UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Físicas y Nanotecnología

TÍTULO: Sociodynamical models on real social networks

Trabajo de integración curricular presentado como requisito para la obtención
del título de Físico

Autor:

Cárdenas Sabando Ronald Andrés

Tutor:

Dr. COSENZA MICELI, MARIO GIUSEPPE , Ph.D.

Urcuquí, mayo de 2021

Urcuquí, 27 de mayo de 2021

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS FÍSICAS Y NANOTECNOLOGÍA
CARRERA DE FÍSICA
ACTA DE DEFENSA No. UITEY-PHY-2021-00004-AD

A los 27 días del mes de mayo de 2021, a las 14:30 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. SVOZILIK JIRI , Ph.D.
Miembro No Tutor	Dr. MOWBRAY , DUNCAN JOHN , Ph.D.
Tutor	Dr. COSENZA MICELI, MARIO GIUSEPPE , Ph.D.

El(la) señor(ita) estudiante **CARDENAS SABANDO, RONALD ANDRES**, con cédula de identidad No. **1723921985**, de la **ESCUELA DE CIENCIAS FÍSICAS Y NANOTECNOLOGÍA**, de la Carrera de **FÍSICA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-39-No.456-2014**, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: **Sociodynamical models on real social networks.**, previa a la obtención del título de **FÍSICO/A.**

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. COSENZA MICELI, MARIO GIUSEPPE , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Tutor	Dr. COSENZA MICELI, MARIO GIUSEPPE , Ph.D.	10,0
Miembro Tribunal De Defensa	Dr. MOWBRAY , DUNCAN JOHN , Ph.D.	9,9
Presidente Tribunal De Defensa	Dr. SVOZILIK JIRI , Ph.D.	9,8

Lo que da un promedio de: **9.9 (Nueve punto Nueve)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

Certifico que *en cumplimiento del Decreto Ejecutivo 1017 de 16 de marzo de 2020, la defensa de trabajo de titulación (o examen de grado modalidad teórico práctica) se realizó vía virtual, por lo que las firmas de los miembros del Tribunal de Defensa de Grado, constan en forma digital.*

CARDENAS SABANDO, RONALD ANDRES
Estudiante

RONALD ANDRES CARDENAS SABANDO
Firmado digitalmente por RONALD ANDRES CARDENAS SABANDO
Fecha: 2021.05.31 16:15:19 -05'00'

Dr. SVOZILIK JIRI , Ph.D.
Presidente Tribunal de Defensa

Firmado electrónicamente por:
JIRI SVOZILIK .

Dr. COSENZA MICELI, MARIO GIUSEPPE , Ph.D.
Tutor

Firmado electrónicamente por:
MARIO GIUSEPPE COSENZA MICELI

UNIVERSIDAD YACHAY TECH

Firmado electrónicamente por:
DUNCAN JOHN MOWBRAY .

Dr. MOWBRAY , DUNCAN JOHN , Ph.D.
Miembro No Tutor

EVELYN CAROLINA CIFUENTES TAFUR
Digitally signed by EVELYN CAROLINA CIFUENTES TAFUR
Date: 2021.05.27 16:49:04 -05'00'

CIFUENTES TAFUR, EVELYN CAROLINA
Secretario Ad-hoc

Autoría

Yo, **Ronald Andrés Cárdenas Sabando**, con cédula de identidad **1723921985**, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.
Urcuquí, Mayo del 2021.

RONALD ANDRES CARDENAS SABANDO
Firmado digitalmente por RONALD ANDRES CARDENAS SABANDO
Fecha: 2021.05.31 17:12:59 -05'00'

Ronald Andrés Cárdenas Sabando
CI: 1723921985

Autorización de publicación

Yo, **Ronald Andrés Cárdenas Sabando**, con cédula de identidad **1723921985** cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, Mayo del 2021.

