupied lattice sites connected by nearest-neighbor distances, for this model each cluster gives a group of agents infected in a given period of time linked by a relation of contagion. Isolated infected agents are regarded as clusters of size unity and any cluster consisting of s connected agents is an $s - cluster$. We borrow the notation from Stauffer's book on percolation theory [81] and define here $n_s = N_s/N$ as the number of s-clusters per agent, where N_s is the number of clusters of size s and N the total number of agents in the system. For different values of λ, in the next section we present the results of the calculation of the first three moments of the cluster size distribution. Namely: $\sum_s n_s$, $\sum_s s n_s$, $\sum_s s^2 n_s$. Those quantities give us, respectively, information about: the total number of clusters, the fraction of infected agents and the mean size of clusters. In order to keep the analogy with percolation, we sum over all values of s excluding the largest cluster (S_{major}). We also present, the calculations of $F_{major} = S_{major}/N$, the fraction of agents that belong to the largest cluster and $F_{IM} = N_{inf}/N$, the fraction of agents that are infected.

5.2.2 Results

For a fixed density, we vary λ (Eq. 5.10) changing the time of infection (Δt_{inf}). Starting with half of the population infected, for rate of infections near λ_c, a given trial may end in the absorbing state after a few time steps or it may *survive*

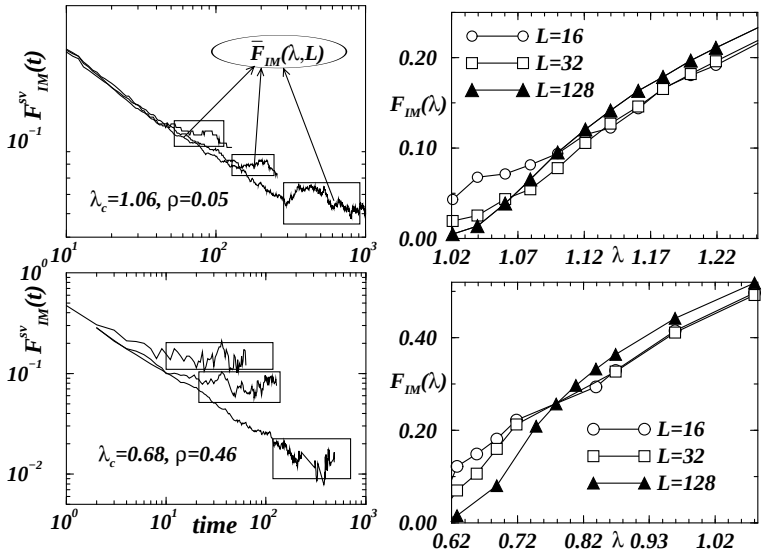

Figure 5.8: Left: Fraction of infected individuals from surviving trials versus time at $\lambda = \lambda_c$, starting with half of the population infected. At the top, the results for $\rho = 0.05$ and $\lambda = 1.06$ and at the bottom $\rho = 0.46$ and $\lambda = 0.68$. System sizes $N = 32 \times 32$, 64×64, 128×128 (from top to bottom). Right: Quasi-stationary fraction of infected agents versus λ for the same densities (Top: $\rho = 0.05$. Bottom: $\rho = 0.46$).

fluctuating with a quasi-stationary fraction of infected agents, marked by windows in the left-side of Fig 5.8. The calculations are made averaging on time at the *quasi-stationary state*, which is described by the surviving trials following an initial transient. The number of time steps of this transient depends on λ and on the system size L (see left side of Fig. 5.8). The data here illustrate how the mean fraction of infected agents $F^{sv}_{IM}(t)$ (the superscript denotes an average restricted to surviving trials) approaches its stationary value $\bar{F}_{IM}(\lambda, N)$ (in the following, we write $\bar{F}_{IM}(\lambda, N)$ just like $F_{IM}(\lambda)$). On the right side of the same figure we see the graph of $F_{IM}(\lambda)$, for three system sizes. We analyze in detail the number of clusters for the two density values $\rho = 0.05$ and $\rho = 0.46$, which have critical rate of spreading $\lambda_c = 1.06$ and $\lambda_c = 0.68$ respectively. Note that at the critical density λ_c, surviving trials tend to stationary values only in the limit $L \to \infty$.

The top of Fig. 5.9 is only for pedagogical reasons, in order to illustrate how the

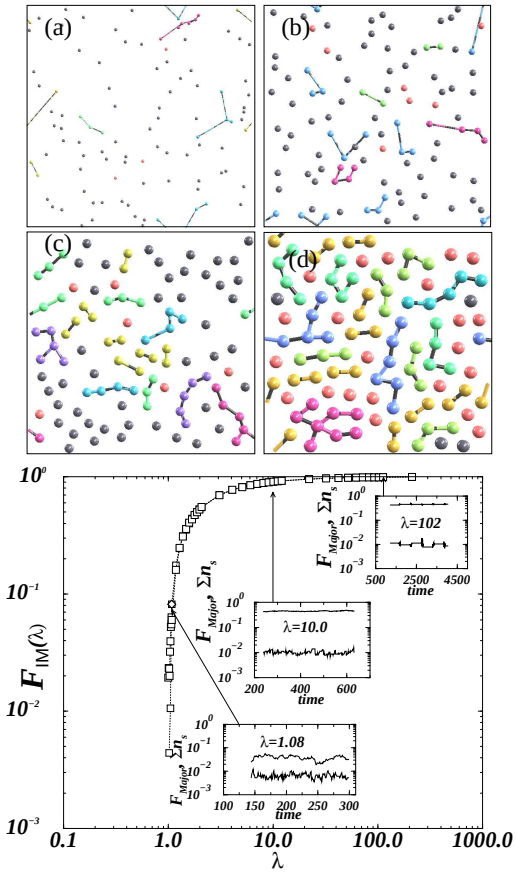

Figure 5.9: Top: Snapshots of cluster sizes of infected agents for systems with different densities: (a)$\rho = 0.05$, (b)$\rho = 0.20$, (c)$\rho = 0.40$ and (d)$\rho = 0.80$, in all cases $\lambda = 1.5$. Bottom: Quasi-stationary fraction of infected agents varying λ over three orders of magnitude (Average over 20 realizations for $\rho = 0.05$ and $N = 32 \times 32$). The insets show the fraction of infected agents in the largest cluster (lower value) and the first moment of the cluster size distribution (upper value) vs. time, at $\lambda = 1.08$, $\lambda = 10.0$ and $\lambda = 108.0$.

number of clusters looks in the quasi-stationary state, we see snapshots of the clusters of infections for *different* system densities and the *same* rate of infection $\lambda = 1.5$, here $N = 10 \times 10$.

For $\rho = 0.05$ and $\lambda \in [1, 200]$, the bottom of Fig. 5.9 shows the variation of $F_{IM}(\lambda)$ and $N = 32 \times 32$ averaged over 20 different realizations. The insets show the change in time of F_{major} and $\sum_s n_s$, for only one realization with $\lambda = 1.08$, $\lambda = 10.0$ and $\lambda = 108$. In contrast to percolation, in this model there is no significant variation of F_{major} with λ, and the relation $F_{major} \ll F_{IM}$ remains. Moreover, the number of clusters $\sum_s n_s$ grows considerably only near λ_c. In

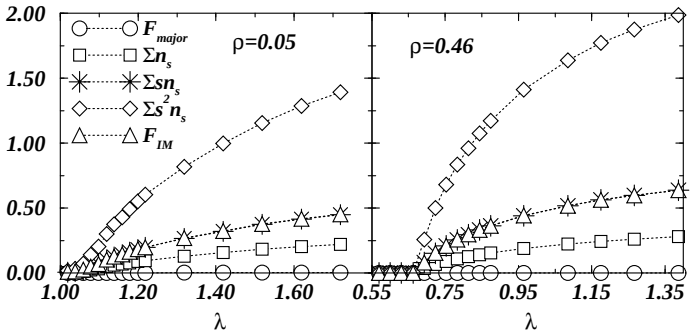

Figure 5.10: First three moments of the cluster size distribution , fraction of agents in the largest cluster (F_{major}) and fraction of infected agents (F_{IM}) vs. λ. Average over 50 trials, system size $N = 64 \times 64$. Left:$\rho = 0.05$. Right:$\rho = 0.46$.

Fig. 5.10 for $\rho = 0.05$ and $\rho = 0.46$, we plot the behavior of the cluster numbers near their respective λ_c. As the largest cluster remains small compared to the total number of agents ($S_{major} \ll N$), we have $F_{IM}(\lambda) \sim \sum_s sn_s$. Additionally one can see that $\sum_s sn_s$ and $\sum_s s^2 n_s$ show the same critical behavior as $F_{IM}(\lambda)$, plotted in detail in Fig. 5.11. We observe that all the moments of the cluster size distribution present exactly the same critical behavior than F_{IM}, namely $(\lambda - \lambda_c)^\beta$, where β depends on the density of the system.

5.2.3 Summary

This section showed that the cluster size distribution of infected individuals is described in terms of the spreading rate (λ) and the same exponents (β) previously known for the total mass of infection. Although the agents are free to move

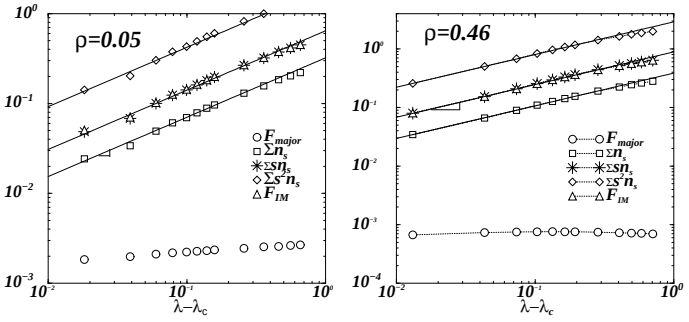

Figure 5.11: Same results of Fig. 5.10 plotted vs. $(\lambda - \lambda_c)$. The solid lines are regressions of the form $m_i(\lambda - \lambda_c)^\beta$ with m_i the coefficient of the ith moment. Left: $\lambda_c = 1.06$, $\beta = 0.66$, $m_0 = 0.321$, $m_1 = 2m_0$, and $m_2 = 6m_0$. Right: $\lambda_c = 0.68$, $\beta = 0.56$, $m_0 = 0.386$, $m_1 = 2.3m_0$, and $m_2 = 7.5m_0$

there is a homogeneous size distribution of infected clusters at the critical rate of infection, and we did not find any critical exponent associated to the cluster sizes. Comparing with the traditional SIS model on a static network we confirm that mobility and spatial correlations change the value of the critical exponent β of the fraction of the infected population, and to the same extent the cluster size distribution.

Chapter 6

Dynamical model of growing social networks

In section 6.1 we present a novel model to simulate real social networks of complex interactions, based in a system of colliding particles (agents). The network is build up by keeping track of the collisions and evolves in time with correlations which emerge due to the mobility of the agents. Therefore, statistical features are a consequence only of local collisions among its individual agents. Agent dynamics is realized by an event-driven algorithm of collisions where energy is gained as opposed to physical systems which have dissipation. The model reproduces empirical data from networks of sexual interactions, not previously obtained with other approaches.

Next, in section 6.2, we investigate the clustering coefficient in bipartite networks where cycles of size three are absent and therefore the standard definition of clustering coefficient cannot be used. Instead, we use another coefficient given by the fraction of cycles with size four, showing that both coefficients yield the same clustering properties. The new coefficient is computed for the empirical networks of sexual contacts, one bipartite and another where no distinction between the nodes is made (monopartite). In both cases the clustering coefficient is similar. Furthermore, combining both clustering coefficients we deduce an expression for estimating cycles of larger size, which improves previous estimates and is suited for either monopartite or multipartite networks, and discuss the applicability of such analytical estimates.

6.1 Model of mobile agents for sexual interactions networks

A social network is a set of people, each of whom is acquainted with some subset of the others. In such a network the nodes (or vertices) represent people joined by edges denoting acquaintance or collaboration. Empirical data of social networks include networks of scientific collaboration[29], of film actor collaborations[13], friendship networks[33] among some others[2]. One kind of social network is the network of sexual contacts[82–85], where connections link those persons (agents) that have had sexual contact with each other. The empirical investigation of such networks are of great interest because, e.g. the topological features of sexual partners distributions help to explain why persons can have the same number of sexual partners and yet being at distinct risk levels of contracting HIV [86].

The simplest way to characterize the influence of each individual on the network is through its degree k, the number of other persons to whom the individual is connected. Sexual contact networks are usually addressed as an example of scale-free networks[1, 83–85], because of having a tail in its degree distribution, which is well fitted by a power-law $P(k) \sim k^{-\gamma}$, with an exponent γ between 2 and 3. However, another characteristic feature, not taken into account, is that the small k-region, comprehending the small-k values varies slowly with k, deviating from the power-law. Moreover, the size of the small-k region also increases in time, yielding rather different distributions when considering the number of partners during a one year period or during the entire life, e.g. for entire-life sexual contacts, the degree distribution shows that at least half of the nodes have degree in the small-k region [82, 83]. A model predicting all these different distributions shapes for different time spans is of crucial interest, because the transmission of diseases occur during the growth of the network.

One of the main difficulties for validating a model of sexual interactions is that typical network studies of sexual contacts involve the circulation of surveys, or anonymous questionnaires, and only the number of sexual partners of each interviewed person is known, not being possible to obtain information about the entire network, in order to calculate degree correlations, closed paths (cycles), or average distance between nodes.

In this section we propose a model of mobile agents in two dimensions from which the network is build by keeping track of the collisions between agents, representing the interactions among them. In this way, the connections are a result not of some *a priori* knowledge about the network structure but of some

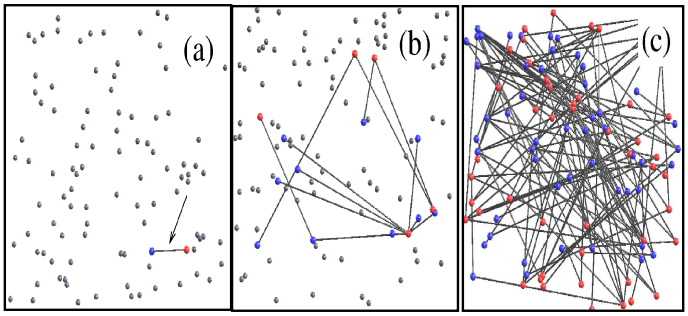

Figure 6.1: Snapshots of the growing network of collisions in a low-density gas with $N = 100$, for **(a)** $n = 0.02N$, **(b)** $n = 0.15N$, **(c)** $n = N$. Two colors represent the two sexual genders and larger symbols emphasize those which belong to the network (linked agents).

local dynamics of the agents from which the complex networks emerge. Below, we show that this model is suited to reproduce sexual contact networks with degree distribution evolving in time, and we validate the model using contact tracing studies from health laboratories, where the entire contact network is known. In this way, we are able to compare the number of cycles and average shortest path between nodes as well as compare the results with the ones obtained with Barabási-Albert scale-free networks [34], which are well-known models, accepted for sexual networks[83–85].

6.1.1 The Model

The model introduced below is a sort of gas [87], where N particles with small radius r represent agents randomly distributed in a two-dimensional system of linear size $L \gg \sqrt{N}r$ (low density) and the basic ingredients are an increase of velocity when collisions produce sexual contacts, two genders for the agents (male and female), and n/N, the fraction of agents that belong to the network, which constitutes an implicit parameter for the resulting topology of the evolving network.

The system has periodic boundary conditions and is initialized as follows: all agents have a randomly chosen gender, position and moving direction with the same velocity modulus $|\mathbf{v}_0(i)|$. We mark one agent from which the network will be constructed. When the marked agent collides for the first time with another

one of the opposite gender, the corresponding collision is taken as the first connection of our network and its colliding partner is marked as the second agent of the network (Fig. 6.1a). Through time, more and more collisions occur, increasing the size n of the network (Fig. 6.1b and 6.1c) till eventually all the agents composing the system are connected.

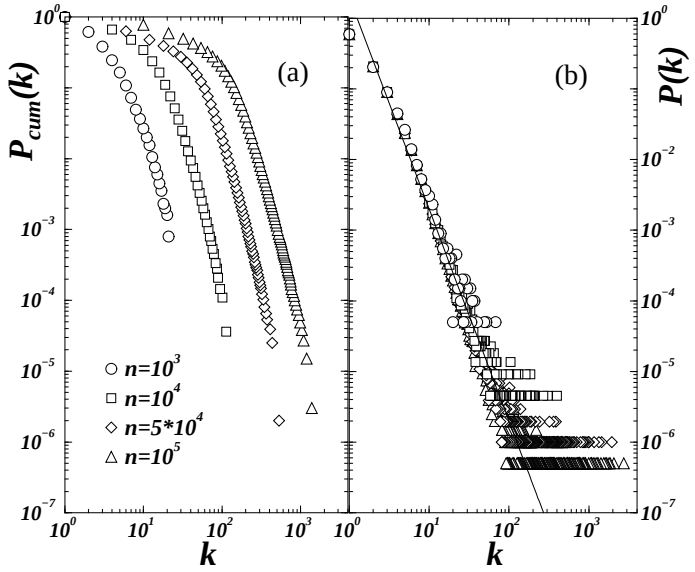

Figure 6.2: **(a)** Cumulative distribution $P_{cum}(k)$ of the number k of partners among agents, when considering type-(i) and -(ii) interactions (see text) for $n = 10^3$ (circles), $n = 10^4$ (squares), $n = 5 \times 10^4$ (diamonds) and $n = 10^5$ (triangles). For the same parameter values **(b)** shows a pure scale-free distribution, obtained when only type-(ii) interactions form links. The solid line indicates the slope $\gamma = 3$ of the scale-free distribution. Here $\alpha = 1$ and $N = 320 \times 320$.

Collisions between two agents take place whenever their distance is equal to their diameter and the collision process is based on an event-driven algorithm, i.e. the simulation progresses by means of a time ordered sequence of collision events and between collisions each agent follows a ballistic trajectory[76].

Since sexual interactions rely on the sociological observation[88] that individuals with a larger number of partners are more likely to get new partners, we choose a collision rule where the velocity of each agent increases with the number k

of sexual partners. The larger the velocity one agent has the more likely it is to collide. Moreover, contrary to collision interactions where the velocity direction is completely deterministic[89], here the moving directions after collisions are randomly selected, since in general, sexual interactions do not determine the direction towards which each agent will be moving afterwards. Therefore, momentum is *not* conserved.

Regarding these observations our collision rule for sexual interactions reads

$$\mathbf{v}(k_i) = (k_i^{\alpha} + |\mathbf{v}_0(i)|)\boldsymbol{\omega}, \tag{6.1}$$

where k_i is the total number of sexual partners of agent i, exponent α is a real positive parameter, $\boldsymbol{\omega} = (\mathbf{e}_x \cos\theta + \mathbf{e}_y \sin\theta)$ with θ a random angle and $\mathbf{e_x}$ and $\mathbf{e_y}$ are orthogonal unit vectors. Collisions which do not correspond to sexual interactions only change the direction of motion. It should be pointed out that the motion of agents occurs in a space which is not the common physical space, but rather a general continuous Euclidean space, whose norm is related with some 'social' distance, depending not only on the physical distance between each pair of agents, but also on their common features and acquaintances.

Collisions corresponding to sexual interactions, i.e. with a velocity update as in Eq. (6.1), are the only ones which produce links, and occur in two possible situations: (i) between two agents which already belong to the network, i.e. between two sexually initiated agents and (ii) when one of such agents finds a non-connected (sexually non-initiated) agent. For simplicity, we do not take into account sexual interactions between two non-connected agents, and therefore our network is connected (see the discussion in Sec. 6.1.3).

When interactions of type (i) and (ii) occur, both the distribution tail and the small-k region are observed, as shown by the cumulative distribution $P_{cum}(k)$ in Fig. 6.2a. Here, we use a system of $N = 320 \times 320$ agents with $\rho = 0.02$, $\alpha = 1$ and distributions are plotted for different stages of the network growth, namely $n = 10^3$, $n = 10^4$, $n = 5 \times 10^4$ and $n = 10^5 \sim N$. As one sees, the exponent of the power-law tail and the transition between the tail and the small-k region increase during the growth process. These features appear due to the fact that at later stages most of the collisions occur between already connected agents. Consequently, the average number of partners increases as well.

If one considers only type-(ii) sexual contacts, the system reproduces a stationary scale-free network, as shown in Fig. 6.2b. In this case the average number of partners, defined as[91] $\langle k \rangle = k_{min}(\gamma - 1)/(\gamma - 2)$ with k_{min} the minimum number of partners, is always 2 ($k_{min} = 1$ and $\gamma = 3$). As we show below, while empirical data of sexual contacts over large periods have distributions like

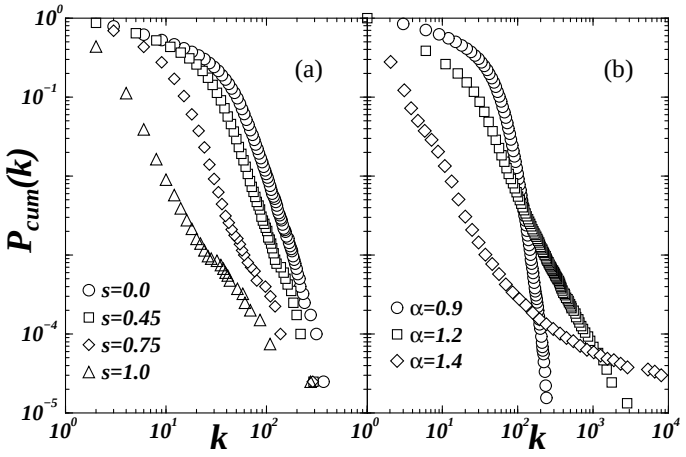

Figure 6.3: Cumulative distributions, when varying **(a)** a parameter s of selectivity which interpolates between Figs.6.2a and 6.2b (see text) for $\alpha = 1$ and **(b)** the exponent α in the update velocity rule, Eq. (6.1), for $s = 0$. Here a fixed stage of growth is considered, namely $n = 5 \times 10^4 = 0.5N$.

the ones for regime (i)+(ii), data for shorter periods ($1 - 10$ years) are scale-free (only (ii)).

With our model one can easily interpolate between both interaction regimes, (i)+(ii) and (ii), by introducing a parameter s of 'selectivity', defined as the probability that sexually initiated agents in case of collision with another initiated agent, have no sexual contact. Physically, this selectivity accounts for the intrinsic ability that a node has to select from all its contacts (collisions) the ones which are sexual. These intrinsic abilities were already used in other contexts, e.g. as a new mechanism leading to scale-free networks in cases where the power-law degree distribution is neither related to dynamical properties nor to preferential attachment[92]. For $s = 0$ one obtains the two regions illustrated in Fig. 6.2a, namely the small-k region and the power-law tail, while for $s = 1$ one obtains the pure scale-free topology illustrated in Fig. 6.2b. In Fig. 6.3a, we show the crossover between these two regimes.

The shape of the cumulative distributions is also sensible to the exponent α in the update velocity rule, Eq. (6.1), as shown in Fig. 6.3b. While for small values of $\alpha \lesssim 1$ one gets an exponential-like distribution, for $\alpha \gtrsim 1.4$ the distribution shows that a few nodes make most of the connections. Henceforth, we fix $\alpha =$

1.2.

Having described the model of mobile agents we proceed to a specific application, i.e. modeling empirical networks of sexual contacts.

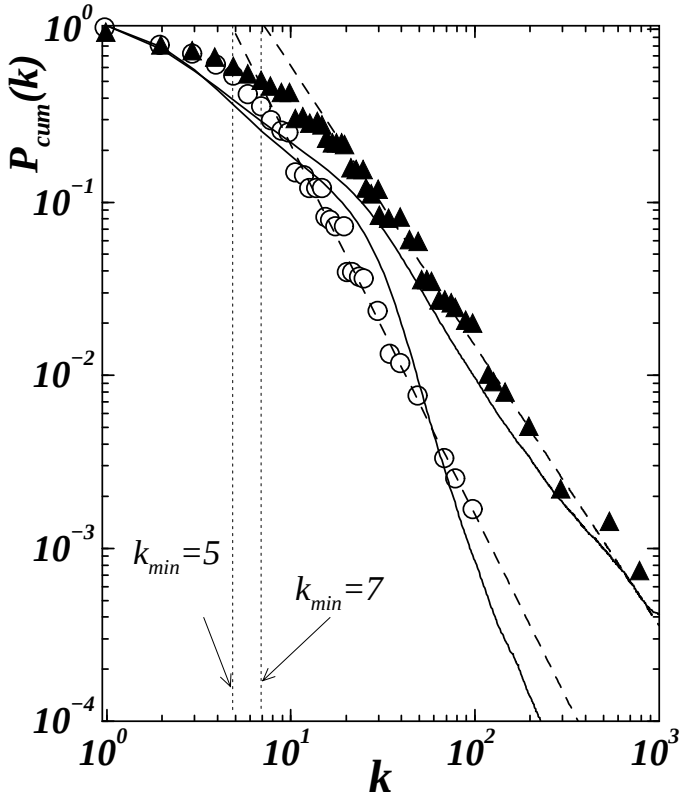

Figure 6.4: Cumulative distribution of sexual partners in a network of heterosexual contacts extracted from Ref. [83], where male (triangles) and females (circles) distributions are plotted separately, with a total of 2810 persons. Solid lines indicate the simulations when plotting the distributions at the same stage $n = 0.2N$, starting with a population composed by 58% of females and 42% of males. Here $N = 10^5$, $s = 0$

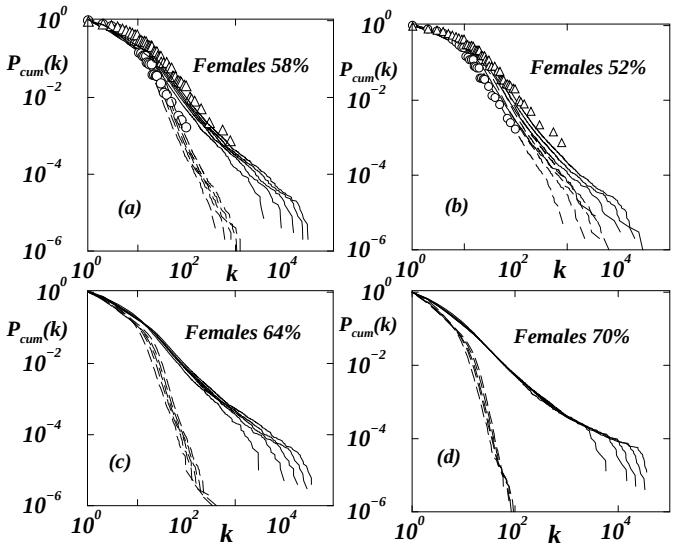

Figure 6.5: Distributions of sexual partners in a network of heterosexual contacts using different amounts of females: **(a)** 58% as in Fig. 6.4, **(b)** 52%, **(c)** 64% and **(d)** 70%. In each plot solid and dashed lines indicate the cumulative distributions of males and females respectively, for five different realizations. Clearly, the exponent of the power-law tail of the distributions decreases when the percentage of females or males increases (see dotted lines). Same conditions as in Fig. 6.4 were used.

6.1.2 Reproducing networks of sexual contacts

We will show that, by properly choosing the parameter values in our model, one can reproduce real data distributions of sexual contact networks. In Fig. 6.4 the cumulative distributions of a real contact network[83] are shown for females (circles) and males (triangles) separately, based on empirical data from 2810 persons from a Swedish survey of sexual behavior, where each person reported the number of sexual partners they had in a given period of time (frequency of sexual relations with each partner is neglected). The solid lines in Fig. 6.4 are the simulated distributions. The simulated power-law tails have exponents $\gamma_m = 2.4$ and $\gamma_f = 4.0$ for males and females respectively, compared with the empirical data $\gamma_m = 2.6 \pm 0.3$ and $\gamma_f = 3.1 \pm 0.3$[83]. To stress that, while the power-law tails are also well fitted by distributions obtained with scale-free networks

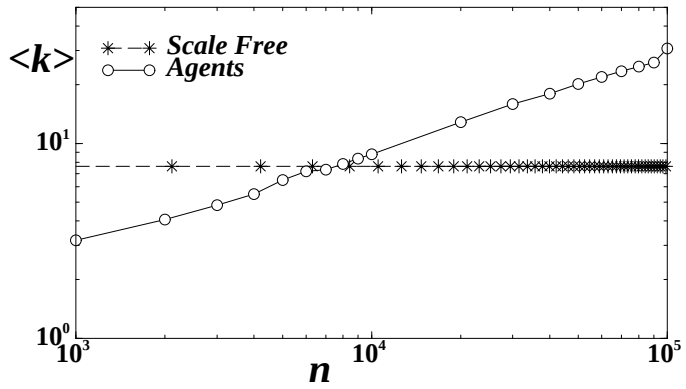

Figure 6.6: Comparing the average number of partners for scale-free networks (stars) and the agent model (circles). For the scale-free network $k_{min} = 4$ and for agent model $N = 520 \times 520$, $s = 0$.

(dashed lines in Fig. 6.4), these distributions have a minimum number of connections (partners) of $k_{min} = 5$ for females and $k_{min} = 7$ for males, contrary to the real value $k_{min} = 1$ also reproduced with our agent model. In fact, the model of mobile agents takes into account not only the power-law tail of these distributions, but also the small-k region which comprehends the significant amount of individuals having only a few sexual partners ($k \gtrsim 1$).

It is important to note that in order to reproduce the difference in the exponents it is necessary to have 58% females and 42% males, which is far from the expected difference among number of females and males in typical human populations, with ratios of females:males of the order of 1.1. The difference in the exponents of the distribution tails for males and females separately, present in the data of sexual surveys, has generated much controversy and is often considered due to a bias either of sampling or honest reporting (see Ref. [82] and references therein). The exponents γ_m and γ_f on a bipartite network are expected to be nearly the same when the percentage of males and females are similar, as shown in Fig. 6.5. In each plot five different realizations are shown for males (solid lines) and females (dashed lines). In Fig. 6.5a, we plot the results for the same conditions as in Fig. 6.4 (58% females and 42% males). Taking the average over the five curves for each gender yields the curves shown in Fig. 6.4. Figures 6.5b-d show the distributions obtained for other percentages illustrating that when decreasing the difference in the ratio females:males the difference in the exponents dissap-

pears. A characteristic feature of our model is that the average number $\langle k \rangle$ of

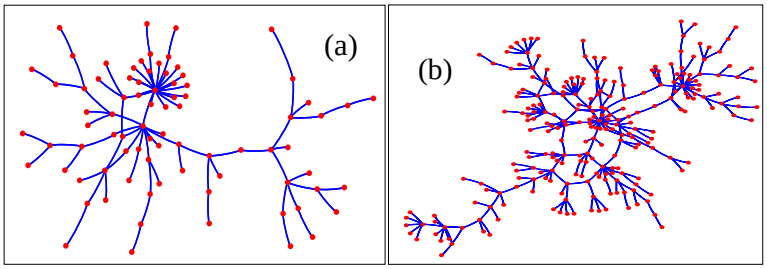

Figure 6.7: Sketch of two real sexual contact networks having **(a)** only heterosexual contacts ($N = 82$ nodes and $L = 84$ connections) and **(b)** homosexual contacts ($N = 250$ nodes and $L = 266$ connections). While in the homosexual network triangles and squares appear, in the heterosexual network triangles are absent (see Table 6.1).

partners increases as the network grows, which is expected to occur in real sexual networks according to the observed differences in the shape of the degree distribution for yearly and entire-life reports of number of sexual partners [83, 84]. This feature is not observed in scale-free networks, as illustrated in Fig. 6.6. Of course, that this growth also indicates non-stationary regimes, where $\langle k \rangle$ diverges with the network growth. In the section 6.1.3 we explain how to overcome this shortcoming.

We compare the model, with two empirical networks of sexual contacts. One network is obtained from an empirical data set, composed solely by heterosexual contacts among $n = 82$ nodes, extracted at the Cadham Provincial Laboratory (Manitoba, Canada) and is a 6-month block data [94] between November 1997 and May 1998 (Figure 6.7a sketches this network). The other data set is the largest cluster with $n = 250$ nodes in the records of a contact tracing study [93], from 1985 to 1999, for HIV tests in Colorado Springs (USA), where most of the registered contacts were homosexual (see Figure 6.7b).

Figures 6.8(a)-(b) show the cumulative distribution of the number of sexual partners for each of the empirical networks. For both cases the agent model and scale-free networks with $k_{min} = 1$ can reproduce the distribution of the number of partners. However, the agent model with $s = 0.7$ reproduces, as well, the clustering coefficient distribution that we measure from the empirical network.

The clustering coefficient $C(i)$ of one agent is defined [13] as the total number of

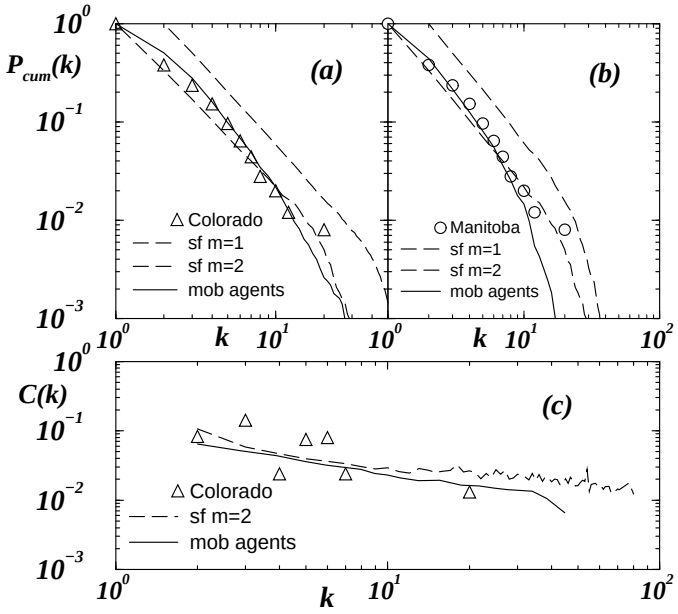

Figure 6.8: **(a)** Cumulative degree distribution of a homosexual contact network[93] with $n = 250$ (triangles). **(b)** Cumulative degree distribution of a heterosexual contact network[94] with $n = 82$ (triangles). Each case is compared with the average degree distribution over 20 iterations, for the BA scale-free model (dashed line) with $k_{min} = 1$, and $k_{min} = 2$ and with our agent model (solid line) with $s = 0.7$. **(c)** Cluster coefficient for the homosexual network empirical data (triangles), the agent model (solid line) and the BA model $k_{min} = 2$. The scale-free $k_{min} = 1$ yields $C(k) = 0$ (not shown).

triangular loops of connections passing at one node divided by the total number of connections k_i. Averaging $C(i)$ over all nodes with k_i neighbors yields the clustering coefficient distribution $C(k)$. While for the scale-free graph which better reproduces these empirical data, the clustering coefficient is zero, our agent model yields a distribution which resembles the one observed in the real network (Fig. 6.8c). This feature is due to the co-existence of a tree-like substructure and closed paths (see Figs. 6.7b).

For both heterosexual and homosexual networks of sexual contacts, the model of mobile agents reproduces other important statistical features, namely the average

	N	L	T	Q	C
Heterosexual	82	84	0	2	0
Homosexual	250	266	11	6	0.029
Heterosexual (Agent Model)	82	83.63	0	1.45	0
Homosexual (Agent Model)	250	287.03	8.23	10.52	0.023
Heterosexual (Null model)	82	84	0	8.47	0
Homosexual (Null model)	250	266	6.94	16.2	0.011
Homosexual (Null m., same T)	250	266	11.0	21.462	0.015
Heterosexual (Scale-free)	82	162	0	159.72	0
Homosexual (Scale-free)	250	498	45.28	256.79	0.082

Table 6.1: Clustering coefficients and cycles in two real networks of sexual contacts (top), one where all contacts are heterosexual and another with homosexual contacts. In each case we present the values of the number N of nodes, the number L of connections, the number T of triangles, the number Q of squares and the average clustering coefficient C. The values of these quantities are computed for networks constructed with the agent model, with two null models (see text), one where the number of triangles is fixed and another where this restriction is not imposed, and with scale-free networks with $k_{min} = 2$. For $k_{min} = 1$, one has better values for the number L of links in both the heterosexual and the homosexual networks, $L = 81$ and $L = 249$ respectively, but there are no cycles (not shown).

clustering coefficient C and the number of loops of a given order. Table 6.1 indicates the number T of triangles (loops composed by three edges), the number Q of squares (loops with four edges) and the average clustering coefficients C given by [13] the average of $C(i)$ over the entire network.

When using the agent model with the same number N of nodes as in the real networks we obtain similar results for L, T, Q and C, as shown in Table 6.1 (middle), where values represent averages over samples of 100 realizations. For

the heterosexual network there are no triangles due to the bipartite nature of the network. In order to ascertain possible non-trivial features in these empirical networks, we compare the topological measures of them with the ones of a null model having the same degree distribution. The null model is a randomized version of the empirical networks, constructed by rewiring randomly selected connections [90]. Namely, whenever one pair of links is selected, say $i \leftrightarrow j$ and $k \leftrightarrow l$, we substitute these links by two other ones, one connecting i and k and the other connecting j and l. Note that in the randomized versions the number of triangles and squares is much larger than in the real networks. In order to compare the number of squares without the effect of the number of triangles in the network, we consider also the case of a null model where additionally to the degree distribution, the number T of triangles is also the same. In this case, altough Table 6.1 shows a better estimate of C, there is yet a much larger number Q of squares. The heterosexual network is not compared to this last kind of null model because it does not have triangles (it is bipartite, as will be discussed in section 6.2).

Note that the clear difference in the number of cycles among the empirical networks and its randomized versions (or null model), implies that the stucture of cycles in the real data are not just a statistical consequence of the degree distribution. Thus, the presented model is a good choice for reproducing the observed cycle structure.

At the bottom of Table 6.1 we also show the values obtained with scale-free networks whose minimum number of connections was chosen to be $k_{min} = 2$. The chosen case $k_{min} = 2$, is that for which the clustering coefficient distributions are as close as possible from the distributions of the real networks. Clearly, the agent model not only yields clustering coefficients much closer to the ones measured in the empirical data, but also does not show the formation of larger amounts of loops (triangles and squares), a feature of the BA scale-free networks w hich is not observed in empirical data.

6.1.3 Summary and discussion

In this section we presented a new model for networks of complex interactions, based on a system of mobile agents whose collision dynamics is governed by an efficient event-driven algorithm that generates the links (contacts) between agents. As a specific application, we showed that the dynamical rules for interactions in sexual networks can be written as a velocity update rule which is a function of a power α of the previous contacts of each colliding agent. For suit-

able values of α and selectivity s, the model not only reproduces empirical data of networks of sexual contacts but also generates networks with similar topological features as the real ones, a fact that is not observed when using standard scale-free networks of static nodes.

Furthermore, our model predicts that the growth mechanism of sexual networks is not purely scale-free, due to interactions among internal agents, having a mean number of partners which increases in time. This should influence the predictions about spreading of infections [91]. The agent model presented here offers a suitable approach to study the emergence of complex networks of interactions in real systems, using only local information for each agent, and may be well suited to study networks in sociophysics, biophysics and chemical reactions, where interactions depend on specific local dynamical behavior of the elementary agents composing the network.

Since the results obtained with the present model are clearly more satisfactory than the ones obtained with previous models, we think that the key feature of mobile agents systems, namely mobility, is a keystone of real contact networks.

While given promising results the model may be improved in two particular aspects. First, it should enable the convergence towards a stationary regime with a growth process starting with all possible collisions instead of one particular agent from which the network is constructed. Second, the dependence of the above results on the velocity rule in Eq. (6.1) should be studied in detail, namely for the case of constant velocity ($\alpha = 0$). In chapter 7 it is shown that a stationary regime is easily obtained with the model described above by introducing a simple aging scheme, while by varying the parameter α one is able to reproduce other non-trivial degree distributions. Moreover, we introduced the selectivity parameter s to select from all possible social interactions (collisions) the ones which are of sexual nature. Without introducing this selectivity, the model of mobile agents is able to reproduce other social networks of acquaintances. Finally, the very few cycles present in the available data sets, here analyzed, could be just an affect of their small size. Therefore, larger data sets should be available in order to analyze and strenghten the applicability of the present model.

6.2 Cycles and clustering coefficient in bipartites networks

One important statistical tool to access the structure of complex networks arising in many systems [2, 6] is the clustering coefficient, introduced by Watts and Strogatz [13] to measure "the cliquishness of a typical neighborhood" in the network and given by the average fraction of neighbors which are interconnected with each other. This quantity has been used for instance to characterize small-world networks [13], to understand synchronization in scale-free networks of oscillators [95] and to characterize chemical reactions [96] and networks of social relationships [97, 98]. One pair of linked neighbors corresponds to a 'triangle', i.e. a cycle of three connections.

While triangles may be abundant in networks of identical nodes they cannot be formed in bipartite networks [97–99], where two types of nodes exist and connections link only nodes of different type. Thus, the standard clustering coefficient is always zero. However, different bipartite networks have in general different cliquishnesses and clustering abilities [98], stemming for another coefficient which uncovers these topological differences among bipartite networks. Bipartite networks arise naturally in e.g. social networks [99, 100] where the relationships (connections) depend on the gender of each person (node), and there are situations, such as in sexual contact networks [69], where one is interested in comparing clustering properties between monopartite (identical nodes) and bipartite (two types of nodes) compositions.

In this section, we study the cliquishness of either monopartite and bipartite networks, using both the standard clustering coefficient and an additional coefficient which gives the fraction of squares, i.e. cycles composed by four connections. As shown below, such a coefficient retains the fundamental properties usually ascribed to the standard clustering coefficient in regular, small-world and scale-free networks. As a specific application, the two examples of networks of sexual contacts will be studied and compared, one being monopartite and another bipartite.

Furthermore, we will show that one can take triangles and squares as the basic units of larger cycles in any network, monopartite or multipartite. The frequency and distribution of larger cycles in networks have revealed its importance in recent research for instance to characterize local ordering in complex networks from which one is able to give insight about their hierarchical structure [101], to determine equilibrium properties of specific network models [102], to estimate the ergodicity of scale-free networks [103], to detect phase transitions in the topology of bosonic networks [104] and to help characterizing the Internet

structure [105]. Since the computation of all cycles in arbitrarily large networks is unfeasible, one uses approximate numerical algorithms [103, 106, 107] or statistical estimates [108, 109]. Here, we go a step further and deduce an expression to estimate the number of cycles of larger size, using both clustering coefficients, which not only improves recent estimates [109] done for monopartite networks, but at the same time can be applied to bipartite networks and multipartite networks of higher order.

We start in Section 6.2.1 by introducing the expression which characterizes the cliquishness of bipartite networks, comparing it to the usual clustering coefficient. In Section 6.2.2 we use both coefficients to estimate cycles of larger size and show how they are applied to bipartite networks, while in Section 6.2.3 we apply both coefficients to real networks of sexual contacts. Discusions are given in Section 6.2.4.

6.2.1 Two complementary clustering coefficients

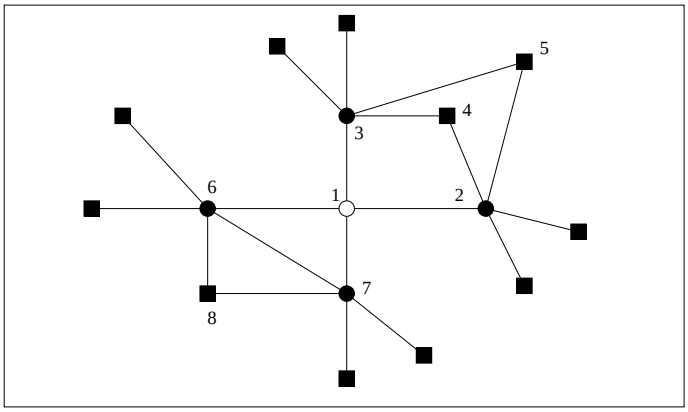

Figure 6.9: Illustration of the neighborhood of a central node (○) composed by its first neighbors (●) and its second neighbors (■), i.e. the neighbors of its neighbors. First and second neighbors are used to compute the complementary clustering coefficient C_4 (see text).

The standard definition of clustering coefficient C_3 is the fraction between the number of triangles observed in one network out of the total number of possible triangles which may appear. For a node i with a number k_i of neighbors the total

number of possible triangles is just the number of pairs of neighbors given by $k_i(k_i-1)/2$. Thus, the clustering coefficient $C_3(i)$ for node i is

$$C_3(i) = \frac{2t_i}{k_i(k_i-1)}. \tag{6.2}$$

where t_i is the number of triangles observed, i.e. the number of connections among the k_i neighbors. As in other studies, here and throughout the manuscript multiple connections between the same pair of nodes are not allowed.

Similarly to $C_3(i)$, a cluster coefficient $C_4(i)$ with squares is the quotient between the number of squares and the total number of possible squares. For a given node i, the number of observed squares is given by the number of common neighbors among its neighbors, while the total number of possible squares is given by the sum over each pair of neighbors of the product between their degrees, after subtracting the common node i and an additional one if they are connected. Explicitly, for a given node i the contribution of a pair of neighbors, say m and n, to $C_4(i)$ reads

$$C_{4,mn}(i) = \frac{q_{imn}}{(k_m-\eta_{imn})(k_n-\eta_{imn})+q_{imn}}, \tag{6.3}$$

where q_{imn} is the number of common neighbors between m and n (not counting i) and $\eta_{imn} = 1+q_{imn}+\theta_{mn}$ with $\theta_{mn} = 1$ if neighbors m and n are connected with each other and 0 otherwise. The numerator in Eq. (6.3) gives the number of squares containing nodes i, m and n, while the denominator counts the total possible number of squares containing these three nodes.

To illustrate the definition given in Eq. (6.3), we show in Fig. 6.9 a simple sketch of a node ($\circ$) neighborhood composed by its first and second neighbors ($\bullet$ and $\blacksquare$ respectively), Considering the neighbors 2 and 3, one has $q_{123} = 2$ squares containing nodes 1, 2 and 3 and there are $k_2 = 5$ and $k_3 = 5$ neighbors of nodes 2 and 3 respectively. Since nodes 2 and 3 are not connected with each other $\theta_{23} = 0$, yielding $\eta_{123} = 3$ and a denominator in Eq. (6.3) which equals 6 possible squares: two squares which are observed and other four squares corresponding to the possible combinations of all pairs of non-common neighbors. For neighbors 6 and 7 a similar calculation can be done, this time with $\theta_{67} = 1$ since the neighbors are connected with each other. The clustering coefficient $C_4(i)$ is easily obtained from Eq. (6.3) just by summing the numerator and denominator separately over the neighbors of i.

While $C_3(i)$ gives the probability that two neighbors of node i are connected with each other, $C_4(i)$ is the probability that two neighbors of node i share a

common neighbor (different from i). Averaging $C_3(i)$ and $C_4(i)$ over the nodes yields two complementary clustering coefficients, $\langle C_3 \rangle$ and $\langle C_4 \rangle$, characterizing the contribution to the network cliquishness of the first and second neighbors respectively. For simplicity we write henceforth C_3 and C_4 for the averages of $C_3(i)$ and $C_4(i)$ respectively.

An important point to stress concerns the denominators in the definitions of both clustering coefficients. The possible number of triangles in Eq. (6.2) does not take into account the topology of the neighborhood, in particular the number of second neighbors. Instead, the standard way [13] to compute C_3, given by Eq. (6.2), is to assume that all possible triangles are observed when the neighbors are fully interconnected. Consequently, possible degree-correlation biases may appear. The same occurs for the definition of C_4. Recently [110] another expression for C_3 was proposed with the aim to filter out these degree-correlation biases by taking into account the minimum number of neighbors of each pair of nodes considered. A similar approach could be done for C_4, substituting the denominator in Eq. (6.3) by a suited function of the minimum number of neighbors of n and m. However, here we consider C_4 as defined above, since it is our purpose to establish a parallel between C_4 and the standard definition of C_3, which itself does not take into account either the correlation removal proposed in Ref. [110].

Figure 6.10 shows both clustering coefficients C_3 and C_4 in several topologies. In all cases C_3 and C_4 are plotted as dashed and solid lines respectively, and are averages over samples of 100 realizations. As an example of regular networks, we use networks with boundary conditions where each node has n neighbors symmetrically placed, i.e. when arranged in a chain, each node has an even number of neighbors, half of them on one side and the other half on the other side. In particular, for $n = 2$ one obtains a chain of nodes connected to its nearest neighbors. For these regular networks, Fig. 6.10a shows the dependence of the clustering coefficients on the fraction n/N of neighbors, with $N = 10^3$ the total number of nodes. As one sees $C_4 < C_3$ and for either small or large fractions of neighbors both coefficients increase abruptly with n. In the middle region C_3 is almost constant, while C_4 decreases slightly. Our simulations have shown that in regular networks the coefficients depend only on n/N, i.e. for any size of the regular network, similar plots are obtained.

Figure 6.10b shows the coefficients for small-world networks with $N = 10^3$ nodes, constructed from a regular network with $n = 4$ neighbors symmetrically placed. The coefficients are computed as functions of the probability p to rewire short-range connections into long-range connections and they are normalized as usual [13] by the clustering coefficients $C^0_{3,4}$ of the underlying regular network. As one sees, C_4 yields approximately the same characteristics as the

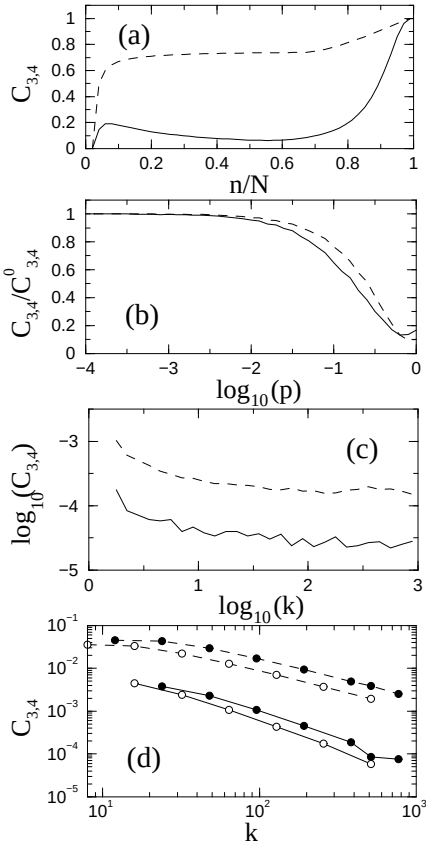

Figure 6.10: Comparisons between the standard clustering coefficient C_3 in Eq. (6.2) (dashed line) and the clustering coefficient C_4 in Eq. (6.3) (solid line) for different network topologies: **(a)** in a regular network with n neighbors symmetrically placed ($N = 10^3$), **(b)** in small-world networks where long-range connections occur with probability p ($N = 10^3$ and $n = 4$) and **(c)** in random scale-free networks where the distribution of the clustering coefficients is plotted as a function of the number k of neighbors ($N = 10^5$ and $m = 2$). In all cases samples of 10^2 networks were used. The distributions $C_3(k)$ and $C_4(k)$ are also plotted for **(d)** Apollonian networks [4] with $N = 9844$ nodes (●) and pseudo-fractal networks [3] with $N = 9843$ nodes (○).

standard clustering coefficient C_3 being therefore able to define the same range of p for which small-world effects are observed. While here the small-world networks were constructed rewiring short-range connections into long-range ones, the same features are observed when using the construction procedure introduced in Ref. [111] where instead of rewiring one just adds long-range connections.

To construct scale-free networks, we use the standard procedure of Albert and Barabási with growth and preferential attachment proportional to the number of neighbors (see e.g. Ref. [2] for details). For such scale-free networks, which we call henceforth random scale-free networks, we plot in Fig. 6.10c the distribution of both coefficients as functions of the number k of neighbors, using networks with $N = 10^5$ nodes and by given initially $m = 2$ connections to each node. Here, one observes that $C_4(k)$ is almost constant as k increases, reproducing the same known feature as the standard $C_3(k)$ apart a scaling factor: $C_4(k)/C_3(k)$ which is approximately constant for all k. In Fig. 6.10d we plot the clustering distributions for two different deterministic scale-free networks recently studied, namely Apollonian networks [4], represented by bullets $\bullet$, and pseudo-fractal networks [3], represented by circles $\circ$. In both cases, the same power-law behavior already known for $C_3(k) \sim k^{-\alpha}$ in these hierarchical networks is also observed for the coefficient $C_4(k)$ with the same value of the exponent α.

All networks in Fig. 6.10 are monopartite, i.e. no distinction between nodes is made, to have the straightforward comparison between both clustering coefficients, C_3 and C_4. Of course, in the case that bipartite counterparts are considered, the standard clustering coefficient C_3 vanishes, and only C_4 is suited to measure the clustering between nodes.

In short, the results shown in Fig. 6.10 give evidence that C_4 is also a suited coefficient to characterize topological features of several complex networks commonly done with the standard clustering coefficient C_3. Furthermore, since C_4 counts squares instead of triangles, it is particularly suited for bipartite networks. Next, we will use this coefficient to compare different models for networks of sexual contacts, where both monopartite and bipartite networks arise naturally.

6.2.2 Estimating the number of large cycles with squares and triangles

Recent studies have attracted attention to the cycle structure of complex networks, since the presence of cycles has important effects for example on information propagation through the network [112] and on epidemic spreading be-

havior [113]. In order to avoid numerical algorithms counting the number of cycles of arbitrary size which implies long computation times, an estimate of the fraction of cycles with different sizes was proposed [109], using the degree distribution $P(k)$ and the standard cluster coefficient distribution $C_3(k)$. However, this estimate yields a lower bound for the total number of cycles and cannot be applied to bipartite networks, as shown below. The aim of this Section is twofold. First, to show that by using both C_3 and C_4 one is able to improve that estimate, being suited at the same time to either monopartite and bipartite networks. Second, to explicitly show some limitations of the estimates below and discuss their applicability.

The estimate in Ref. [109] considers the set of cycles with a central node, i.e. cycles with one node connected to all other nodes composing the cycle. Figure 6.11a illustrates one of such cycles, where the central node and each pair of its consecutive neighbors forms a triangle, within a total amount of four adjacent triangles. In such set of cycles, to estimate the number of cycles with size s one looks to the central node of each cycle which has a number, say k, of neighbors. The number of different possible cycles to occur is $n_0(s,k) = \binom{k}{s-1}\frac{(s-1)!}{2}$, since one has $\binom{k}{s-1}$ different groups of s nodes and in each one of these groups there are $(s-1)!/2$ different ways ordering the s nodes into a cycle. The fraction $n_0(s,k)$ of cycles which is expected to occur is $p_0(s,k) = C_3(k)^{s-2}$, since the probability of having one edge between two consecutive neighbors is $C_3(k)$ and

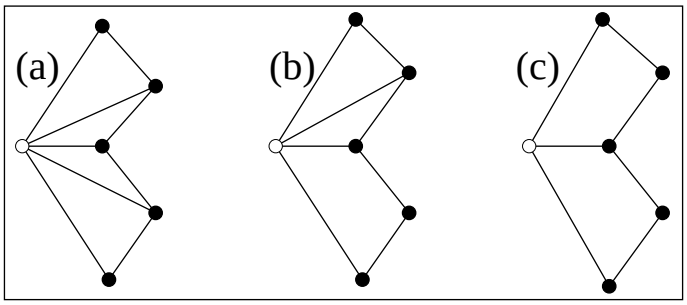

Figure 6.11: Illustrative examples of cycles (size $s = 6$) where the most connected node ($\circ$) is connected to **(a)** all the other nodes composing the cycle, forming four adjacent triangles. In **(b)** the most connected node is connected to all other nodes except one, forming two triangles and one sub-cycle of size $s = 4$, while in **(c)** the same cycle $s = 6$ encloses two sub-cycles of size $s = 4$ and no triangles (see text).

one must have $s-2$ edges between the $s-1$ neighbors. Therefore, the number of cycles of size s is estimated as

$$N_s = N g_s \sum_{k=s-1}^{k_{max}} P(k) n_0(s,k) p_0(s,k), \tag{6.4}$$

where $P(k)$ is the degree distribution and g_s is a factor which takes into account the number of repeated cycles. This geometrical factor can be computed for each particular case of s-cycles but the estimates can be carried out without the explicit computation of the factor [109].

The estimate in Eq. (6.4) is a lower bound for the total number of cycles since it considers only cycles with a central node. For instance, in Fig. 6.11b while cycles of size $s=4$ can be estimated with Eq. (6.4), the cycle $s=6$ cannot since it has no central node, and in Fig. 6.11c the above equation cannot estimate any cycle of any size. In fact, Fig. 6.11c illustrates the type of cycles appearing in bipartite networks, where no triangles are observed. For such cycles $C_3(k)=0$ and therefore all terms in Eq. (6.4) vanish yielding a wrong estimate of the number of cycles.

To take into account cycles without central nodes (Figs. 6.11b and 6.11c), one must consider the clustering coefficient $C_4(k)$ defined in Eq. (6.3). One first considers the set of cycles of size s with one node ($\circ$) connected to all the others *except* one, as illustrated in Fig. 6.11b. In this case, since there are $s-2$ nodes connected to node $\circ$ one has $n_1(s,k) = \binom{k}{s-2}(s-2)!/2$ different possible cycles of size s, with k the number of neighbors of node $\circ$. The fraction of the $n_1(s,k)$ cycles which is expected to be observed is given by $p_1(s,k) = C_3(k)^{s-4} C_4(k)(1-C_3(k))$, since the probability of having $s-4$ connections among the $s-2$ connected nodes is $C_3(k)^{s-4}$, the probability that a pair of neighbors of node $\circ$ has to share a common neighbor (different from node $\circ$) is $C_4(k)$ and the probability that these same pairs of neighbors are not connected is $(1-C_3(k))$. Writing an equation similar to Eq. (6.4), where instead of $n_0(s,k)$ and $p_0(s,k)$ one has $n_1(s,k)$ and $p_1(s,k)$ respectively and the sum starts at $s-2$ instead of $s-1$, one has an additional number N'_s of estimated cycles which is not considered in estimate (6.4). Notice that, since for N'_s one considers at least one sub-cycle of size $s=4$, this additional estimate contributes only to the estimate of cycles with size $s \geq 4$. We call henceforth sub-cycle, a cycle which is enclosed in a larger cycle and which does not contain itself any shorter cycle.

Still, the new estimate $N_s + N'_s$ is not suited to bipartite networks, since it yields nonzero estimate only for $s=4$. To improve the estimate further one must

consider not only cycles composed by one single sub-cycle of size $s = 4$, as done in the previous paragraph, but also cycles with any number of sub-cycles of size $s = 4$. Figure 6.11c illustrates a cycle of size $s = 6$ composed by two sub-cycles of size 4. In general, following the same approach as previously, for cycles composed by q sub-cycles of size 4 one finds $n_q(s,k) = \frac{(s-q-1)!}{2}\binom{k}{s-q-1}$ possible cycles of size s calculated for a node with k neighbors and a fraction $p_q(s,k) = C_3(k)^{s-2q-2}C_4(k)^q(1-C_3(k))^q$ of them which are expected to be observed. For $q = 0$ one considers cycles as the one illustrated in Fig. 6.11a, while for $q = 1$ and $q = 2$ one considers the set of cycles with one and two sub-cycles of size 4, as illustrated in Figs. 6.11b and 6.11c respectively. Summing up over k and q yields our final expression

$$N_s = Ng_s \sum_{q=0}^{[s/2]-1} \sum_{k=s-q-1}^{k_{max}} P(k) n_q(s,k) p_q(s,k). \tag{6.5}$$

where $[x]$ denotes the integer part of x. In particular, the first term ($q = 0$) is the sum in Eq. (6.4). The upper limit $[s/2] - 1$ of the first sum results from the fact that the exponent of $C_3(k)$ in $p_q(s,k)$ must be non-negative: $s - 2q - 2 \geq 0$. The estimate in Eq. (6.5) not only improves the estimated number computed from Eq. (6.4), but also enables the estimate of cycles up to a larger maximal size. In fact, since in the binomial coefficient $\binom{k}{s-1}$ of Eq. (6.4) one must have $s - 1 \leq k \leq k_{max}$, one only estimates cycles of size up to $k_{max} + 1$, while in Eq. (6.5) the maximal size is $2k_{max}$, as can be concluded using both conditions $s - 2q - 2 \geq 0$ and $s - q - 1 \leq k_{max}$.

Figure 6.12 compares two cases treated in Ref. [109], both with a degree distribution $P(k) = P_0 k^{-\gamma}$ and coefficient distributions $C_3(k) = C_3^{(0)} k^{-\alpha}$, using one value of $\alpha < 1$ (Fig. 6.12a) and another one $\alpha > 1$ (Fig. 6.12b). Dashed lines indicate the estimate using Eq. (6.4), while solid lines indicate the estimate using Eq. (6.5). In both cases, the latter estimate is larger. For $\alpha < 1$ the difference between both estimates decreases with the size s of the cycle. For $\alpha > 1$ the difference between the estimates increases with s beyond a size $s^* \lesssim k_{max}$. Clearly, from Fig. 6.12b one sees that $k_{max} + 1$ is the larger cycle size for which Eq. (6.4) can give an estimate, while for Eq. (6.5) the estimate proceeds up to $2k_{max}$ (partially shown). In both cases, the typical size for which N_s attains a maximum is numerically the same for both estimates, as expected. Moreover, for $\alpha > 1$ (Fig. 6.12b), beyond a size of the order of k_{max}, $N_s/(Ng_s)$ in Eq. (6.5) decreases exponentially with s, slightly different as observed for Eq. (6.4). In fact, the deviation of Eq. (6.4) from the exponential tail, is due to the fact that for very large cycle sizes ($s \sim k_{max}$) Eq. (6.4) can only consider very few terms in its sum.

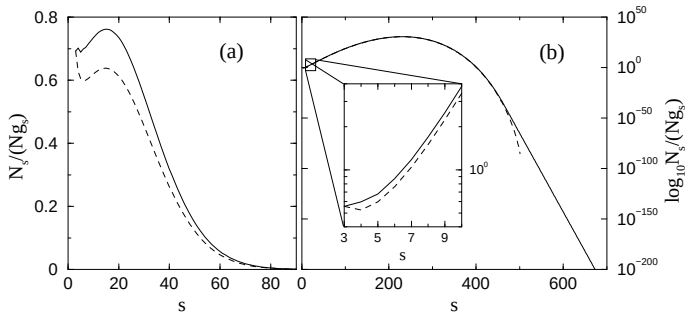

Figure 6.12: Estimating the number of cycles using Eq. (6.4), dashed lines, and Eq. (6.5), solid lines. Here we impose a degree distribution $P(k) = P_0 k^{-\gamma}$ with $P_0 = 0.737$ and $\gamma = 2.5$, and coefficient distributions $C_{3,4}(k) = C_{3,4}^{(0)} k^{-\alpha}$ with **(a)** $C_3^{(0)} = 2$, $C_4^{(0)} = 0.33$, $\alpha = 0.9$ and **(b)** $C_3^{(0)} = 1$, $C_4^{(0)} = 0.17$, $\alpha = 1.1$. In all cases $k_{max} = 500$.

Another advantage of the estimate of Eq. (6.5) is that it counts cycles in bipartite networks. For bipartite networks there are no connections between the neighbors, i.e. all subgraphs are similar to the one illustrated in Fig. 6.11c. Therefore all terms in Eq. (6.5) vanish except those for which the exponent of $C_3(k)$ is zero, i.e. for $s = 2(q+1)$. Consequently, since q is an integer, Eq. (6.5) shows clearly that in bipartite networks there are only cycles of even size, as already known [99]. Moreover, substituting $q = (s-2)/2$ in Eq. (6.5) yields a simple expression for the number of cycles in bipartite networks, namely

$$N_s^{\text{Bipart}} = N g_s \sum_{k=s/2}^{k_{max}} P(k) \frac{(s/2)!}{2} \binom{k}{s/2} C_4(k)^{s/2-1}. \tag{6.6}$$

A simple example to illustrate the validity of Eq. (6.5) is the fully connected network, where each node is connected to each other one. In this case the number of cycles with size s is given by $N_s = \binom{N}{s} \frac{(s-1)!}{2}$. The factor $(s-1)!$ counts the arrangements between $s-1$ nodes in each combination of s nodes, while the division by two is due to the undirected links. To compute N_s from Eq. (6.5) one has for the particular case of the fully connected network, $P(k) = C_3(k) = C_4(k) = \delta_{k-N+1}$, $k_{max} = N-1$ and $g_s = 1/s$. Consequently the only nonzero term in the first sum is the one for $q = 0$, while the nonzero term in the second sum is the one for $k = N-1$, yielding the same result as above.

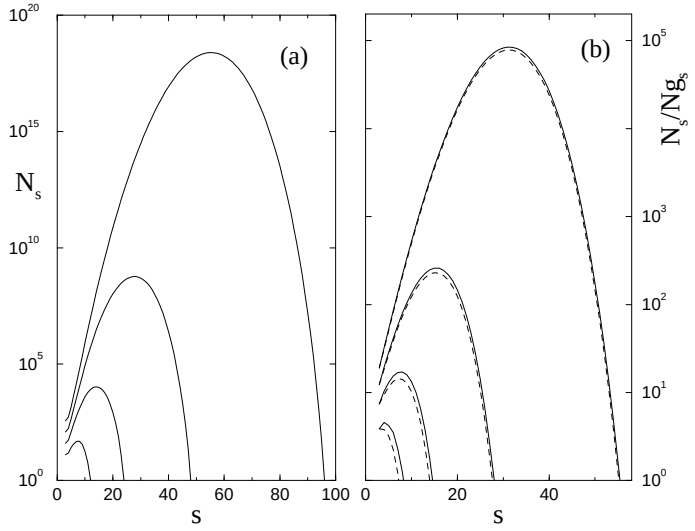

Figure 6.13: **(a)** The exact number of cycles as a function of the size for the pseudo-fractal network [3] compared with **(b)** N_s/Ng_s of the analytical expressions in Eqs. (6.4), dashed lines, and (6.5), solid lines. From small to large curves one has pseudo-fractal networks with $m = 2, 3, 4, 5$ generations (see text).

Both the expression in Eq. (6.4) and the one in Eq. (6.5) are particularly suited for networks or subnetworks where nodes are highly connected to each other, since in those situations there is a very large number of centrally connected cycles as the ones illustrated in Fig. 6.11. Highly connected subnetworks appear, for instance, in social networks which are composed by communities [21]. In Ref. [109], for instance, the estimate of small cycles from Eq. (6.4) is compared to the true values computed for several empirical networks, namely the Internet, the co-authorship web and semantic networks. While for $s = 3$ and 4 the estimate is clearly good, for $s = 5$ there is a clear underestimation, due to the appearance of no centrally connected cycles. Of course one expects that, similarly to what is observed in Fig. 6.12, the estimate in Eq. (6.5) improves the one used in [109] for such situations.

Next we illustrate this point using a particular network, the pseudo fractal network, already described in section 4.1.1. For these networks, the exact number

of cycles with size s can be written iteratively [114] as

$$N_s^{(m+1)} = \sum_{l=3}^{s} \binom{l}{s-l} N_s^{(m)}, \tag{6.7}$$

for $s \geq 4$ and $N_3^{(m+1)} = N_3^{(m)} + 3^m$.

Figure 6.13 shows the real number of cycles of the pseudo-fractal network (Fig. 6.13a) with the quantity $N_s/(Ng_s)$ (Fig. 6.13b) for the pseudo-fractal network. In Fig. 6.13b solid lines indicate our estimate, while dashed lines indicate the previous estimate in Ref. [109]. In both cases, the underestimation is very significant when compared to the exact number from Eq. (6.7). Nevertheless, even in this case, the estimates predict the shape of the cycle distributions. Up to our knowledge, complex networks for which the exact number of cycles may be computed having most nodes highly connected are not known.

It is important to notice that triangles and squares may appear in any multipartite network (except in bipartite ones, where triangles are absent). Therefore, the estimates described and studied in this Section can be applied not only to bipartite networks but also to any multipartite network of any order. In the next Section we will focus on the applicability of the clustering coefficient C_4 in empirical sexual networks (monopartite and bipartite) with the aim to compare simulated results for such networks.

6.2.3 Cycles and clustering in sexual networks

In this Section we apply both coefficients C_3 and C_4 in Eqs. (6.2) and (6.3) to analyze the two real networks of sexual contacts previously presented. As Figure 6.7 sketches, in these two networks one can see that cycles of different sizes appear. While the network with only heterosexual contacts is clearly bipartite, the network with homosexual contacts is monopartite.

For the two networks in Fig. 6.7, Table 6.2 presents the coefficients C_3 and C_4 with the same simulations of the model shown in Table 6.1. As one sees, although the heterosexual network has less squares than the homosexual network due to its smaller size, C_4 is much larger. Another feature common to both networks is $L/N \sim 1$, i.e. an effective coordination number of $2L/N \sim 2$.

While in the heterosexual network the number of squares and consequently the value of C_4 is overestimated, for the homosexual network the null model yields

	$\langle C_3 \rangle$	$\langle C_4 \rangle$
Heterosexual (Fig. 6.6a)	0	0.00486
Homosexual (Fig. 6.6b)	0.02980	0.00192
Heterosexual (Null model)	0	0.0451
Homosexual (Null model)	0.011	0.00373
Homosexual (Null model, same T)	0.0145	0.00477
Heterosexual (Agent Model)	0	0.01273
Homosexual (Agent Model)	0.02302	0.01224
Heterosexual (Scale-free)	0	0.12859
Homosexual (Scale-free)	0.08170	0.02787

Table 6.2: Clustering coefficients and cycles in two real networks of sexual contacts (top), illustrated in Fig. 6.7, one where all contacts are heterosexual (Fig. 6.7a) and another with homosexual contacts (Fig. 6.7b). In each case one indicates the values of both clustering coefficients C_3 and C_4 in Eqs. (6.2) and (6.3) respectively. The values of these quantities are compared with the ones of a null model (see text) with the same degree distribution for two cases, one where the number of triangles is fixed and another where this restriction is not imposed, and also with networks constructed with the agent model presented in the previous section [69] and with the BA scale-free network with $m = 2$. Samples of 100 realizations were used in each case.

reasonable results for both clustering coefficients, although the large discrepancy between the number of triangles and squares. In order to compare the number of squares without the effect of the number of triangles in the network, as in section 6.1.2, we consider also the case of a null model where additionally to the degree distribution, the number T of triangles is also the same. In this case, Table 6.2 still shows an underestimation of C_3 and a much larger number Q of squares. Notice that, while the total number of triangles is the same, the standard clustering coefficient $\langle C_3 \rangle$ can be nevertheless different, since it is an average

over the local clustering coefficient of each node, which depends not only on the number of triangles the node belongs to but also on its degree.

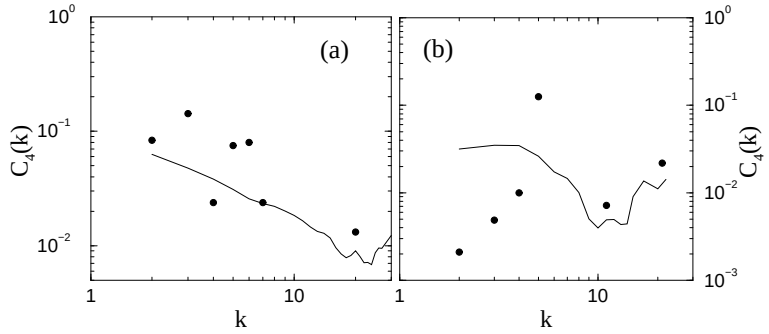

Figure 6.14: Comparing clustering coefficient $C_4(k)$ in Eq. (6.3) between networks obtained from the agent model (solid lines) used to reproduce real networks of sexual contacts (bullets): **(a)**Monopartite case, $N = 250$. **(b)**Bipartite case, $N = 82$. Same simulations of the model as presented in Fig. 6.8.

Using the same number of nodes as in the real networks illustrated in Fig. 6.7 and considering two types of nodes for the heterosexual (bipartite) case, we obtain with the agent model similar results for C_3 and C_4, as shown in Table 6.2 where values represent averages over samples of 100 realizations. As one sees, in general, the agent model yields values much closer to the ones for the empirical networks, than the two null models considered above. Remarkably, for the bipartite case not only the number of connections and the number of squares are numerically the same (as was shown in table 6.1), but also C_4 is of the same order of magnitude. Similar values of the topological quantities are also obtained for the monopartite case, with the exception of C_4.

In Fig. 6.14a we show the clustering coefficient distributions for the monopartite network of sexual contacts sketched in Fig. 6.7b, while in Fig. 6.14b we plot the distribution for the bipartite network (Fig. 6.7a). In both figures bullets indicate the distributions of the empirical data, while solid lines indicate the distributions of the networks obtained with the agent model, imposing the same size as the real network, i.e. stopping the simulation when the number of connected agents equals the size of the corresponding empirical network, and taking averages over a sample of 100 realizations.

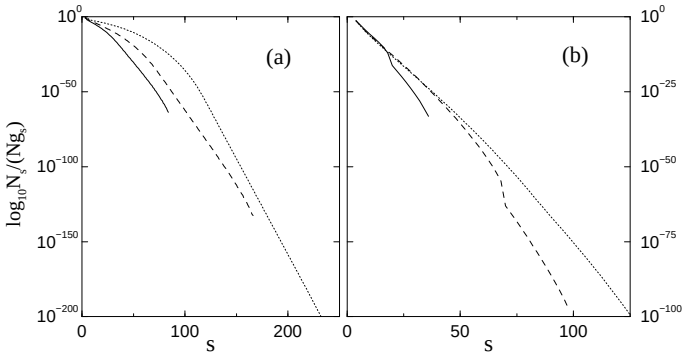

Figure 6.15: Estimating the number of cycles for the agent model using Eq. (6.5) for $N = 1000$ (solid lines), $N = 5000$ (dashed lines) and $N = 10000$ (dotted lines) in **(a)** a monopartite network and in **(b)** a bipartite network, both obtained with the agent model.

The results above concern small empirical networks. To improve the particular study of sexual networks reproduced by our model, larger networks of sexual contacts should be also studied and comparisons with a null model [90] must be carried out to validate the agent model. The main point here is that the results above already show that the complementary clustering coefficient C_4 is suited to quantify the cliquishness of neighborhoods in either monopartite and bipartite counterparts of the same complex networks, while the standard clustering coefficient is not.

With the agent model one is able to construct larger networks than the empirical ones. In such large networks cycles of different size may appear and one important question is to know the frequency of cycles of any order. Using the agent model for large networks, and computing only their degree distribution and the two clustering coefficients we can estimate the distribution of cycles in those networks. Figure 6.15 shows the distribution of the fraction $N_s/(Ng_s)$ of cycles as a function of their size s, for a monopartite network (Fig. 6.15a) and a bipartite network (Fig. 6.15b) composed of $N = 1000, 5000$ and 10000 nodes. Here, while monopartite networks show an exponential tail preceded by a region where the number of cycles is large, bipartite networks are composed of cycles whose number depends exponentially on their size. Furthermore one observes a clear transition at a characteristic size, which seems to scale with the network size.

6.2.4 Sumary and discussion

We introduced a clustering coefficient similar to the standard one, which instead of measuring the fraction of triangles in a network measures the fraction of squares, and showed that with this clustering coefficient it is also possible to characterize topological features in complex networks, usually done with the standard coefficient. We showed explicitly that the range of values of the probability to acquire long-range connections in small-world networks and the typical clustering coefficient distributions of either random scale-free and hierarchical networks are approximately the same. In addition, we showed that this second clustering coefficient enables one to quantify the cliquishness in bipartite networks where triangles are absent. Thus, one should take triangles and squares simultaneously as the two basic cycle units in any network.

An application of both clustering coefficients was proposed, namely to estimate the number of cycles in any network, either monopartite or multipartite. Using a recent estimate which yields a lower bound of the number of cycles in monopartite networks up to a size $s < k_{max} + 1$ where k_{max} is the maximum number of neighbors in the network, we deduce a more general expression which not only improves the previous estimate but is also suited for bipartite networks and enables one to estimate cycles of size up to $2k_{max}$. Furthermore, in the particular case of bipartite networks our estimate yields as a natural consequence that only cycles of even size may appear.

To illustrate the applicability of the complementary clustering coefficient in bipartite networks, we studied a concrete example of two sexual networks, one where only heterosexual contacts occur (bipartite network) and another with homosexual contacts (monopartite). The results obtained with the two real networks were found to be similar to the ones obtained with the agent model introduced in Sec. 6.1.1.

All in all, our analytical expression gives a simple way to extract information concerning the distribution of cycles in multipartite networks, and in particular the clustering coefficient C_4 can be regarded as a suited measure of neighborhood cliquishnesses in bipartite networks.

Chapter 7

Model of mobile agents for social interaction networks

The topological features of networks of acquaintances fundamentally differ from other networks [2, 115]. First, they are single-scale networks and present small-world effect [33]. Second, they are divided into groups or communities [2]. Additionally, while those networks are dynamical, evolving in time, their evolution process differs from standard growth models as those that govern e.g. the World Wide Web, or from copying mechanisms proposed for biological networks. An interesting development in this area is given in [116] where a simple procedure of transitive linking is proposed to generate small-world networks. While each one of the mentioned features can be reproduced with some previous model, there is still no single model that incorporates simultaneously dynamical evolution, clustering and community structure.

In this chapter we show that all these characteristics can be reproduced in a very natural way, by using standard concepts and techniques from physical systems. Namely, we propose an approach to dynamical networks based on a system of mobile agents representing the nodes of the network. We will show that, due to this motion, it is possible to reproduce the main properties [1, 2] of empirical social networks, namely the degree distribution, the clustering coefficient (C) and the shortest path length, by choosing the same average degree measured in the empirical networks, and adjusting only one parameter, the density of the system. The community structure emerges naturally, without labeling *a priori* the community each agent belongs to, as in previous works [117]. Moreover, this ap-

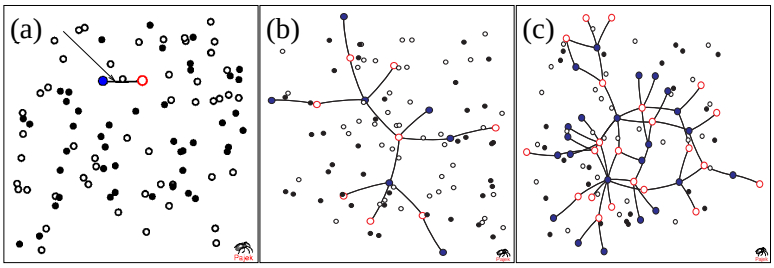

Figure 7.1: Snapshots of the system of mobile agents, where the edges between two agents indicate that they already collided with each other. **(a)** Snapshot after the first collision, **(b-c)** two subsequent snapshots within one cluster of collisions. Filled nodes (blue) and unfilled nodes (red) represent two different types of nodes, e.g. males and females (see Sec. 7.5 for details).

proach gives some insight to further explain the structure of empirical networks, from a recently available large data set of friendship networks [118] concerning 90118 students, divided among 84 schools from USA, constructed from an In-School questionnaire (see details of this database in chapter 8). The acquaintance between pairs of students was rigorously defined. Each student was given a paper-and-pencil questionnaire and a copy of a list with every student in the school. The student was asked to check if he/she participated in any of 5 activities with the friend: like going to (his/her) house in the last seven days, or meeting (him/her) after school to hang out or go somewhere in the last seven days, etc. Other studies [33] have used a slightly different definition of friendships and obtained the same kind of degree distribution, an indication of the robustness of the concept of friendship.

7.1 The Model

In this chapter we examine a stationary version of the model presented in Sec. 6.1.1 and used to model growing networks of sexual contacts. Here social contacts are introduced by setting a link joining the two agents after a collision, till its *removal* when one of the agents leaves the system when reaching certain age. Therefore, during the evolution of the system, each agent i is characterized by its number k_i of links and by its age $\mathcal{A}_i$.

Since collisions represent social contacts their dynamical rules should fulfill some sociological requirements. Namely, it is known [88] that many social interactions occur more commonly between individuals having already a large number of previous contacts. For instance, here we keep the velocity up-date rule (Eq. 6.1).

The exponent α in Eq. (6.1) controls the velocity update after each collision. For $\alpha = 0$ the velocity of each agent is constant in time, and consequently the kinetic energy density $\frac{1}{2}\rho v^2$ of the system is constant. For $\alpha > 0$ the velocity increases with degree k. In this range, the value $\alpha = 1$ ($|\mathbf{v}| \propto k$) marks a transition between a sub-linear regime ($\alpha < 1$) and a supra-linear regime ($\alpha > 1$) with different degree distributions[69, 123]. Throughout this chapter we will consider $\alpha = 1$ in most of the situations, showing that it produces the suitable dynamics to reproduce real networks of social contacts. It should be noticed that while positive values of α yield dynamical laws which fulfill sociological requirements, the equation of motion (6.1) is also able to consider completely different situations, where $\alpha < 0$, i.e. where the ability to acquire new contacts *decreases* with the number of previous contacts. In this research, we have focused in the regimes for $\alpha > 0$.

As one may notice, contrary to collision interactions where the velocity vector is completely deterministic [89], here momentum is *not* conserved. This is a consequence not only of the increase of $\mathbf{v}$ but also of the fact that after one collision the moving direction is randomly selected. The main reason for this random choice is that, as a first approximation it is plausible to assume that social contacts do not determine which social contact will occur next.

Concerning the residence time or 'age' parameter $\mathcal{A}$, during which the agents remain in the system, if $\mathcal{A} \to \infty$, each agent will eventually collide with all the other agents forming a fully connected network. Whereas, when the average residence time of the agents is finite the system will reach a nontrivial quasi-stationary states [33, 119], as described in section 7.2.

The aging scheme considered here is simply parameterized by some threshold in the age of the agents. More precisely, each agent i is initialized with a certain age $A_i(0)$ which is a random number uniformly distributed in the interval $[0, T_\ell]$ with T_ℓ being the maximal age an agent may have. Being updated according to $\mathcal{A}_i(t+\Delta t) = \mathcal{A}_i(t)+\Delta t$, the age eventually reaches $\mathcal{A}_i(t) = T_\ell$, when the agent i leaves the system, yielding a total residence time $T_\ell - \mathcal{A}_i(0)$. Computationally the replacement of an old agent by a new one is carried out simply by removing all the connections of the old agent and updating its velocity to the initial value v_0 with a new random direction.

Before proceeding in characterizing the behavior of such a system it is important to a/-ddress three last points to understand the parallel between the model and real systems.

First, the two-dimensional continuous space where nodes move is *not* the physical space where individuals travel, meet or establish social acquaintances. Instead, it represents a projection of a highly dimensional Euclidean space whose metric may be related to a so called the social distance [115, 117]: two close nodes have similar affinities (same tastes, same behavior, etc.) and therefore it is probable for them to establish an acquaintance, i.e. to collide. It should be stressed that the metric is *related* to social distance, but also incorporates effects of random factors promoting two persons to meet or establish friendship connections. We have no rigorous explanation for the fact that a two-dimensional projection of such a 'large'-dimensional space suffices to reproduce empirical data of acquaintances. But the fact is that it does, and therefore for simplicity we will consider a two-dimensional system. One could also consider higher dimensional systems of mobile agents but then the 'projected' velocity in Eq.(6.1) would change.

Second, since the system of mobile agents is used to extract a complex network, one could ask where the main parameters ρ and T_l appear in the network structure. One easily concludes that increasing the density ρ confines the accessible region of agents thus promoting the occurrence of collisions among them which are more confined in space. In other words, increasing the density one increases the clustering coefficient [13]. As for T_ℓ, the larger the residence time, the larger the number of collisions an agent may have. Thus, increasing T_ℓ increases the average number $\langle k \rangle$ of connections.

Finally, in a space of affinities what is the meaning of a velocity?. The velocity can be interpreted as measure of the accessible region of a given agent within this social space which increases, decreases or remains constant after each collision, depending on the value of α.

7.2 The quasi-stationary regime

In the QS state the dynamical and topological quantities fluctuate around an average value after a transient time of the order of $2T_\ell$. Figure 7.2 illustrates the convergence toward the QS state for two different values of T_ℓ, namely for $T_\ell = 30.75$ and $T_\ell = 73.35$. Here, the convergence is characterized by plotting

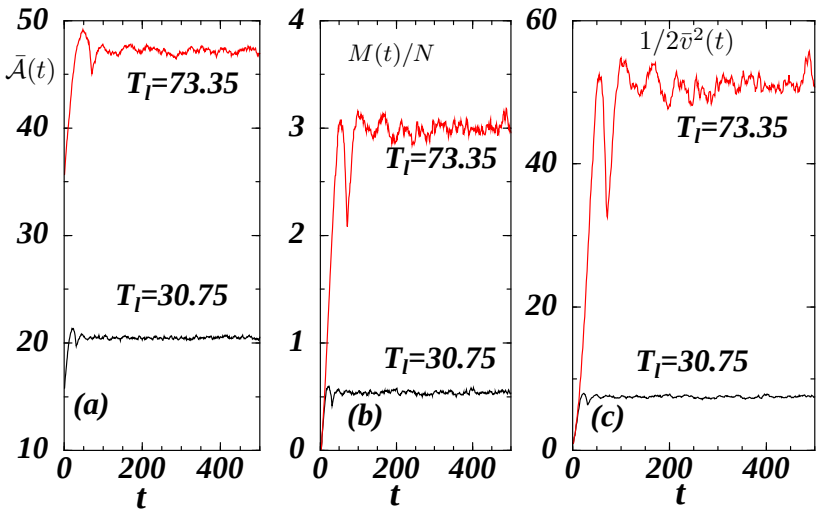

Figure 7.2: **(a)** Average age $\bar{\mathcal{A}}(t)$ vs. time. **(b)** Number of links per agent, M/N, vs. time. **(c)** Average energy vs. time. Two different time of life are shown: $T_l = 73.35$ and $T_l = 30.75$, at one realization of a system of $N = 4096$, with $v_0 = \sqrt{2}$ and $\rho = 0.02$

the effective coordination given by the fraction $M(t)/N$ between the total number M of connections and the total number N of agents (Fig. 7.2a), the mean energy $1/2\bar{v}^2(t)$ (Fig. 7.2b) and the mean age $\bar{\mathcal{A}}(t)$ (Fig. 7.2c), as function of time t, the bar '$-$' for $\bar{v}(t)$ and $\bar{\mathcal{A}}(t)$ represents the average over all the agents in a given time, t. In all three cases, the above quantities increase in the earlier stages of the network growth attaining a maximum value around $t \sim T_\ell/2$ where the agents start dying, resulting in the decrease of their values till a minimum at $t = T_\ell$. For the studied initialization of $\mathcal{A}_i$, uniformly distributed in range $[0, T_l]$, the maximum at $t \sim T_\ell/2$ is due to the fact that the agents have on average a life-time equal to $T_\ell/2$, while the minimum is due to the extinction of all the first 'generation' of agents.

When the average residence time is too small, two agents will have no time to collide at least once, and consequently no network is formed. On the contrary, when T_ℓ is too large, each agent will cross the entire system and a fully connected network appears. To avoid these two extreme regimes we consider an average residence time which is neither very small nor large when compared to the characteristic time τ between collisions.

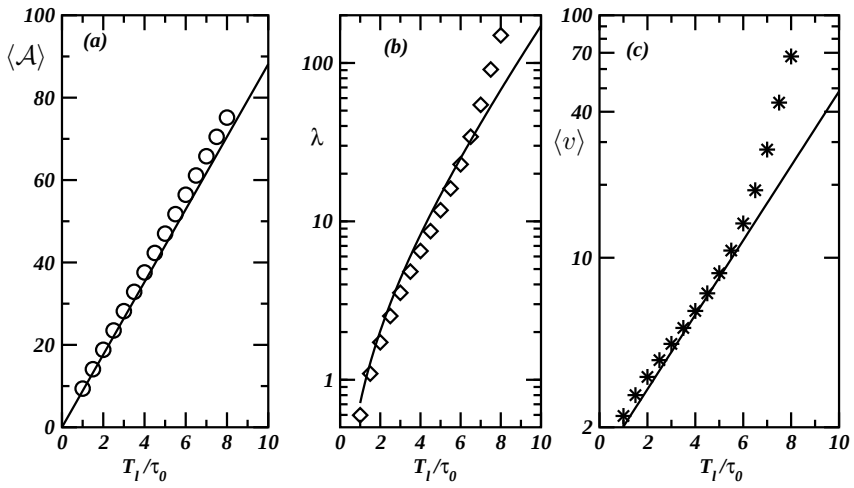

Figure 7.3: **(a)** Numerical simulation of average age, $\langle \mathcal{A} \rangle$ vs. T_l/τ_0 (circles), the straight line has slope $5/8\tau_0$ (Eq. 7.3). **(b)** Results of the numerical simulation of λ vs T_l/τ_0 (diamonds) compared to the result of Eq. 7.12 (solid line). **(c)** velocity modulus from the simulation (stars), compared to Eq. 7.10 (solid line). The simulations represent averages in the quasi-stationary regime, over 100 realizations, in a system of $N = 10^4$, with $\rho = 0.02$, $v_0 = \sqrt{2}$, for different values of T_l.

For that, we define a collision rate, as the fraction between the average residence time $T_\ell - \bar{\mathcal{A}}(0)$, and the characteristic time τ of the mean free path defined as

$$\tau = \frac{1}{\sqrt{\pi}\rho 2r\langle v \rangle}, \tag{7.1}$$

where $\langle v \rangle$ is the average velocity of the agents. With this assumptions our collision rate reads

$$\lambda = \frac{T_\ell - \bar{\mathcal{A}}(0)}{\tau} = \frac{1}{2v_0\tau_0}\langle v \rangle(T_\ell), \tag{7.2}$$

where $\langle\rangle$, means an average over the value for each agent at different snapshots in the quasistationary state (QS), τ_0 is the characteristic time of the system with the initial velocity v_0, i.e. $\tau_0 \equiv 1/(2\sqrt{\pi}r\rho v_0)$ and $\bar{\mathcal{A}}(0) = T_l/2$ is the average initial age of the agents.

At $t = 0$ we have $\bar{\mathcal{A}}(0) = T_l/2$, later at $t = T_l/2$, half of the population has been replaced by new individuals which have $\langle \mathcal{A} \rangle_{1st} = T_l/2$, while the other half has

$\langle \mathcal{A} \rangle_{2nd} = 3T_l/4$, these processes of birth and death produce a mean age which is the average of the two values $\langle \mathcal{A} \rangle_{1st}$ and $\langle \mathcal{A} \rangle_{2nd}$:

$$\langle \mathcal{A} \rangle = 5T_l/8. \tag{7.3}$$

In Fig. 7.3a we show numerical results of $\langle \mathcal{A} \rangle$ vs. T_l/τ_0 (circles) and the straight line has slope $5/8\tau_0$, which confirms Eq. 7.3.

The number of collisions of an agent $n_{c,i}$ in a two dimensional diluted gas, increases with time and is proportional to the radius of the agents (r), its velocity v_i and the density of agents (ρ) [120]:

$$n_{c,i}(t) = \rho r \sqrt{2\pi} v_i(t) t, \tag{7.4}$$

The inverse of the coefficient in Eq. 7.4 is the characteristic time between collisions. For simplicity we write it in terms of the initial velocity (v_0), and thus we have:

$$\frac{dn_{c,i}}{dt} = \frac{1}{\tau_0 v_0} v_i(t). \tag{7.5}$$

Because links are produced due to collisions, we write the degree of an agent as a function of $n_{c,i}$,

$$k_i(t) = (1 - f_i(t)) n_{c,i}(t), \tag{7.6}$$

where $f_i(t)$ is the fraction of neighbors of agent i that leave the system due to their age. Changes in the degree of an agent affect its velocity (according to the velocity up-date rule of Eq. 6.1):

$$v_i(t) = v_0 + v_1 k_i(t) \tag{7.7}$$

With Eqs.[7.4−7.7], we can write a differential equation for $v_i(t)$

$$\frac{dv_i}{dt} = v_1 \frac{df_i}{dt} \sqrt{2\pi} \rho r v_i(t) t + \frac{v_1}{\tau_0 v_0} (1 - f_i(t)) v_i(t). \tag{7.8}$$

Because we do not know an exact expression for $f_i(t)$, we have to make some approximations. We asume $f_i(t) = -t/(2T_l)$, which implies that after $t = T_l$ the degree of an agent equals half of its collisions , $k_i(T_l) \approx n_{c,i}(T_l)/2$. Substituting $f_i(t)$ into Eq. 7.8, we obtain:

$$\frac{dv_i(t)}{dt} = \frac{v_1}{\tau_0 v_0} (1 - \frac{t}{T_l}) v_i(t), \tag{7.9}$$

summing over all the particles and integrating it over the interval $[0, T_l]$ gives:

$$\langle v(T_l)\rangle = v_0 e^{\frac{T_l v_1}{2\tau_0 v_0}}. \tag{7.10}$$

Note that the resulting value is a constant that depends on T_l and the selected initial conditions τ_0, v_0. Figure 7.3c compares the expression of Eq. 7.10 to the average velocity modulus obtained from the simulations. The result of the simulation represents an average in the quasi-stationary regime, over 100 realizations, in a system of $N = 10^4$, with $\rho = 0.02$, $v_0 = \sqrt{2}$, for each value of T_l.

From Eq. 7.2 we define λ,

$$\lambda \equiv \langle n_c\rangle \equiv \frac{T_l}{2v_0\tau_0}\langle v\rangle \tag{7.11}$$

as a characteristic quantity, which gives the number of collisions during the average life time of agents, and using Eq. 7.10, its approximation is given by,

$$\lambda \equiv \langle n_c\rangle = \frac{T_l}{2\tau_0} e^{\frac{T_l v_1}{2\tau_0 v_0}} \tag{7.12}$$

Figure 7.3b presents the comparison between the numerical calculation of λ and Eq. 7.12. Both expressions Eq. 7.10 and Eq. 7.12, estimate better the results for lower values of T_l/τ_0.

In Figure 7.4 we compare the values obtained from simulations of the average degree $\langle k\rangle$ vs. T_l/τ_0 with Eq. 7.13. This is calculated substituting Eq. 7.12 with the previous approximation $\langle k(T_l)\rangle \approx \lambda/2$, which gives:

$$\langle k\rangle = \frac{T_l}{4\tau_0} e^{\frac{T_l v_1}{2\tau_0 v_0}}. \tag{7.13}$$

Equation 7.13 works better for lower densities and lower values of T_l/τ_0.

In contrast, as we show in the inset of Fig. 7.4b, we find that $\langle k\rangle = \lambda/2$ (solid line) is valid for all densities. In fact, given this linear relation $k(\lambda)$, λ being the appropriate control parameter to determine the emergence of a giant connected component in the system. By varying the collision rate λ, one finds a critical value λ_c marking a transition from a state composed by several small clusters to a state where a giant cluster emerges after attaining the QS. In the next Section 7.3 we characterize the critical transition occurring at λ_c.

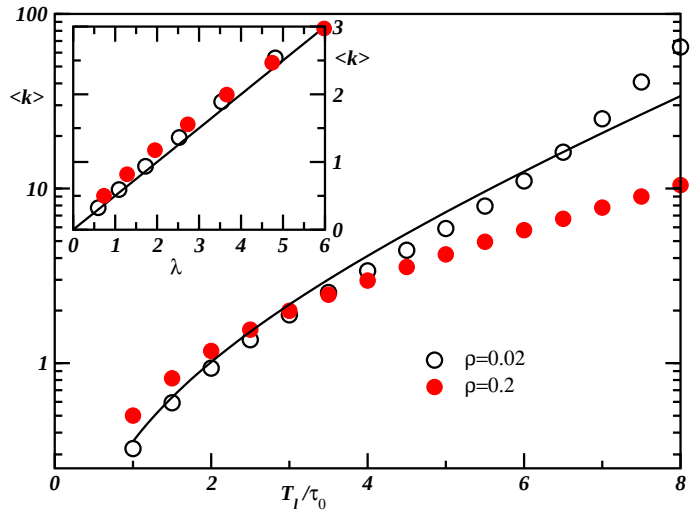

Figure 7.4: Comparison of the average degree $\langle k \rangle$ of the components of the network coming from the simulation (circles) with Eq. 7.13 (solid line). **Inset:** Numerical results of $\langle k \rangle$ vs λ (circles), the straight line is a guide to the eye with slope 0.5. The simulations conditions are the same as those described for Fig. 7.3

7.3 Critical behavior

A cluster is defined as a group of agents connected by links. Note that in contrast to lattice models, were clusters are given by occupied lattice sites connected by nearest-neighbors distances, for this model, each cluster gives a group of agents in contact at a certain time. Isolated agents are regarded as clusters of size unity and any cluster consisting of s connected agents is called an s-cluster. We borrow again the notation from Stauffer's book on percolation theory [81] and define here $n_s = N_s/N$ as the number of s-clusters per agent, where N_s is the number of clusters of size s and N the total number of agents in the system. In Fig. 7.5 we show from left to right: the fraction of agents in the largest cluster G/N, in Fig. 7.5a, the total number of clusters $\sum n_s$, in Fig. 7.5b and the mean size of clusters $\sum s^2 n_s$. As we see qualitatively from the figure for $\lambda = 2.04$ clusters of small sizes coexist in the QS and for $\lambda = 11.04$, there is one giant cluster with a large fraction of the population. For a given value of T_l, we calculate λ numerically in the QS state and average at different times the value of G. The result is plotted in Fig. 7.6b, for different number of agents, $N = L \times L$, with $L =$

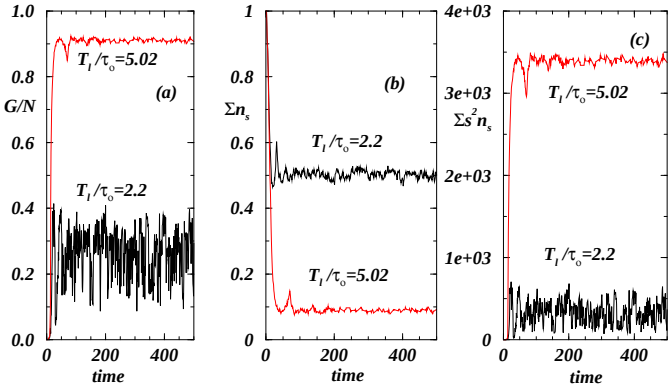

Figure 7.5: **(a)** Fraction of agents in the largest cluster G/N vs. $time$. **(b)** Total number of clusters $\sum n_s$ vs. $time$. **(c)** Mean size of clusters $\sum s^2 n_s$ vs. $time$. Two values of T_l/τ_0 are presented, namely, $T_l/\tau_0 = 2.2$ and $T_l/\tau_0 = 5.02$, other simulations conditions like in Fig. 7.2.

32, 46, 92, 128 and 256. At the same time $\chi \equiv \sum s^2 n_s$ is calculated excluding the largest cluster from the sum (Fig. 7.7b). According to the standard method of scaling theory [121], we expect G and χ to follow

$$G = L^{-\beta/\nu}\mathcal{F}[(\lambda - \lambda_c)L^{-1/\nu}] \tag{7.14}$$

$$\chi = L^{\gamma/\nu}\mathcal{G}[(\lambda - \lambda_c)L^{-1/\nu}]. \tag{7.15}$$

This is confirmed by the collapses of the curves near the transition ($\lambda - \lambda_c \sim 0$), in the Figs. 7.6a and 7.7a, for the values of ν, β, and γ reported in the central column of table 7.1, with $\lambda_c = 2.04$.

In order to test the scaling relation for the cluster numbers. For a system with $N = 2^{16}$ agents, we calculate bins of the cluster size distribution at different values of λ. Following Ref. [81], we take 4 bins ($16-31$, $32-63$, $64-127$, $128-255$) and plot them for different λ, like shown in the inset of Fig. 7.8. For each bin we take the ratio $n_s/n_s(\lambda_c)$ and plot it vs. $(\lambda - \lambda_c)s^\sigma$ using s as the geometric average over the two extremes of the bin size and

$$\sigma = \frac{1}{\beta + \gamma}, \tag{7.16}$$

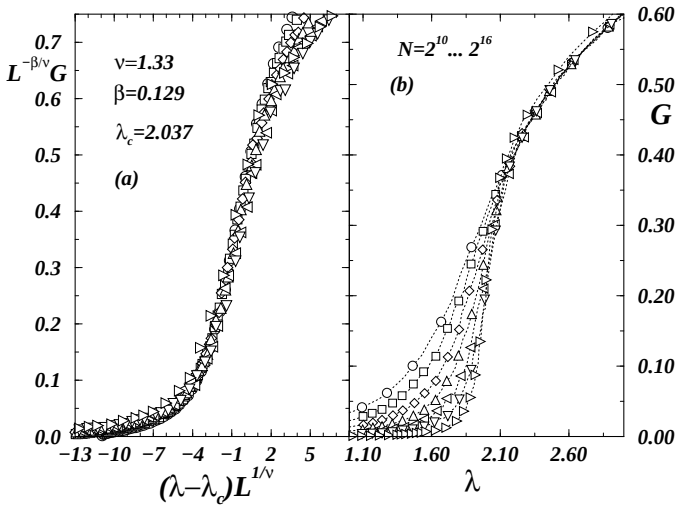

Figure 7.6: **(a)** Confirmation of the scaling relation in Eq. 7.14, for the systems sizes of the right part and the values of β and ν reported in table 7.1 and $\lambda_c = 2.04$. **(a)** Fractions of agents in the largest cluster G/N vs. λ (numerical results in the QS for a fixed value of T_l/τ_0). Results for different numbers of agents: $N = L \times L$, with $L = 32$, 46, 92, 128 and 256.

	MF	mobile agents	Percolation ($2d$)[81]
ν	0.5	1.3 ± 0.1	1.33333
γ	1	2.4 ± 0.1	2.38888
β	1	0.13 ± 0.01	0.13888
σ	0.5	0.40 ± 0.01	0.3956

Table 7.1: Critical exponents related to the emergence of the giant cluster in a random graph model (percolation MF), for the network of mobile agents presented here, compared to the exact results of $2d$ percolation.

using the values of β, γ and λ_c obtained previously. Eq. 7.16, confirms that the scaling relation of the phase transition belongs to the universality class of $2d$ percolation. It is known that the emergence of a giant cluster for a random graph depends on $\langle k \rangle$ with its critical value at $\langle k \rangle_c = 1$, and the phase transition belongs in the same universality class as mean field percolation (see exponents

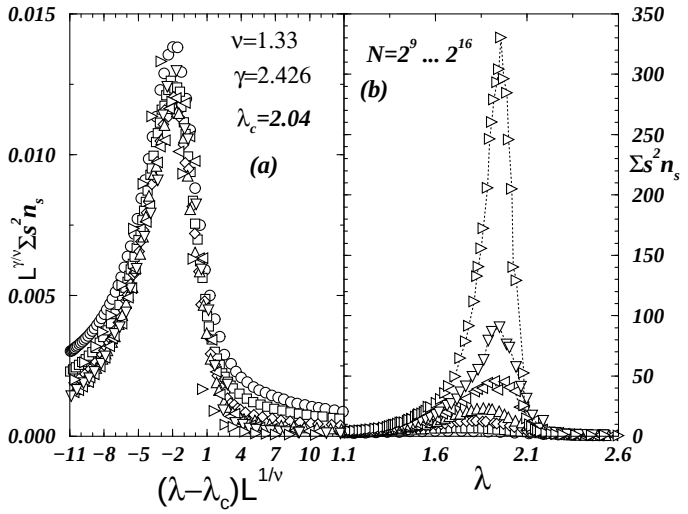

Figure 7.7: **(b)** Confirmation of the scaling relation of Eq. 7.15, for the system sizes of the right part and the values of γ and ν reported in table 7.1 and $\lambda_c = 2.04$. **(a)** Mean cluster size $\sum s^2 n_s$ vs. λ (same simulations as in Fig. 7.6).

in the first column of table 7.1) [122]. For our model we observe, in the same way, the emergence of the giant cluster at $\langle k \rangle_c = 0.5\lambda_c = 1.02$, and in contrast the universality class corresponds to $2d$ percolation, as shown in table 7.1).

Before ending this Section, we stress that the correlation exponent ν presented in the Tab. 7.1 is the one obtained from finite size scaling. This exponent can be also explicitly calculated as the linear size of clusters (see [81]), namely by computing the correlation length $\xi(\lambda)$,

$$\xi^2 = \frac{2\sum_s R_2^s s^2 n_s}{\sum s^2 n_s}. \tag{7.17}$$

Being $2R_s^2$ the average squared distance between two cluster sites, which is calculated for each cluster explicitly from r_i the spatial position of the agents with respect to the position of the center of mass of the cluster $r_0 = \sum_{i=1}^{s} \frac{r_i}{s}$, as follows:

$$R_s^2 = \sum_{i=1}^{s} \frac{|r_i - r_0|^2}{s}. \tag{7.18}$$

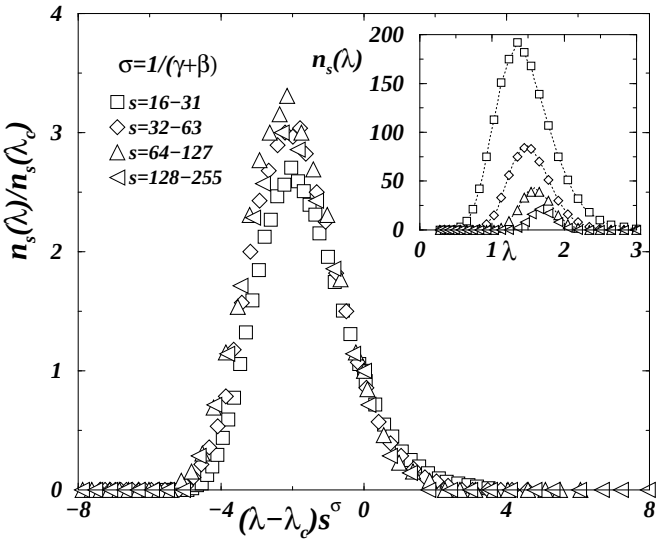

Figure 7.8: Collapse of curves from the inset, plotting $n_s/n_s(\lambda_c)$ vs. $(\lambda - \lambda_c)s^\sigma$ (see details in text). **Inset:** Bins of the cluster size n_s vs. λ, for s =$16 - 31$, $32 - 63$, $64 - 127$, and $128 - 255$.

Thus, apart from a constant factor, the correlation length is the radius of those clusters which give the main contribution to the second moment of the cluster size distribution near the percolation threshold. We expect ξ to diverge as λ approaches λ_c, as

$$\xi(\lambda) \sim |\lambda - \lambda_c|^{-\nu}. \tag{7.19}$$

The result of the numerical calculations of Eq. 7.19 is shown in Fig. 7.9, where the solid line has a slope of -1.3 yielding a correlation exponent in agreement with the previous results (see Tab. 7.1). Since the agents move on a two-dimensional plane and have only a finite life time, they can only establish connections within a restricted vicinity, and this effect corresponds to a connectivity which is short range at each snapshot of the system. Thus, although the clusters in the agent model are not quenched in time, the underlying dynamics yields a short range $2d$ percolation.

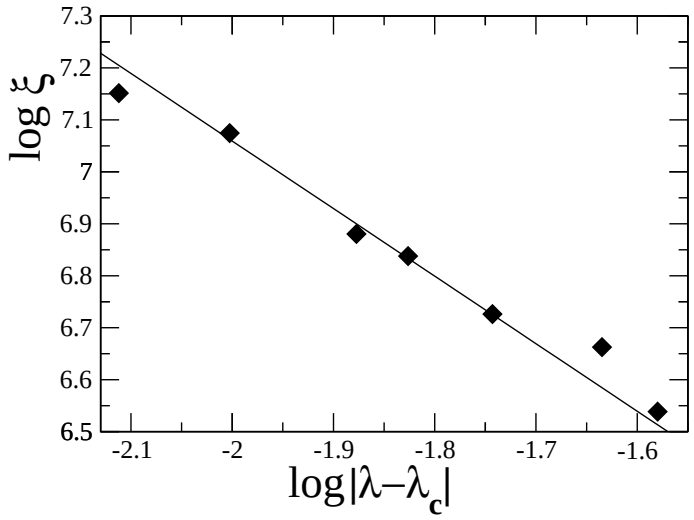

Figure 7.9: Correlation length ξ as a function of $\lambda - \lambda_c$. Symbols indicate the result of simulations performed for different values of λ and the solid line has a slope of $\nu = -1.3$.

7.4 Properties of the network

In this section we will study the degree distribution $P(k)$, the clustering distributions $C(k)$ characterizing the cliquishness of the agents neighborhoods, and the average path length ℓ in the QS state.

In Fig. 7.10 we compute the degree distribution $P(k)$, by counting the fraction of nodes having k neighbors. Two different velocity updates are illustrated here, namely $\alpha = 0$ (triangles) and $\alpha = 1$ (diamonds). Clearly, the degree distribution depends strongly on the collision update rule, i.e. on the value of α in Eq. (6.1).

More precisely, for $\alpha = 0$, the velocity of each agent is always constant, the resulting degree distribution being a consequence of the effective time T_l to create links and of the collision rate, yielding a Poissonian,

$$P(k) = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}. \tag{7.20}$$

For this value $\alpha = 0$, the distribution of degree was calculated for a fixed value of $T_l/\tau_0 = 9.5$ the resulting network has $\langle k \rangle = 2.52$, introducing this value

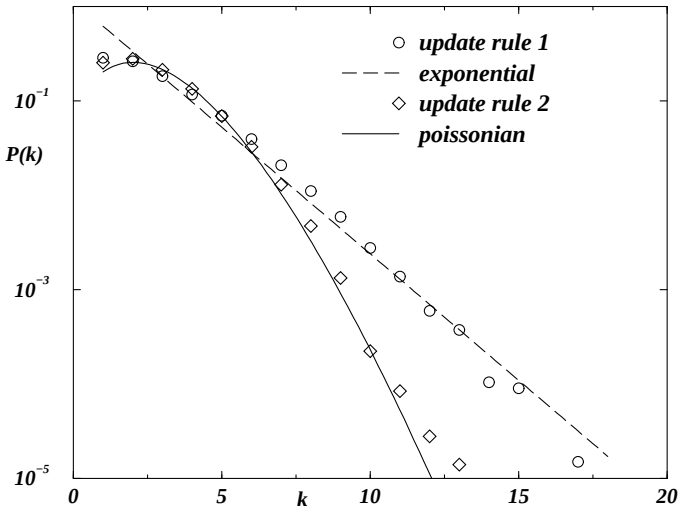

Figure 7.10: Degree distribution of the giant cluster in the quasi-stationary state, averaged over 100 iterations for a system of $N = 10^4$ agents. Results are presented for the velocity update rule 1 with $T_l/\tau_0 = 9.5$ (diamonds), compared to Eq. 7.20 (solid line), evaluating the resulting $\langle k \rangle = 2.52$. And also for the velocity update rule of section 6.1.1, with $T_l/\tau_0 = 3.0$ (circles), compared to Eq. 7.21 evaluated at the resulting $\langle k \rangle = 2.62$.

in Eq.7.20, we obtain the solid line in Fig. 7.10, which shows that the degree distribution of the network is well approximated by a Poissonian. In this way, one can argue that for $\alpha = 0$ the system produces a two-dimensional geometric random graph [124] in the QS state.

For $\alpha = 1$ the velocity in Eq. (6.1) increases linearly with the number of previous collisions. As we will see, this kind of dynamics reveals to be most suited to reproduce the statistical features of real social networks (see Sec. 7.5). In this case, the initial age of agents uniformly distributed (see Sec. 7.1) yields an exponential velocity distribution and consequently an exponential degree distribution

$$P(k) = \frac{1}{\langle k \rangle - 1} e^{-(\frac{k-1}{\langle k \rangle - 1})}. \tag{7.21}$$

In Fig. 7.9 the dashed line indicates Eq. (7.21) evaluated at $\langle k \rangle = 2.62$, which results being $T_l/\tau_0 = 3.0$ used in the numerical simulation (triangles). So, for

this value $\alpha = 1$ the agent model is capable of producing a two-dimensional exponential graphs in the QS state.

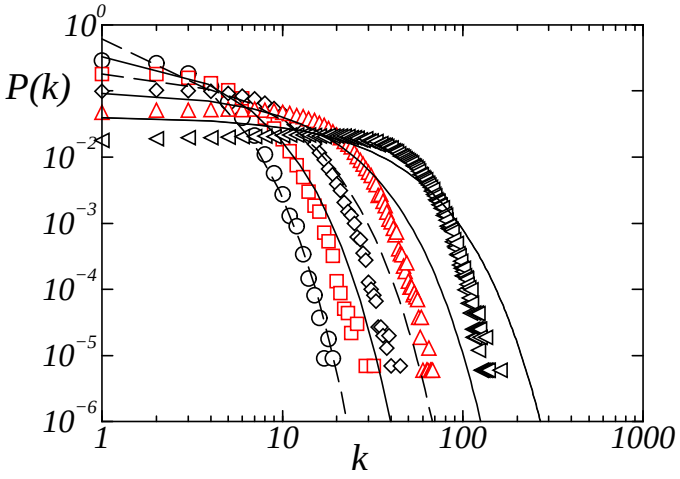

Figure 7.11: **(a)** Degree distribution of the giant cluster in the QS, for several values of $\lambda = 5.24, 8.04, 13.04, 23.76$ and 57.36 (symbols). Lines indicate the corresponding exponential fit with an exponential (see Eq. (7.21)). Here $N = 64 \times 64$ and $\alpha = 1$. For other system sizes the results are similar.

In the following we will consider the value $\alpha = 1$ in Eq. (6.1) and study how the degree distribution depends on the collision rate λ.

Figure 7.11 shows the degree distribution in the QS state for $\lambda = 5.24$ (circles), 8.04 (squares), 13.04 (diamonds), 23.76 (up triangles) and 57.36 (left triangles), plotting with lines the exponential in Eq. (7.21) evaluated with the corresponding $\langle k \rangle$. As one sees, while for small λ the exponential expression fits well the observed degree distributions, for larger $\langle k \rangle$ the numerical results have a lower cutoff than the analytical expression. Therefore, one concludes that the agent model is able to reproduce exponential distributions for low values of the collision rate ($\lambda < 10$) and that other non-trivial distributions are generated increasing the collision rate. The latest ones are the ones observed in empirical social networks (see Sec. 7.5).

Another property of interest to characterize in the network is its clustering coefficient (see Sec. 3.1.4). Figure 7.12 shows the mean degree-dependent clustering coefficient $C(k)$ for the same mobile agents systems in Fig. 7.11. It is interest-

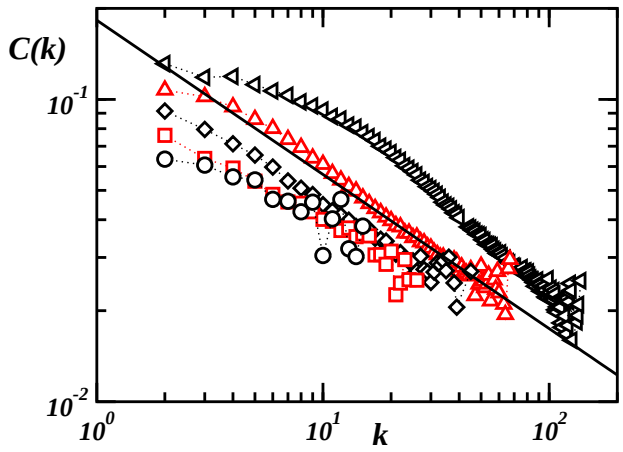

Figure 7.12: Clustering coefficient distribution for the same simulations of Fig. 7.11, namely for (left to right) $\lambda = 5.24, 8.04, 13.04, 23.76$ and 57.36. The solid line is a guide to the eye with slope $1/2$ (see details in text). Here $\alpha = 1$, $N = 128 \times 128$ and averages over 100 realizations were taken.

ing to note that, in contrast to random graphs, which have a clustering coefficient independent on k, here we observe a dependence of the form $\sim k^{-\alpha}$ with $\alpha \in [0.4, 0.6]$. In other words, Fig. (7.12) shows that the clustering coefficient decreases with the degree, a feature which indicates the existence of an underlying hierarchical structure[101]. An analytical approach to understand this structure, remains to be performed and is beyond the scope of this work.

To investigate small-world effects in our system, we compare the networks of the model of mobile agents with random networks having the same number N of nodes and where each pair of nodes is connected with a probability $p = 2M/(N(N-1))$. With this probability one obtains a random network with approximately the same effective coordination M/N (same $\langle k \rangle$) as the one observed in the model of mobile agents.

As seen from Fig. 7.13a, the average path length of the model is small compared to the system size. Moreover, as seen in Fig. 7.13b the cluster coefficient in the agent model (circles) is much larger than in the random counterparts (triangles). However, the networks generated by the model of mobile agents are not small-world, since to have the small-world property[1] it is also required that the increase of the shortest path length with the system size is not faster than $\ln N$.

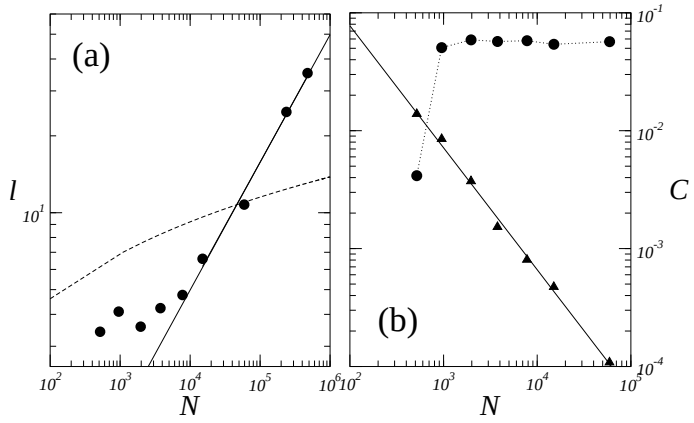

Figure 7.13: Thermodynamic limit of the networks in the system of mobile agents. **(a)** Shortest path length ℓ as a function of the system size N, for $T_l/\tau_0 = 5.02$, compared to $\ln N$ (dashed lines). For large network sizes the solid line indicates a fit $\ell = e^{-3}\sqrt{N}$ (see text). **(b)** Average clustering coefficient (C) as a function of N for the agent model (bullets) compared to the corresponding random graph with the same $\langle k \rangle$ (triangles), having a fit (solid line) of $C \sim 1/N$. Here $\rho = 0.02$.

From Fig. 7.13a one sees that this is only true for small system sizes ($N < 10^4$). For larger systems, the fitted numerical results yield $\ell = e^{-3}\sqrt{N}$.

In Fig. 7.13b one sees the behavior of the average cluster coefficient when the system size is increased. Interestingly, one clearly sees an independence on N, beyond $N > 10^3$, where $C \sim 0.08$. For higher density values ρ this N-independent value of C increases. This result is quite in contrast to ER random graphs which scales with $1/N$, as illustrated in Fig. 7.13b with triangles.

7.5 Real world network of social interactions

In this section we show that the presented model is suitable to reproduce empirical networks. As we showed in the previous section, while for small $\langle k \rangle$ the degree distribution of the giant cluster is exponential of the form of Eq. (7.21), for larger $\langle k \rangle$ it deviates from this shape. The same deviation as $\langle k \rangle$ increases is in fact found in empirical data, from the friendship networks of the 84 schools.

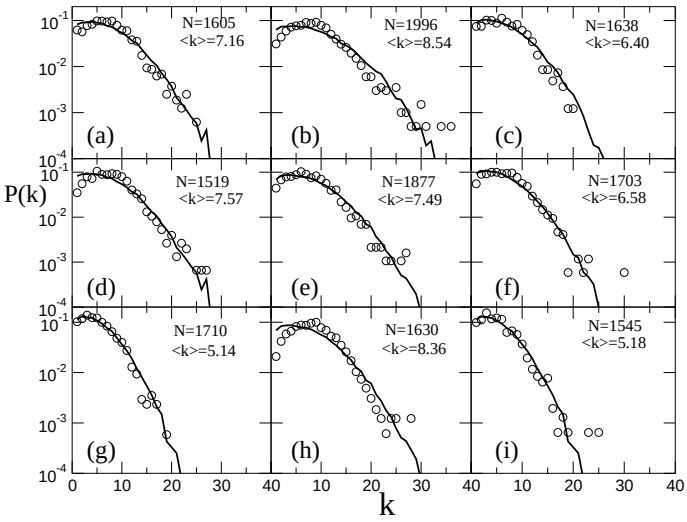

Figure 7.14: Reproducing the distribution of friendship acquaintances in empirical social networks [118] with the agent model. In all cases circles represent the empirical data, while solid line indicate the result of the model of mobile agents.

Figure 7.14 shows several plots of the degree distributions from nine of the schools (circles) compared to the distributions obtained with the agent model (solid lines) for the same number of agents. The values for the collision rate in the simulations were taken from $\lambda = 2\langle k \rangle$ where the average degree is the one found in the corresponding school. As one sees, in all cases the tails of the distributions are well fitted by the model. We found very convincing agreement with the model and the entire database of empirical networks.

For each of the schools, Fig. 7.15a shows the average shortest path length l (circles), and the clustering coefficient C (triangles). Solid lines indicate the results obtained for the agent model using the same range of values of $\langle k \rangle$, averaged over 100 realizations with $N = 2209$ and $\rho = 0.1$. Since l depends on the network size, it is divided by the shortest path length l_0 of a random graph with the same average degree and size. Clearly, the agent model predicts accurately both the clustering coefficient and the shortest path length for the same average degree.

By computing the average degree $\langle k \rangle$ of each school one is able to obtain the value of T_l/τ_0 for which the agent model reproduces properly the empirical data, as illustrated in Fig. 7.15b. Here solid lines indicate the prediction curve for the

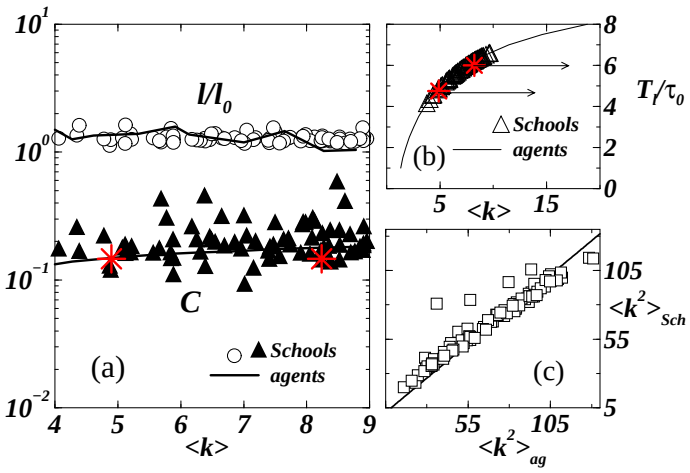

Figure 7.15: Reproducing statistical and topological features of empirical social networks with the model of mobile agents. **(a)** Average shortest path length l and clustering coefficient C as functions of the average degree $\langle k \rangle$. Symbols represent the results of empirical data and dashed lines the simulations. **(b)** Plot of T_l/τ_0 as a function of $\langle k \rangle$ for the agents models (solid line). The curve is used as a prediction for the suitable value of T_l/τ_o to reproduce a given school by knowing its average degree $\langle k \rangle$. Stars illustrate two particular schools for Figs. 7.16 and 7.18 having $T_l/\tau_0 = 4.75$(school 1) and 6.0 (school 2) respectively. **(c)** Second moment $\langle k^2 \rangle$ for each school vs. the second moment calculated for the corresponding simulation with the agent model, the solid line is a guide to the eye with slope one.

agent model, while triangles indicate the values of T_l/τ_0 chosen to reproduce the social network of the schools with the resulting value of $\langle k \rangle$. Moreover, the second moment $\langle k^2 \rangle_{ag}$ obtained with the simulations of the agent model is a rescaling of the same quantity $\langle k^2 \rangle_{Sch}$ measured for the empirical school networks, as shown in Fig. 7.15c.

Figure 7.16a shows the degree distribution averaged over all the schools, compared to the average of the ones obtained from the agent model simulations using the chosen values of T_l according to the relation sketched in Fig. 7.15b. As one clearly sees, the degree distribution obtained with the agent model fits much better the empirical data than the exponential (dotted line) or Poisson (dashed line) distributions for a given $\langle k \rangle$. The inset in figure 7.16a shows the comparison of

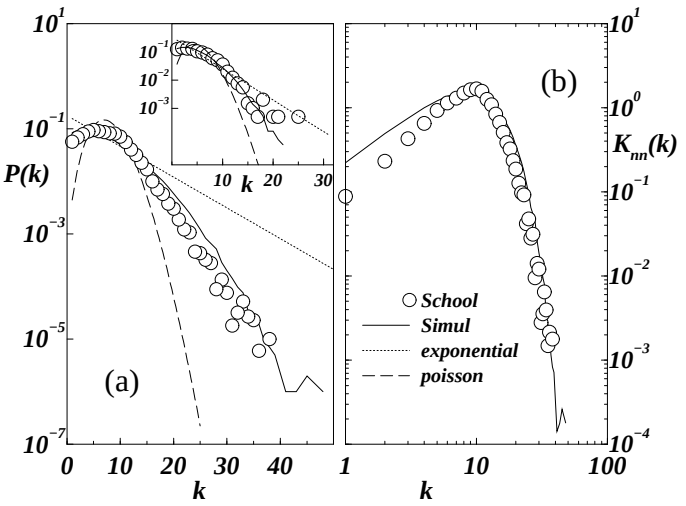

Figure 7.16: **(a)** Degree distribution $P(k)$ averaged over all the schools (symbols) compared to $P(k)$ of the simulations averaged over 10 realizations (solid line) with the given selection of T_l/τ_0 (see Fig. 7.15). The inset shows the results for a particular school (school 1). **(b)** Average degree K_{nn} of the nearest neighbors as a function of k. Dashed and dotted lines indicate the Poisson and exponential distributions respectively, for the same average degree $\langle k \rangle$.

the network of one particular school (school 1 in Fig. 7.15), and the average over 20 realizations of its corresponding model (with $T_l/\tau_0 = 4.75$).

It has been recognized that social networks show degree correlations, in the sense that the degree of the nodes at the end points of a link are not independent [2]. This can be quantified by computing $K_{nn}(k)$, the average degree of the nearest neighbors of a vertex of degree k [115]. Figure 7.16b shows a good agreement of this value between real data and the model for the same networks of Fig. 7.16a. Similar to other social networks the mixing is assortative [2], i.e. K_{nn} increases with k, but in contrast to networks with scale free degree distribution (i.e. collaboration networks), $K_{nn}(k)$ for friendship networks presents a cutoff due to the rapid decay in the degree distribution.

Further, the typical community structure found in social networks, namely the existence of several subnetworks highly interconnected with a few connections between each other, is also reproduced by the agent model. Here, we use a precise

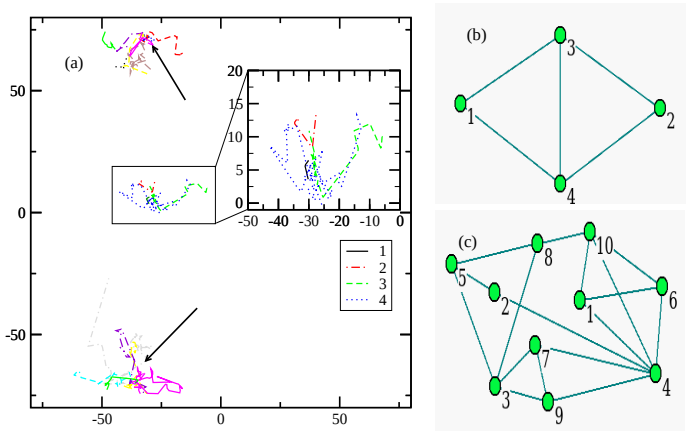

Figure 7.17: **(a)** Enclosed in a box is an illustrative example of trajectories of 4 agents (labeled from 1 to 4) forming a 3-clique sketched in **(b)**. The inset enlarges the region bounded by the box. The arrows indicate two other regions of the system, where one 3-clique community is observed. In these cases the communities are composed by 10 nodes as sketched in **(c)**.

definition of network community recently proposed [21] based on the concept of k-clique community. In Fig. 7.17 we plot the system of mobile agents, drawing only the trajectories of the agents which belong to two 3-clique communities, having 4 and 10 agents and sketched in Fig. 7.17b and Fig. 7.17c respectively. Agents that form a community share a region in space and agents with larger trajectories are responsible for building up the community. It should be pointed out that the agent motion in the system does not have the straightforward meaning of human motion in physical space, but may be better related with affinities among individuals.

Figure 7.18a shows the size distribution of 3-clique communities in a particular school (school 2) compared to the simulation for the suitable value of T_l/τ_0 (see Fig. 7.15), while in Fig. 7.18b the average over all schools is compared to the average over 10 realization of the corresponding model for each school. In both cases, the agent model reproduces the distribution of community sizes observed for the empirical data, particularly the feature related to the existence of a big community having a large fraction of the population, namely $s \sim 10^3$ agents.

In the particular case of sexual contacts it has been reported that the degree distri-

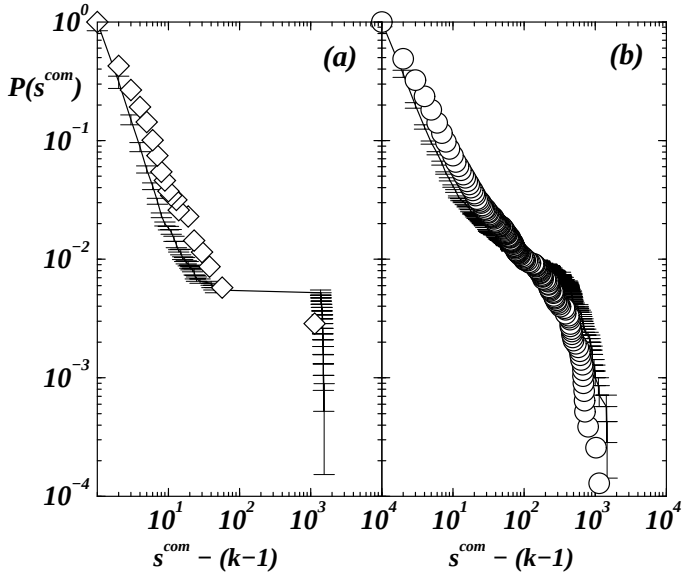

Figure 7.18: **(a)** Distribution of community size s of 3-clique communities for one particular school (school 2) **(b)** the corresponding average over the 84 schools of the data set. Symbols indicate the distribution of the empirical data, while lines indicate the results of the corresponding simulation, with error bars.

bution presents a power-law [83]. The agent model may also reproduce networks with power-law distributions as subsets of the social networks simulated above, by assigning to each agent an additional property. Figure 7.19 shows with triangles the cumulative degree distribution of the sexual contact network presented in Fig. 6.7b from Sec. 6.1.2. The dashed line indicates the degree distribution of a social contact network simulated with the agent model while the solid line is the degree distribution of a subset of contacts from the social network. The contacts in the subset are chosen by assigning to each agent an intrinsic property which enables one to select from all the social contacts the ones which are sexual. Namely, when two agents form a link, as stated before, this link is now marked as a 'sexual contact' if the sum of the property values of the two agents is greater than a given threshold. These values are assigned to the agents with an exponential distribution and the conditional threshold is $\ln N/2$, following the scheme of intrinsic fitness proposed in another context by Caldarelli et. al. [92]. Interestingly, one is able to extract from the typical distributions of social con-

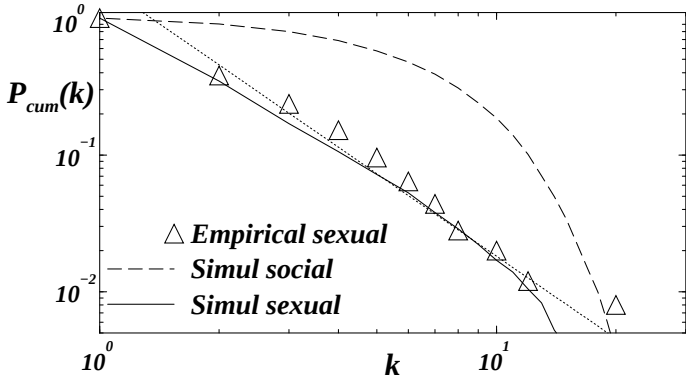

Figure 7.19: Cumulative degree distribution of the number k of sexual partners in a real empirical network of sexual contacts (triangles) with 250 individuals, compared to the simulation of the agent model (solid line), the dotted line is a guide to the eye with slope 2. The simulated sexual contact distribution (dashed line) is in fact a subset of the total social contact distribution (see text). Here $N = 4096$, $T_l/\tau_0 = 5.5$ and $\langle k \rangle = 7.32$ and the average size of the resulting sexual network is 220.

tacts shown throughout the paper, power-law distributions in QS which resemble much the ones observed in real networks of sexual contacts.

7.6 Summary

We presented a novel approach to construct contact networks, based on a system of mobile agents. For a suitable collision rule and aging scheme we have shown that one is able to produce quasi-stationary states which reproduce accurately the main statistical and topological features observed in recent empirical social networks. The QS state of the agent model is fully characterized by one single parameter and yields a phase transition belonging to the universality class of two-dimensional percolation. Moreover, we showed that, by introducing an additional property labeling the ability to select a particular type of social contact, e.g. sexual contacts, the degree distributions reduce to power-law distributions as observed in real sexual networks. Summarizing, we gave evidence that motion of the nodes is a fundamental feature to reproduce social networks, and therefore

the above model could be important to improve the study and may serve as a novel approach to model empirical contact networks.

Chapter 8

Community structure and and role of attributes in friendship networks

8.1 Motivation

Schools are a primary social system, besides the family in which any adolescent participates. They constitute a representative social system where it is possible to identify recurrent patterns of human interactions. The identification of *grouping* patterns in schools can provide a valuable source of information; used to elaborate preventive/organizational plans in direct benefit of the analyzed schools and their surrounding communities. Typical in-school questionnaires provide information about friendship nominations made by students [32, 125]. Such data sets can be organized and analyzed as social networks: nodes are the students and links are the reported friendships.

Previous analysis in the social sciences [8, 32, 125–127] use p models to identify how some of the attributes of the network members are correlated with their inclinations in choosing group relationships. In particular, desegregation of schools as a function of the racial diversity has been a topic of analysis in multi-ethnic countries in Western Europe [125] and the USA [32, 118]. These studies suggested that the way schools organize students could affect the level of racial friendship

Grade	7th	8th	9th	10th	11th	12th
%	17 ± 1	16 ± 2	19 ± 2	17 ± 1	15 ± 1	14 ± 1

Race	White	Black	Hispanic	Asian	Other
%	58 ± 3	14 ± 3	13 ± 2	4 ± 1	10 ± 1

Gender	Female	Male
%	49.6 ± 0.7	49.7 ± 0.7

Clustering Coefficient	Friendship nets.	Community nets.
$\langle C \rangle$	0.32 ± 0.09	0.2 ± 0.1

Shortest-path length	Friendship nets.	Community nets.
$\langle \ell \rangle$	4 ± 1	3 ± 1

Table 8.1: Average over all the schools of the percentage of the students population in each grade ($7th$-$12th$), by different races: white, black, hispanic, asian, mixed (or other), and per Gender (Female and Males). The average of two topological characteristics of the networks of friendship and communities (see explanation in text) is also shown: the clustering coefficient and the average distance among students in the network.

segregation in a school.

Despite these advances, there is clear need to decipher the main statistical features of overlapping community structure in such data sets which is a key question to interpret the global organization of networks. This chapter approaches this issue proposing a different set of tools from those used previously [128–131]. Recent works in the context of statistical physics have approached the analysis of empirical social networks [20, 33, 90, 132–134]. These contributions are strongly related to the study of the structural and hierarchical ordering of the networks, with the aim of characterizing emerging collective phenomena. In this chapter, in addition to a topological study, we incorporate the analysis of attributes, or (*colors*) of the nodes, and show that this is a crucial aspect for identifying preferences from the different ethnic groups. To a *national representative* sample of school networks presented in [32, 118], we apply threshold analysis, do network characterization and community extraction and propose a novel approach for quantifying ethnic preferences. The presented method of analysis gives a new insight into the data.

8.2 Networks of Friendships

The friendship networks presented here are constructed from an in-school questionnaire of the Add-Health [118] friendship nomination study; 90118 students took this survey in 1994-1995. The analyzed data are limited to students who provided information from schools with response rates of 50% or higher. Each student was given a paper-and-pencil questionnaire and a copy of a list with every student in the school. In the data files the activities are summed to create a weighted network. Weights are in the range from 1, meaning the student nominated the friend but reported no activity, to 6 meaning the student nominated the friend and reported participating in all five activities with (him/her).

The structure files contain information on 75871 nodes divided in 84 networks (schools). As a first view into the data, we identify how the whole population is distributed according to the different known attributes, i.e. grade, race and sex, the results are shown in Table 8.1. The amount of students through the different grades and per gender are on average uniform, with percentages of $\sim 16\%$ and $\%50$ respectively. A relevant characteristics, however, is that in most of the analyzed samples of schools the majority of the population is white. In Figs. 8.1a-c, we visualize the friendship networks for two schools with pajek [135]. Nodes represent students, with a different color according to his/her race. A link is drawn among nodes if at least one of the student nominates the other like a friend. In order to provide the maximum information in the visualization of the networks, we introduce artificially by hand the spatial distribution of nodes, corresponding to the different grades, placed counter clockwise, starting with the 7th grade at lower right corner and ending with the 12th grade. One already can see that the links among the grades clearly differentiate the relations among the upper grades (high school) and the lower grades (middle school). Separation of colors within the 6 groups, however, is not artificially introduced; the apparent clustering of nodes according to the same color is due to the fact that they are more densely interconnected. Fig. 8.1a is a characteristic sample of the 84 schools, we call it here, school 1, in this school the great majority of the population is white (70%), which contrasts to two non-characteristic samples, school 2 and school 3 visualized in Figs. 8.1b and c, where blacks (40%) and hispanics (50%) are overrepresented respectively with respect to the average.

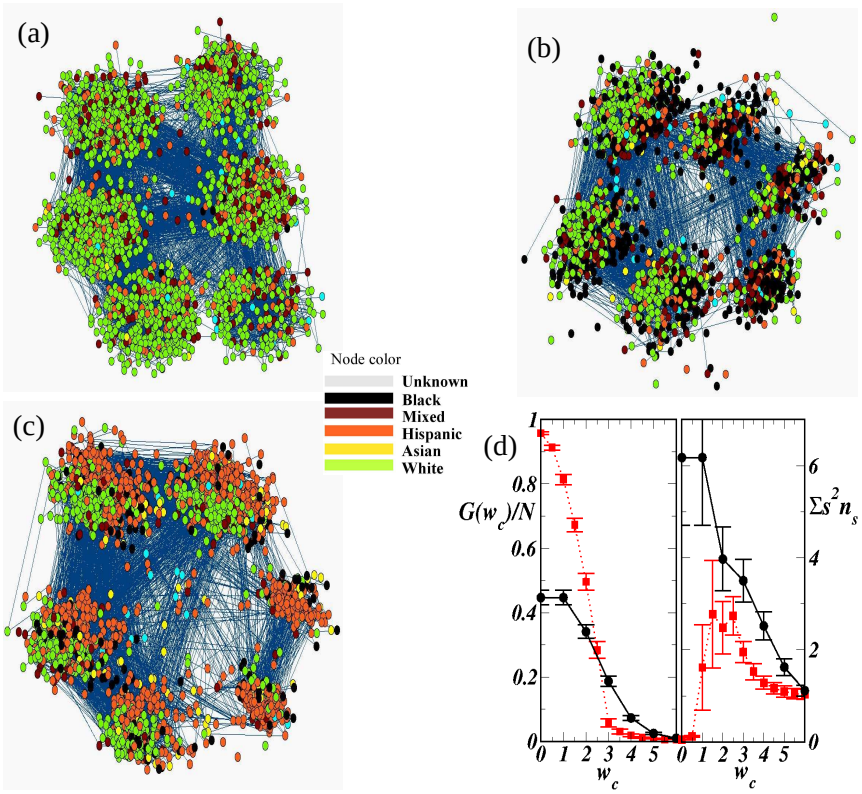

Figure 8.1: **(a)-(c)** Networks of friendships from schools 1, 2 and 3 (respectively). Nodes represent students, with a different color according to his/her race. Spatial distribution of nodes corresponds to the different grades, placed counter clockwise, from 7th to 12th grade. **(d) Left:** G/N fraction of sites in the largest connected component G for the networks with mutual links only (circles) and networks with mutual and not mutual links (squares) versus threshold weight w_c. Only links with average weight in both directions $w \geq w_c$ are kept. **Right:** Average cluster size excluding the largest component for the same analysis as in the left part.

8.3 Role of weights and directionality

In a friendship network, one would expect the presence of mutual links, those are links with nomination in both directions and even the weights should be close to

each other. We apply threshold analysis to measure the influence of weights and directionality in the links.

First, we analyze the network formed only by mutual links, which should have the more reliable information about stronger relations or tight friendships inside the networks. We introduce the mean of the weight in both directions to characterize the weight of each link (w). We examine different thresholds of (w_c) for creating links, i.e. a link is created only if there is a mutual link and $w \geq w_c$. The values of the weights is from 1 to 6, the lighter possible restriction is $w = 1$, which includes any mutual link present in the network with the minimum weight for each direction. In the left part of figure 8.1d (circles) we present the calculations of G/N, the fraction of nodes that belong to the largest cluster vs. w_c. In the right side, $\sum_s s^2 n_s$, the mean cluster size excluding the largest cluster (G) is presented. Interestingly the size of G, when considering only mutual connections is roughly half of the population, and the network is split in various components.

In order to examine the influence of weights, we make the cluster analysis for a threshold weight but now considering the network as undirected, a link is formed if at least one nomination exists, and ($w > w_c$); w is taken as zero if the link in that direction is missing. For this case, we find that the population is connected in a giant component only if weaker links below a critical value of $w_c = 2$ are included. This effect is shown in both sides of Figs.8.1d (red squares).

With the direction and weight analysis we conclude that the population is connected in a single component due to the weak undirected links.

In our further analysis of community detection we assume that a link exists if any of the students nominates the other, and we do not consider any threshold for the weight, a link exists due to the nomination and is independent on the number of activities realized. Imposing the minimum restriction possible for the creation of a link, perhaps we are dealing with a network of acquaintances or affinities instead of real friendships; but those networks tell us something about the preferred social relationships of the students at the schools.

8.4 Community detection: The role of triads

It is widely accepted that most of social networks have a community structure, i.e., posses groups of vertices that have a high density of edges within them, with a lower density of edges between the groups [136, 137]. In order to detect the community structure within each school we use the “clique percolation method”

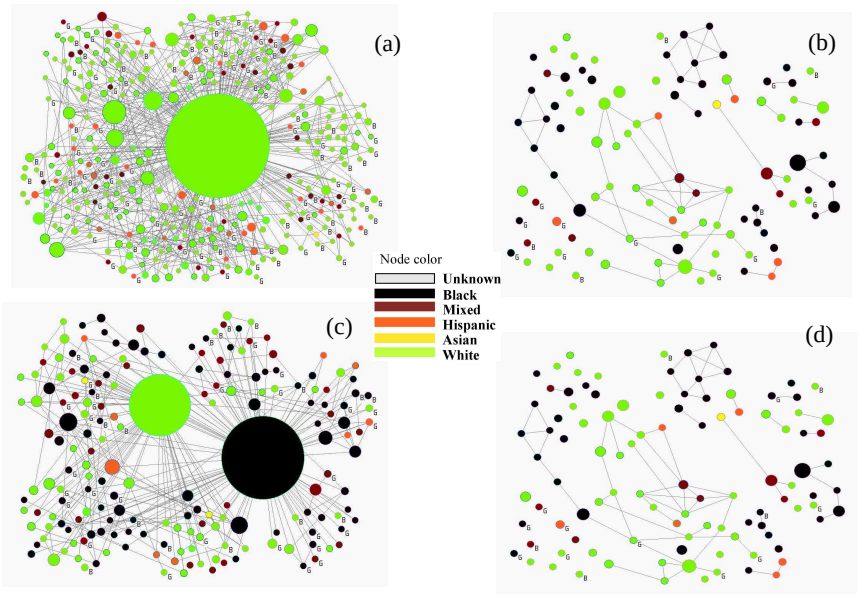

Figure 8.2: Network of 3-clique communities at school 1 (**(a)**) and school 2 (**(c)**). Compared to the corresponding networks of 4-clique communities (**(b)** and **(d)** respectively). The color is assigned according to the race of the majority of nodes in the community and the node size is proportional to the square root of the number of nodes in the community.

, recently proposed in Refs. [21] and [22]. A community is defined as a group of fully connected graphs that share many of their nodes. Thus we refer to a k-clique community, as a union of all k-cliques (complete subgraphs of size k) that can be reached from each other through a series of adjacent k-cliques (i.e. sharing $k-1$ nodes). The set of identified communities from each friendship network constitutes itself, a network: Each community is a node, where the links are represented by the original nodes (students) shared by two communities and the size of each community is the number of its nodes. We show that this method of community detection is a powerful tool for the visualization and representation of the qualitative information from social networks, for this purpose we analyze in detail the information that can be extracted from the two particular schools mentioned above.

Fig. 8.2 shows the network of k-clique communities extracted from the friendship

networks of school 1 and school 2. Figs. 8.2a and c are the 3-clique communities of friendship networks of school 1 and school 2, respectively. In turn, Figs. 8.2b and d are the 4-clique communities extracted from the same schools. The area of the circles represents the number of nodes and each node color is related to a race. Although, each community can have students from different races, we assign to it the color of the majority of the members of the community. We have analyzed also the grades of the majority in each community (not shown) and keep the counter clockwise distribution of nodes per grade. The separation among middle and high schools friendships is observed clearly for school 2 which has two large communities with majority of students from upper ($9th - 12th$) and lower ($7th - 8th$) grades. For school 1, the largest community has students from all the grades.

A particular characteristic that we find for the friendship networks, is that $k = 3$ is the optimal value where a highly structured network of communities are observed (Figs. 8.2a and c) and the great majority of the students belong to any of those communities. In contrast, $k = 4$ detect more cohesive communities and less than 20% of the school populations is present in those communities. For other studied networks [21], like protein networks or collaboration networks, the optimal value for detecting communities is $k = 4$ or 5, in contrast, triads (triangles) are found to be the elementary unit of high school friendship networks.

One of the interesting aspects of such a study is that on the level of more cohesive groups the number of communities becomes balanced even for cases when the ratio of the sizes of the ethnic groups is far from 1. From here (see Figs. 8.2b and d) we conclude that when in minority, the students tend to form stronger and extensive ties, thus, the number of more densely interconnected communities becomes over-represented compared to the $k = 3$ case.

8.5 Complex networks of Communities

Further, we characterize the structure of the extracted networks of 3-clique communities over all the schools. We calculate the degree and size distribution, clustering coefficient, average degree of the nearest neighbors and shortest path length; and compare the results with the original friendship networks.

In Fig. 8.3a we present the cumulative degree distribution, defined as the fraction of nodes having degree larger than k, averaged over the 84 friendship networks at the schools, and compare it with the average of the corresponding 84 networks of communities. The networks of friendships have a scale, corresponding to the

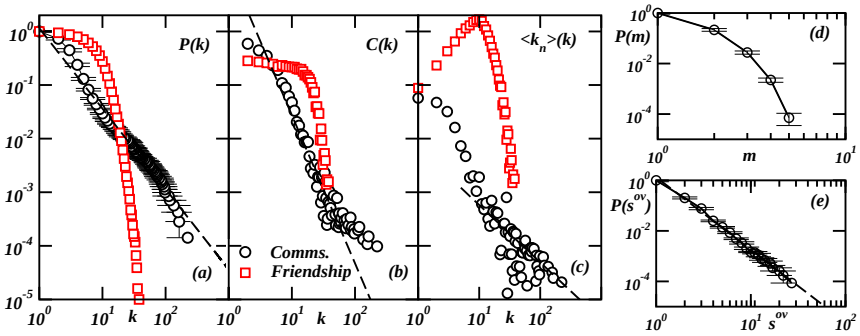

Figure 8.3: Different network properties averaged over the complete dataset of schools, for the community networks (circles) and for friendship networks (squares): **(a)** Cumulative degree distribution. **(b)** Degree-dependent clustering coefficient. **(c)** Average degree of the nearest neighbor. **(d)** Cummulative distribution of the membership number (m) and of **(d)** the overlap size (s^{ov}) for the community networks.

natural cutoff in the number of friends; e.g. $\langle k \rangle \sim 4$. Similar evidence was reported previously for other friendship networks [33]. Interestingly, the degree of the community networks have a much broader scale, following a power-law of the form $\sim k^{-\gamma}$ with $\gamma \sim 1.5$. Thus at the level of communities the 'rich get richer' [138], this information is relevant for the understanding of community formation and should be taken into account for the formulation of models of social networks [133].

Another property of interest to characterize the network is the clustering coefficient. The local clustering coefficient (C_i) of a vertex i with degree k_i, is defined as the ratio of the number of triangles connected to it and all the possible number of triangles ($k_i(k_i-1)/2$). The mean degree-dependent clustering coefficient is the average of the local clustering over all vertices with degree k. This quantity is analyzed for the two types of networks and presented in Fig. 8.3b. For the friendship networks $C(k)$ varies slightly with k for most of the observed k-range; decaying only for larger degrees. In contrast, for the network of communities $C(k)$ follows a power law, $\sim k^{-\alpha}$, with $\alpha \sim 2.8$. This kind of dependence of the clustering coefficient as an inverse power of the node degree, can be signature of hierarchical structure, related to the self-similarity of some complex networks [15, 139, 140].

An interesting quantity in social networks is the average degree of nearest neigh-

bors, $k_{nn}(k)$. Assortativity, is a feature characteristic of social networks e.g., $k_{nn}(k)$ follows a power law with positive exponent, which means that nodes of larger degree tend to be connected among themselves. We find that the analyzed friendship networks, can be identified as assortative because k_{nn} increases with k (Fig. 8.3c), but in contrast to networks with scale free degree distribution (i.e. collaboration networks), $k_{nn}(k)$ present a cutoff due to the rapid decay in the degree distribution. The networks of communities, in contrast, are disassortative, i.e., k_{nn} follows a negative power law, $\sim k^{-\beta}$, with $\beta \sim 1.1$. This indicates that communities with larger degree do no tend to share members. As observed in the two sample schools, large communities or the hubs in the community networks, represent groups with different kind of affinities , i.e. high and middle/junior schools. Those communities in turn share students (have links) with other smaller communities from people of the same age (grade), but do not share many students among themselves. That is why in the networks of communities hubs are not strongly connected and the network is disassortative. Thus, identifying differentiated community hubs with different interests of its members, we expect that other social networks analyzed at the level of communities to be disassortative as well.

For the available data we also calculate the membership (m) of each student, which is the number of communities that the students belongs to. Fig. 8.3d displays the cumulative distribution of the membership number $P(m)$, which shows that on average, each student belongs to a limited number of communities (less than 5). In turn, any two communities can share s^{ov} nodes, which defines the overlap size between these communities. Fig. 8.3e shows the average of the overlap distribution for all the schools, which is a power law with exponent 2.9. We can conclude that a student belongs at most to 4 different clique-communities inside the school, and that there is no characteristic overlap size in the networks. Similar characteristics have been observed in other social and biological networks but not in their randomized versions [21].

Additionally, as shown in the table 8.1, the clustering coefficient, $\langle C \rangle$ for friendship networks and community networks have both a similar average value near 0.3, which is larger than an equivalent random graph of the same size. In turn, the shortest path length, $\langle \ell \rangle$ is less than the logarithm of the number of nodes (not shown). It means both the friendship and community networks present characteristics of small-worlds.

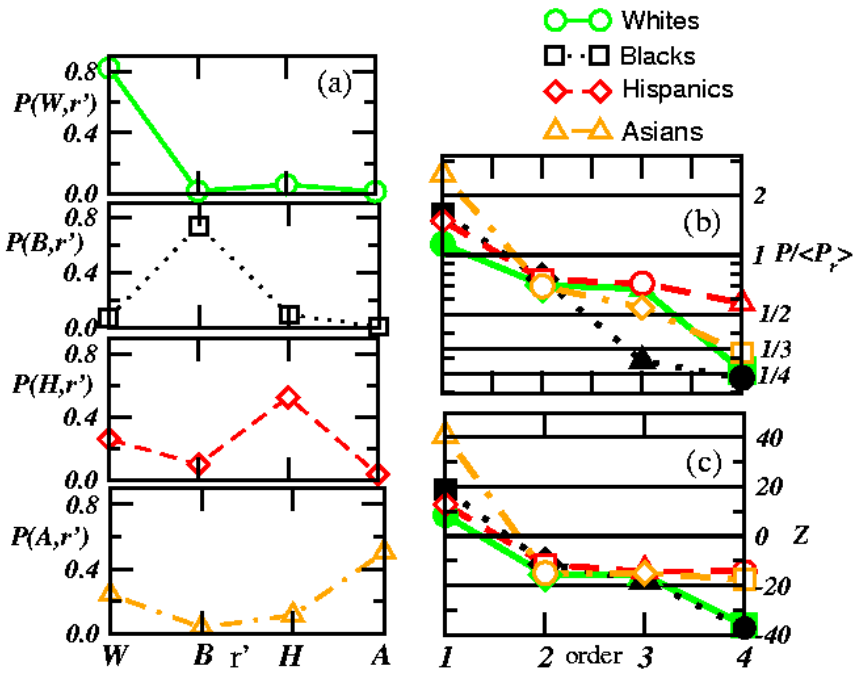

Figure 8.4: Measuring preferences of inter-racial connections $r - r'$. **(a)**Probability of inter racial connections $r - r'$. From top to bottom: probability of directed links from whites ($r = W$), blacks ($r = B$), hispanics ($r = H$) and asians to races $r' = W$ (circles), B (squares), H (diamonds), and A (triangles). Racial preferences manifest themselves as systematic deviations of the ratio $P(r, r')/\langle P_r(r, r')\rangle$ from 1. **(b)** $P/\langle P_r\rangle$ in decreasing order from 1 to 4, for the nominations made from r to r'. **(c)** Statistical significance Z of these deviations. The combination of **(b)** and **(c)** reveals relations $r - r'$ that are significatively absent. The results are the average over the 84 school networks.

8.6 Ethnic preferences

We propose a quantitative method to measure '*preferential*' nominations as a function of the attributes of the students. A nomination can be considered *preferential*, if pairs of nodes with given attributes are significatively more recurrent within the empirical networks than those in their randomized versions. In the studied sample of friendship networks, we find the striking appearance of quan-

titatively preferential nominations among students of the same race, as a characteristic behavior present in each grade and common to each racial group from all schools. Here we present in detail the measure of preferences in the school networks as a function of the race known for the nodes, without separating the information by grade.

In each directed network we identify the frequency of each of the 25 possible race pairs, formed from the 5 races attributed to the nodes. To focus on those pairs that are significatively recurrent, we compare the real network to suitable randomized networks.

The randomized networks have the same single nodes characteristics as those of the real networks: Each node in the randomized network keeps its race and the same number of incoming and outgoing edges as the corresponding node has in the real network. For randomizing the networks we employ a Markov-chain algorithm, based on starting with the real network and repeatedly swapping randomly chosen pairs of connections ($A \to B$, $C \to D$ is replaced by $A \to D$, $C \to B$) until the network is randomized [90, 134]. Switching is prohibited if either of the connections A and D or C and B already exit in any direction. Thus we preserve the number of mutual edges, switching only those with different mutual edges. these deviations. The combination of **(a)** and **(b)** reveals

In Fig. 8.4a from top to bottom we present results for the main four races identified at the schools: white, black, hispanic and asian. For each race r, we plot the probability $P(r, r')$ of existing a directed link $r \to r'$, to a node with race r'. The presented results are the average over the 84 schools, P shows that the common behavior for each racial group is to nominate as friends students with the same race (intra ethnic nominations) more likely than students from any of the other races (inter ethnic nominations). The comparison to randomized networks takes into account the effects of differences in the amount of each race population. Racial preferences manifest themselves as systematic deviations of the ratio $P(r, r')/\langle P_r(r, r')\rangle$ from 1. In Fig. 8.4b, we present $P/\langle P_r\rangle$ in decreasing order from 1 to 4, for the nominations made for each race r and indicating with a different symbol the race of the nominated nodes r'. From this profile becomes clear not only the preferences for intra-ethnic nominations, but also that symmetrically, some inter-ethnic nominations are found 4 times less than in the randomized versions, e.g. those from asians $\leftrightarrow$ blacks and blacks $\leftrightarrow$ whites. In Fig. 8.4c, we show the statistical significance Z of these deviations, defined as:

$$Z(r, r') \equiv \frac{P(r, r') - P_r(r, r')}{\sigma_r(a, a')}, \tag{8.1}$$

where $\sigma_r(r, r')$ is the standard deviation of $P_r(r, r')$ in 100 realizations of a ran-

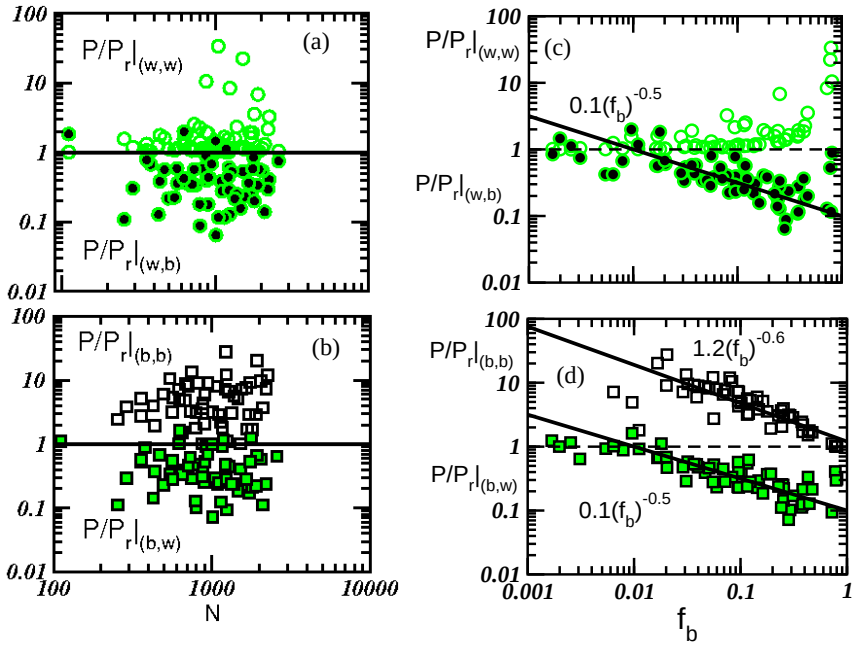

Figure 8.5: Relative significance of nominations (P/P_r) vs. the number of nodes in each school network (N) **(a)**For $white \rightarrow white$ and $white \rightarrow black$ nominations. **(b)** For $black \rightarrow black$ and $black \rightarrow white$ nominations. **(c)** and **(d)** display same results as in **(a)** and **(b)** vs. fraction of the minority, i.e. black population (f_b). P/P_r can be fitted by a negative power law of the form $f_b^{-\alpha}$, with $\alpha = 0.6$. for $black \leftrightarrow black$ nominations and $\alpha = 0.5$ for $black \leftrightarrow white$ nominations. This shows that although heterogeneity decreases the relative frequency of $b \leftrightarrow b$, it does not favor inter-ethnic relations $b \leftrightarrow w$.

domized network. The combination of these two plots reveals relations $r \leftrightarrow r'$ that are significantly absent.

Next, we illustrate how the measured quantity $P(r, r')/\langle P_r(r, r')\rangle$, can be used to obtain certain characteristics of the groups of schools as a whole. In the following, we focus on the relations of two ethnic groups: blacks (b) and whites (w). In Fig. 8.5a we represent the obtained value of P/P_r vs. N, the number of nodes in each school network. We present the values for the nominations from whites, intra ethnic $w \rightarrow w$ and inter ethnic $w \rightarrow b$. Equivalently, in Fig. 8.5b the corresponding nominations from blacks $b \rightarrow b$ and $b \rightarrow w$. These figures show

a sample of 64 schools which have at least 0.2% of any of both races (white and black). In Fig.8.5a and Fig.8.5b, it is clear that intra ethnic nominations occur equally or more frequently than in the randomized networks ($P/P_r \geq 1$), while inter ethnic nominations are less likely to occur ($P/P_r < 1$), and these results do not depend on the total number of the population (N).

When we plot the same quantities as function of the fraction of the minority it is possible to extract some relevant tendencies from the entire sample. Note that P/P_r vs. f_b, for $b \to b$ is greater than unity and tend to one only when $f_b \sim 1$ (top of Fig. 8.5c), just for such values P/P_r of $w \to w$ is then considerably greater than unity (top of Fig. 8.5b). These figures show that both races present the following behavior: When the population of a given race, is the majority (fraction $f \sim 1$), then their intra-ethnic nominations resemble those of the randomized networks $P/P_r \sim 1$, but when they represent a minority ($f \ll 1$) such population tends to make intra-ethnic nominations of friends ($P/P_r \gg 1$).

In contrast, P/P_r for inter-ethnic nominations $w \leftrightarrow b$ can for both groups, be fitted by a negative power-law of the form f_b^{α}, with $\alpha \sim 0.5$ (lower parts of Fig. 8.5b and of Fig.8.5c). These results suggest that the increase of racial heterogeneity does not favor the inter-ethnic nominations among the increasing minority and the race of the majority, but has the opposite effect. A similar conclusion has been reached for the same Add Health data base of friendship nominations with a totally different method of analysis [32].

8.7 Summary

To conclude, a network-theory approach coupled with the analysis of attributes is demonstrably useful in analyzing the structure of friendship nominations at a national representative sample of schools in the USA. The study enables us to unveil a number of significant results. We measured that although racial heterogeneity decreases the relative frequency of intra-ethnic friendship nominations of the minority, it does not favor inter-ethnic relations. Triads are the basic building block of highly structured network of communities which are scale free and dissasortative. Mutual strong nominations of friendship do not form a single connected component in the friendship samples.

The non-trivial community structure has implications in the spreading of information at schools. The reported features related to triads and ethnic preferences give insight into the social organization and should be taken into account in the

formulation of agent based models of social systems. We think that the results and the set of analytical tools presented here will be of interest for further studies.

Chapter 9

General Conclusion

In this thesis different aspects of dynamical interactions on models of complex systems have been investigated. In a first stage, we prescribed the agent interactions through the topology of static networks, i.e. each node interacts with a fixed set of neighbors. We demonstrated that the results of *opinion dynamics* on *stochastic* networks, which are common representations of social networks, can be successfully mapped into *deterministic hierarchical* networks. This result implied both numerical and analytical advantages. First we showed that the use of hierarchical networks considerably lowers the computational time for simulating realistic spreading dynamics, e.g. election processes. Further, we used the deterministic structure to develop an analytical expression that accurately predicts the results of the dynamics of opinion simulated on stochastic networks. The studied kind of interaction came from previously accepted models of opinion dynamics: at each simulation step the resulting *opinion* or state of a node is represented by an *integer* number assigned as a function of its previous state and from the interaction with its neighbors. As a further aplication, we proposed a general opinion model, where each node can have *multiple* opinions each one represented with a *real* number and a *noise* term is added. Even with those generalizations, the kind of transition into the absorbing state of the dynamics showed to depend on the effective dimension of the interaction, and to be independent of the topological correlations of the prescribed networks. Thus, a hierarchical representation is shown to be an appropriate representation to solve analytically general kinds of opinion models.

Another kind of spreading dynamics here investigated was a model of *infection*

propagation. Each agent could be either *susceptible* or *infected* through contagion. An infection lasted a selected *time of infection*, after which, the agent became susceptible again. Resembling real systems, a contagion was only possible through physical contacts among pairs of the type: '*infected-susceptible*', and the possible contacts were *not* prescribed a priori. The agents were free to move in a two dimensional cell and the infection propagated trough *collisions* in the cell. In a first, theoretical approximation, the motion rule was simplified, and given by a Lennard-John interaction potential. We showed that for different densities of cell occupation the dynamics of infection spreading generates a *set* of critical exponents. That is, we obtained a continuous variation of exponents vs. the density of agents, that allowed us to recover in its limits previously known results. In the limit of low densities the exponents coincide with the reported values for *mean-field* calculations, while in the limit of high densities, the obtained exponents reproduce the results reported for *regular* $2d$ *lattices*.

A mobile agent approach showed to be a very convenient way of generation of contact networks. A contact network was generated starting from a colliding pair of agents and keeping track of the successive *contacts*, which in turn, increased the magnitude of the velocity and created the *links* of the resulting network. Introducing a parameter of *selectivity* defined as the probability to *reject* a collision as new *contact*, one could obtain either perfect growing scale-free networks or networks with accelerated growth, in which the average degree of the network increases linearly with time. As an interesting application, the generated networks compare favorably to the topology of empirical data from scale-free networks of sexual contacts, in terms of *degree* distributions and *degree-dependent clustering* distributions. It was shown that the observed numbers of cycles of the empirical networks, are considerably less than from those of their *randomized versions*. In other words, the cycle structure or *clustering* observed in the real networks is not just a direct consequence of its degree distribution, but a non trivial property of that kind of social networks. This observed structure is well reproduced by our proposed model. Both the spatial correlation generated by the movement and the increasing velocity with collisions happen to be the fundamental ingredients in the generation of social contact networks.

This aspect was verified in a stationary version of the model. When introducing a *maximal age* after which an *old* agent is replaced by a *new* one, the resulting contact networks have *single scale* degree distributions and are able to reproduce all the observed topological features of a large data base of friendship networks. With the two free parameters of the models: *the maximal age* and the *density* of agents in the cell, the average *degree* and average *clustering coefficient* could be adjusted to the observed values in the empirical data sets. Due to the spa-

tial correlations, *community structures* emerge naturally in the contact networks generated by the model, which in turn, agree in terms of degree distribution and degree correlations with the empirical friendship networks. The resulting model constitutes a successful approach to reproduce the various non-trivial topological characteristics of social networks with the least number of adjusting parameters.

In order to properly analyze our empirical data sets we had in some cases to develop appropriate tools of analysis, which constituted at the same time, some contributions in the field of analysis of complex networks. For the case of bipartite networks coming from the data of empirical sexual contacts, we defined a clustering coefficient C_4 based on squares or cycles of size four, as a complement to the standard C_3 based on triangles. As a function of both clustering coefficients C_3 and C_4 we developed an expression that gives a good estimate to the numbers of cycles of different sizes in a network. In particular, the given expression is suited for bipartite networks, like it is the case of heterosexual contact networks.

In another context, we developed a method of analysis that uncovered novel aspects in the large data base of analyzed friendship networks. We showed that the weight and directionality of links play an important role in the data set and that mutual strong links do not form a large connected component in the empirical friendship networks. The friendship networks present a rich community structure, making it possible to identify *network of communities*, where each node represents a community or group of friends and two communities are connected by the persons they share. We showed that the measured networks of communities are scale-free and present non trivial degree correlations. A crucial aspect of the data set resulted to be the correlation of pairs of nodes as a function of its colors or *attributes*. Comparing the empirical networks to its randomized versions, we developed an appropriate method to quantify the degree of racial segregation in the analyzed sample of friendship networks. Measures based on correlations among the colors or attributes have been far less investigated than the correlations based on degree or topological patterns. As we showed for the case of friendship networks, this kind of analysis may constitute also a very important source of information in the characterization of other social networks, and also networks in other branches of research.

9.1 Outlook

The results presented along this thesis constitute the standing blocks for some issues that could be further investigated. We outline each of them separated by topics[1]:

- **Opinion dynamics**
  Our results presented some advances in the problem of solving opinion dynamics on static networks. However, the observed independence among the dynamics to consensus and the details of the network structure, as a constant for diverse models of opinions, makes us to question weather a static picture of a network is really an appropriate treatment for this kind of models. Recent literature on affiliation networks has shown that the structure of group memberships plays a major role in the spread of attitudes or opinions which may affect, for instance, individual orientations toward social interactions. These considerations rise the question: ***Does the network structure determine the opinion interactions or the opinion affinity determine the network structure?***. A dynamical interplay of opinion and social network formation is a topic that deserves to be investigated and may lead to more realistic representations of opinion dynamics and their associated networks.[2]

- **Infection Spreading**
  We showed that a dynamical treatment of the interplay among infection time and the time between contacts presents a rich behavior which clearly affects the spreading of infections within a modeled population. An important application of this aspect, concerning sexually transmitted infections, is a systematic study of the infection dynamics incorporating the empirical observations concerning the time ("gap") between the end of an individual's partnership with one sexual partner and the start of the next partnership. The gap can be positive, indicating that there is a non zero interval between the two partnerships, or negative, indicating that the partnership are concurrent, i.e., that they overlap [141]. ***The incorporation of measured gaps with the required topology for the network of sexual contacts and its interplay with infection periods***, is a research topic that deserves more attention.

[1] Currently we are investigating part of the open questions mentioned here.

[2] While writing this manuscript a recent contribution in this area has appeared in http://arxiv.org/pdf/physics/0603023. A model is presented where with a probability ϕ a node convinces a neighbor and with probability $1-\phi$ it creates a new link with a node sharing its same opinion. They showed a continuous phase transition to consensus as a function of ϕ.

- **Networks of Mobile agents**
  The models presented here provide a framework for the study of complex networks as a result of the collisions of mobile agents. The origin of the found properties can be traced back to the very presence of communities, due to the fluctuations in the number of individuals in an abstract representation of a social space. The presented approach establishes a wonderful bridge between the developments of granular gas theory and the modeling of complex networks. In the quasi-stationary regime it is possible to adapt diverse rules of collisions and driving mechanisms investigated in the context of granular gases [89, 142]. For example, a kind of question would be: ***How is characterized the network emerging from collision rules that model the coordination or grouping of animals during their motion, and the swarming processes that are observed ubiquitously in nature?***.

  It is important to mention here that the aim in our method is to develop the simplest model possible, looking at the aggregate level and not at the level of its constituent units. In this sense, our models contrast or perhaps could be seen as a convenient complement of the models in the discipline *agent-based modeling* (ABM) [143] which deals with *individual* complex behavior, including *learning* and *adaptation*. Our *mobile agents* successfully incorporate elements of granular gas theory in the modeling of complex networks. This is clearly a challenging, even disturbing idea.

  With this idea in mind I would like to finish this thesis, borrowing a remark form Philip Ball [*Nature* Editor]:

  '*By seeking to uncover the rules of collective human activities, today's statistical physicists are aiming to return to their roots: Social statistics also guided Maxwell and Boltzmann towards the utilization of probability distributions in the development of the kinetic theory of gases*' Physica A **314** (2002) 1-4.

Bibliography

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

[2] M.E.J. Newman, *SIAM Review* **45** 167 (2003).

[3] S.N. Dorogovtsev, A.V. Goltsev and J.F.F. Mendes, *Phys.Rev. E*, **65**, 066122 (2002).

[4] J.S. Andrade Jr., H.J. Herrmann, R.F.S. Andrade, and L.R. da Silva, *Phys. Rev. Lett.*, **94**, 018702 (2005).

[5] B. Bollobás and O.M. Riordan, *"Modern Graph Theory"* (Springer, Heidelberg, 1998).

[6] B. Bollobás and O.M. Riordan, *"Handbook of Graphs and Networks: From the Genome to the Internet"* (Wiley-VCH, Weinheim, 2003).

[7] S. Boccaletti, V. Latora, Y.Moreno, M. Chavez and D.-U. Hwang, *Physics Reports*, **424**, 175 (2006).

[8] S. Wassermann and K. Faust *"Social Network Analysis: Methods and Applications"*, (Cambridge University Press, 1994).

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

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

[11] R. Sedgewick, *"Algorithms in C"*, 3rd Ed., (Addison Wesley Professional, 1997).

[12] H.J. Herrmann, D.C. Hong, H.E. Stanley, *J. Phys. A: Math. Gen.*, **17**, L261 (1984).

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

[14] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E.*, **68**, 036112 (2003).

[15] E. Ravasz and A-L. Barabási, *Phys. Rev. E*, **67**, 026112 (2003).

[16] S. Shen-Ohrr, R. Milo, S. Mangan and U. Alon *Nature Gen.*, **31**, 64 (2002).

[17] R. Milo, S. Shen-Ohrr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, **298**, 824 (2002).

[18] M. Girvan and M.E.J. Newman, *Proc. Natl. Acad. Sci. USA*, **99**, 7821 (2002).

[19] M. E. J. Newman, *Phys. Rev. E*, **69**, 066133 (2004).

[20] M.A. Porter, P.J. Mucha, M.E.J. Newman and C.M. Warmbrand, *Proc. Natl. Acad. Sci. USA*, **102**, 7057 (2006).

[21] G. Palla, I. Derényi, I. Farkas and Tamás Vicsek, *Nature*, **435**, 814 (2005).

[22] I. Derényi, G. Palla, and T. Vicsek, *Phys. Rev. Lett.*, **94**, 160202 (2005).

[23] J. Duch and A. Arenas, *Physical Review E*, **72**, 027104 (2005).

[24] M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* (2006) in press.

[25] J.L. Morenos, *"Who shall survive? A new approach to the problem of human interrelations."* (Beacon House, New York, (1953)).

[26] P. Marioliss, *Social Science Quarterly*, **56**, 425 (1975).

[27] P.S. Bearman, J. Moody, and K. Stovel, *American Journal of Sociology*, **110**, 44 (2004).

[28] http://imdb.com/.

[29] M.E.J. Newman, *Proc. Natl. Acad. Sci. USA*, **98**, 404 (2001).

[30] W. Aiello, F. Chung, and L. Lu, in J. Abello, P.M. Pardalos and M.G.C Resende eds., *"Handbook of Massive Datasets"*, **97**, (Kluwer, Dordrecht, 2002).

[31] H. Ebel, L.-I Mielsch, and S. Bornholdt, *Phys. Rev. E*, **66**, 035103 (2002).

[32] J. Moody, *Am. Jour. of Soc.*, **107**, 679 (2001).

[33] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, *Proc. Natl. Acad. Sci. USA*, **97**, 11149 (2000).

[34] R. Albert and A.-L. Barabási, *Science*, **286**, 509 (1999).

[35] E. M. Jin, M. Girvan, and M. E. J. Newman, *Phys. Rev. E*, **64**, 046132 (2001).

[36] M. E. J. Newman, *Proc. Natl. Acad. Sci. USA*, **101**, 5200 (2004).

[37] W.W. Zachary, *J. Anthropol. Res.*, **33**, 452 (1977).

[38] L. Donetti and M. A. Muñoz, *J. Stat. Mech.*, P10012 (2004).

[39] M.E.J. Newman, J. Park, *Phys. Rev. E*, **68**, 036122 (2003).

[40] S. Wolfram, *"Cellular Automata and Complexity"*, (Adison-Weasly, Reading, MA, 1994).

[41] J. Marro and R. Dickman, *"Nonequilibrium phase transitions in lattices models"*, (Cambridge University Press, Cambridge , 1999).

[42] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E*, **63**, 066117 (2001).

[43] K. Sznajd-Weron and J. Sznajd, *Int. J. Mod. Phys. C* **11**, 1157 (2000).

[44] G. Defuant, D. Neau, F. Amblard, G. Weisbuch, *Adv. Complex Syst.*, **3**, 87 (2000).

[45] M. Hegselmann and J. Krause, *JASS*, **5**, 3 (2002).

[46] P.L. Krapivsky and S. Redner, *Phys. Rev., Lett.*, **90**, 238701 (2003).

[47] S. Galam, *Physica A*, **285**, 66 (2000).

[48] R. Axelrod, *J. Conflict. Resolut.*, **41**, 203 (1997).

[49] D. Stauffer, S. Moss de Oliveira, P.M.C. de Oliveira, and J.S. Sa Martins, *"Biology, Sociology, Geology by Computational Physicists"*, (Elsevier, Amsterdam, 2006); D. Stauffer, *JASS*, **5**, 1 (2002).

[50] A.-L. Barabási, E. Ravasz and T. Vicsek, *Physica A*, **299**, 3 (2001).

[51] **http://en.wikipedia.org/wiki/Opinion**

[52] G. Weisbuch, G. Defuant, F. Amblard, J.-P. Nadal, *Complexity* **7**, 55 (2002); G. Defuant, F. Amblard, G. Weisbuch, T. Faure, *JASS*, **5**, 4 (2002); R. U. Krause, *"Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren"*, (Krause and M. Stöckler eds., Bremen University, 1997).

[53] S. Galam, *Eurphys. Lett.*, **70**, 705 (2005).

[54] A.T. Bernardes, D. Stauffer, and J. Kertész, *Eur. Phys. J. B*, **25**, 123 (2002).

[55] D. Stauffer, A.O. Sousa, and S. Moss de Oliveira, *Int. J. Mod. Phys. C*, **11**, 1239 (2000).

[56] J. Bonnekoh, *Int. J. Mod. Phys. C*, **14**, 1231 (2003).

[57] C. Schulze, *Int. J. Mod. Phys. C*, **14**, 95 (2003).

[58] R. Ochrombel, *Int. J. Mod. Phys. C*, **13**, 1091 (2001).

[59] K. Sznajd-Weron and J. Sznajd, *Int. J. Mod. Phys. C*, **13**, 115 (2002).

[60] **http://www.tse.gov.br/ele/**.

[61] R.N. Costa Filho, M.P. Almeida, J.S. Andrade, Jr., and J.E. Moreira, *Phys. Rev. E.*, **60**, 1067 (1999); M.L. Lyra, U.M.S. Costa, R.N. Costa and J.S. Andrade, *Europhys. Lett.*, **62**, 131 (2003); R.N. Costa, J.E. Moreira and J.S. Andrade, *Physica A*, **322**, 698 (2003).

[62] **http://www.indian-elections.com/states/**.

[63] A.O. Sousa, *Physica A*, **348**, 701 (2005).

[64] C. Schulze, *Int. J. Mod. Phys. C*, **15**, 867 (2004).

[65] D. Stauffer and P.M.C. de Oliveira, *Eur. Phys. J. B*, **30**, 587 (2002).

[66] F. Slanina and H. Lavička, *Eur. Phys. J. B*, **30**, 587 (2002).

[67] M.C. González, A.O. Sousa, and H. J. Herrmann, *Int. J. Mod. Phys. C*, **15**, 45 (2004).

[68] P.J. Reynolds W. Klein and H. E. Stanley, *J. Phys C*, **10**, L167 (1977).

[69] M.C. González, P.G. Lind and H.J. Herrmann, *Eur. Phys. J. B*, **49**, 371 (2006). `physics/0508145`.

[70] M.C. González, A.O. Sousa and H.J Herrmann, *Eur. Phys. Jour. B*, **49**, 253 (2006).

[71] H. Hinrichsen, *Adv. Phys.*, **49**, 815 (2000).

[72] D. Mollison, *J.R. Statist. Soc. B*, **39**, 283 (1977).

[73] A. R. Kerstein, *Phys. Rev. B*, **30**, 2980 (1984).

[74] P. G. de Gennes, *Hebd. C.R. Séan. Acad. Sci.*, **286B**, 131 (1978).

[75] B. Hede, *J. Phys. A*, **22**, L439 (1989).

[76] D. C. Rapaport, *"The Art of molecular dynamics simulation"* (Cambridge University Press, Cambridge, 1995).

[77] P. Grassberger, *J. Phys. A: Math. Gen.*, **22**, 3673 (1989).

[78] M. Lägues, *C. R. Acad. Sci.*, **288**, 339 (1979).

[79] B. Loh and R. Jn Song, *Dengue Bulletin*, **25**, 74 (2001).

[80] M. C. González and H. J. Herrmann, *Physica A*, **340**, 741 (2004).

[81] D. Stauffer, *"Introduction to Percolation Theory"* (Taylor & Francis, London, 1985).

[82] M. Morris, *Nature*, **365**, 437 (1993).

[83] F. Liljeros, C.R. Edling, L.A.N. Amaral and H.E. Stanley, *Nature*, **411**, 907 (2001).

[84] A. Schneeberger, C. Mercer, S.A.J. Gregson, N.M. Ferguson, C.A. Nyamukapa, R.M. Anderson, A.M. Johnson, G.P. Garnett, *Sex. Trans. Dis.*, **31**, 380 (2004).

[85] V. Latora, A. Nyamba, J. Simpore, B. Sylvette, S. Diane, B. Sylvére, S. Musumeci, *J. Med. Virology*, **78**, 724 (2006).

[86] A.S. Klovdahl, *Soc. Sci. Med.*, **28**, 25 (2001).

[87] T. Pöschel and S. Luding, *"Granular Gases"*, (Lecture Notes in Physics, 564, Springer-Verlag, 2001).

[88] E.O. Laumann, J.H.Gagnon, R.T Michaels, *"Organization of Sexuality"*, (University of Chicago Press, 1994).

[89] D. ben-Avraham, E. Ben-Naim, K. Lindenberg and A. Rosas, *Phys. Rev. E*, **68**, 050103 (2003).

[90] S. Maslov and K. Sneppen, *Science*, **296**, 910 (2002).

[91] K.T.D. Eames and M.J. Keeling, *Proc. Nat. Ac. Sci.*, **99**, 13330 (2002).

[92] G. Caldarelli, A. Capocci, P. De Los Rios, M.A. Muñoz, *Phys. Rev. Lett.*, **89**, 258702 (2002).

[93] J.J. Potterat, L. Phillips-Plummer, S.Q. Muth, R.B. Rothenberg, D.E. Woodhouse, T.S. Maldonado-Long, H.P. Zimmerman, J.B. Muth, *Sex. Transm. Infect.*, **78**, i159 (2002).

[94] J.L. Wylie and A. Jolly, *Sex. Transm. Dis.* **28**, 14 (2001).

[95] P.N. McGraw and M. Menzinger, *Phys. Rev. E*, **72**, 015101(R) (2005).

[96] P.F. Stadler, A. Wagner and D.A. Fell, *Adv. Complex Sys.*, **4**, 207 (2001).

[97] M.E.J. Newman, *Social Networks*, **25**, 83 (2003).

[98] P. Holme, C.R. Edling and F. Liljeros, *Social Networks*, **26**, 155 (2004).

[99] P. Holme, F. Liljeros, C.R. Edling and B.J. Kim, *Phys. Rev. E*, **68**, 056107 (2003).

[100] R. Guimerà, X. Guardiola, A. Arenas, A. Díaz-Guilera, D. Streib and L.A.N. Amaral, *"Quantifying the creation of social capital in a digital community"*, private report.

[101] G. Caldarelli, R. Pastor-Satorras and A. Vespignani, *Eur. Phys. J. B*, **38**, 183 (2004).

[102] E. Marinari and R. Monasson, *J. Stat. Mech.*, P09004 (2004).

[103] H.D. Rozenfeld, J.E. Kirk, E.M. Bollt and D. ben-Avraham, *J. Phys. A* **38**, 4589 (2005).

[104] G. Bianconi and A. Capocci, *Phys. Rev. Lett.* **90**, 078701 (2003).

[105] G. Bianconi, G. Caldarelli and A. Capocci, *Phys. Rev. E*, **68**, 066116 (2005).

[106] C.P. Herrero, *Phys. Rev. E*, **71**, 016103 (2005).

[107] S.-J. Yang, *Phys. Rev. E*, **71**, 016107 (2005).

[108] G. Bianconi and M. Marsili, *JSTAT*, P06005 (2005).

[109] A. Vázquez, J.G. Oliveira and A.-L. Barabási, *Phys. Rev. E*, **71**, 025103(R) (2005).

[110] S.N. Soffer and A. Vázquez, *Phys. Rev. E*, **71**, 057101 (2005).

[111] M.E.J. Newman and D.J. Watts, *Phys. Rev. E*, **60**, 7332 (1999).

[112] H.-J. Kim and J.M. Kim, *Phys. Rev. E*, **72**, 036109 (2005).

[113] T. Petermann and P. de los Rios, *Phys. Rev. E*, **69**, 066116 (2004).

[114] K. Klemm and P.F. Stadler, *Phy. Rev. E*, **73**, 025101 (2006).

[115] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, *Phys. Rev. E.*, **70**, 056122 (2004).

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

[117] D.J. Watts, P.S. Dodds and M.E.J. Newman, *Science*, **296**, 1302 (2002).

[118] This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant from the National Institute of Child Health and Human Developtment (P01-HD31921). Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).

[119] S.N. Dorogovtsev and J.F.F. Mendes, *Phys. Rev. E*, **62**, 1842 (2000).

[120] L.E. Reichl, *"A modern course in statistical physics"*, (Ed. John Wiley & Sons, Canada, 1995, 2nd ed.).

[121] M.E.J.Newman and G. T. Barkema *"Monte Carlo methods in statistical physics"*, (Clarendon Press, Oxford, 1999).

[122] K. Christensen, R. Donangelo, B. Koiler, and K. Sneppen, *Phys. Rev. Lett.*, **81**, 2380 (1998).

[123] P.G. Lind, M.C. González and H.J Herrmann , *Phys. Rev. E*, **72**, 056127 (2005).

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

[125] C. Baerveldt, M.A.J. Van Duijn, L. Vermeij and D.A. Van Hemert, *Social Networks*, **26**, 55 (2004).

[126] C.J. Anderson, S. Wassermann and B. Crouch, *Social Networks*, **21**, 37 (1999).

[127] C. Baerveldt, R. Van Rossem, M. Vermande, and F. Weerman, *Connections*, **26**, 11 (2004).

[128] J. Moody, *Social Networks*, **23**, 261 (20001).

[129] I. Jansson, *Social Networks*, **21**, 339 (1999).

[130] I. Jansson, *Social Networks*, **19**, 285 (1997).

[131] D.D. Brewer and C.M. Webster, *Social Networks*, **21**, 361 (1999).

[132] V. Colizza, A. Flammini, M. A. Serrano and A. Vespignani, *Nature Physics*, **2**, 110 (2006).

[133] M.C. González, P.G. Lind, H.J. Herrmann, *Phys. Rev. Lett.*, **96**, 088702 (2006).

[134] R. Milo, S, Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Ohrr, I. Ayzenshtat, M. Scheffer and U. Alon, *Science*, **303**, 1538 (2004).

[135] http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[136] J. Scott, *"Social Network Analysis: A Handbook"* 2nd Ed., (Sage, London, 2000).

[137] B.S. Everitt, *Cluster Analysis* 3rd Ed., (Edward Arnold, London, 1993).

[138] P. Pollner, G. Palla and T. Vicsek, *Europhys. Lett*, **73**, 478 (2006).

[139] G. Szabó, M.J. Alava and J. Kertész, *Phys. Rev. E*, **67**, 056102 (2003).

[140] E. Song, S. Havlin, H.A. Makse, *Nature*, **433**, 392 (2005).

[141] B. Foxman, M.E.J Newman, B. Percha, K. K. Holmes, and S. O. Aral, *Sex. Trans. Dis.*, **33**, 209 (2006).

[142] R. Cafiero, S. Luding and H.J. Herrmann, *Phys. Rev. Lett.*, **84**, 6014 (2000).

[143] E. Bonabeau, *Proc. Natl. Acad. Sci. USA*, **99**, 7280 (2002).

Acknowledgments

Working with my advisor, professor Hans Herrmann, has been for me a great honor and also a pleasure. I arrived to his group a bit more than three years ago with big expectations about research on Statistical Physics. Today, thanks to his advise, I am absolutely satisfied with the results and I think I got from him the taste of what really means doing good research in this wonderful area. I thank him for the combination of being a very generous person and an extraordinary bright mind.

Next to thank is P.G. Lind, my super collaborator in most of the results of this research. He is the best co-worker one can ever imagen to have: Excellent professional and a wonderful person.

I also had fruitful collaborations with other great professionals, some of them, now my best friends: A.D. Araújo, V. Schämmle, A.A. Moreira, A. Mora and A.O. Sousa.

I had the fortune of doing some work in collaboration with some other professors, from whom I learned a lot and enjoyed the hospitality within their groups, my special thanks to T. Vicsek and J. Kertész for my visit to Budapest and A.S. Andrade for my stay in Fortaleza, Brazil.

I have beneficiated from generous support and fruitful discussions from many great researchers. I am grateful to: D. Stauffer, M. Lässig, M. Barthélémy, M. Paczuski, K. Sznajd-Weron, J.A.C. Gallas, R. García-Rojo, H-J. Seybold, I. Farkas, R.C. Hidalgo, L.E. Guerrero, M. Haase, B. Biswal, L. Trujillo and F. Osanloo.

Now, my everyday co-workers. I thank without exception all the members of the ICP at Stuttgart University, they self-organize and create a perfect environment of joy and solidarity at work which I will surely miss. I am in debt to V. Schämmle, C. Kunert and M. Hecht for their valuable support in the corrections to the German resume of this thesis. I thank also from there, the best system administrator I have ever met, Frank Huber, who kindly helped me with problems concerning

the system and everything else. Marlies Parsons, is our super retired secretary at the ICP, for me she is like a relative in Germany, always a source of warm support and the perfect advise. I thank H. Patzelt, the new secretary, in a short time, also a great support for me.

I am in debt to the Studentenwerk Stuttgart, a wonderful organism that really facilitates in all aspects the Students Life in Stuttgart. Special thanks to Mrs. Grime and the Hausmeister.

I acknowledge a scholarship (A/02/22803) from the Deutscher Akademischer Austausch Dienst (DAAD), Germany. They are a top class agency in support of the Scientific Research world wide, with extraordinary Professionalism. Special thanks to Mrs. V. Metje.

No hay suficientes palabras para agradecer a Ricardo, mi gran compañero, mi camarada y cómplice de vida. Emiliano, simplemente por ser él, sólo por existir. Mis padres, hermanos y demás familiares, por su amor y apoyo incondicional a lo largo de este proyecto.

Dedicated to Ricardo and Emiliano,
love, passion and inspiration.
to my parents,
for their never ending support.

Marta C. González
May 4th, 2006
Stuttgart, Germany