RONALD ANDRES CARDENAS SABANDO

Firmado digitalmente por RONALD ANDRES CARDENAS SABANDO
Fecha: 2021.05.31 17:13:28 -05'00'

Ronald Andrés Cárdenas Sabando
CI: 1723921985

Dedication

I dedicate this thesis to my mother. Without your immeasurable personal sacrifice and unconditional love, I would have not become the individual that I am today. I hope this is one of many achievements that we will celebrate together.

Acknowledgements

First and foremost, I am extremely grateful to my advisor, Professor Mario Cosenza, Phd, for his invaluable advice, continuous support, motivation, enthusiasm, and patience during the creation of this thesis. Without his persistent help, the goals of this project would not have been realized. I could not have imagine having a better advisor than him.

I would like to thank the Corporación Ecuatoriana para el Desarrollo de la Investigación y Academia (CEDIA) for allowing me the use of their *High Performance Computing* (HPC) cluster through project CEPRA-XIII-2019 "Sistemas Complejos". All the computational simulations of this thesis were carried out in the HPC cluster. Without it, all the numerical calculations would not been achieved in the time and dimensions they did.

I am very grateful to my Professors of the School of Physical Sciences and Nanotechnology of University Yachay Tech for their excellent teachings in the academic field and for their permanent guidance along my undergraduate years that have helped me to reach my goals.

Lastly, I am deeply grateful to family for their unwavering support and belief in me. Especially to my mother, who has always guided me to learn, grow and develop, and who has been a source of encouragement and inspiration for me and my brothers throughout our life.

Resumen

En este trabajo investigamos un proceso adaptativo de recableado de enlaces para explicar la formación de módulos o estructura de comunidades que es comúnmente observado en varias redes naturales sociales y tecnológicas. Nuestro trabajo se basa en un marco previamente propuesto para coevolución de topología y dinámica en redes. Coevolución consiste en la coexistencia de dos procesos dinámicos en una red, cambio de estados y recableo de enlaces entre nodos, que pueden darse con diferentes escalas de tiempo o probabilidades. Para caracterizar el surgimiento de una estructura de comunidad desde una configuración aleatoria inicial, nosotros introducimos una cantidad estadística correspondiente al producto del cambio de modularidad multiplicado por el tamaño promedio normalizado del componente más grande de una red. Esta cantidad es numéricamente calculada en el espacio de los parámetros que representan las acciones de conexión y desconexiónque constituyen el proceso de recableado adaptativo. La estructura de comunidad surge en una red conectada para un rango de valores intermedio para estos parámetros, denotados como fase II. Esta región separa una fase I, donde la red se mantiene aleatoria y conectada, de una fase III, donde la red es fragmentada en componentes pequeños. Nuestros resultados muestran que la existencia de un proceso de recableado adaptativo es suficiente para inducir la formación de comunidades en redes, incluso en la ausencia de dinámicas de nodos. Nosotros hemos establecido una relación entre el surgimiento de comunidades y la densidad de enlaces activos, enlaces que conectan nodos en diferentes estados en la red. Como una contribución principal, hemos encontrado una solución analítica para la evolución de enlaces activos usando un enfoque de campo medio. Además, basado en el marco general, hemos propuesto un nuevo modelo de coevolución, donde un proceso de recableado adaptativo es acoplado a una dinámica de formación de opiniones para nodos. En este modelo, nodos representan agentes sociales que tienen opiniones en una escala discreta y pueden interactuar acorde a una condición umbral, una situación típica en varios sistemas sociales. Nosotros mostramos que nuestro modelo contiene varios modelos previos como casos especiales. Un producto útil de esta tesis es la elaboración de un código computacional para simulaciones de modelos generales para coevolución de topología y dinámicas en redes.

Palabras claves: Redes Complejas, Sistemas Complejos, Sociofísica, Modelo de Formación de Opiniones, Transisiones sin equilibrio.

Abstract

We investigate a process of adaptive rewiring of links as a mechanism to explain the formation of modular or community structure that is commonly observed in many natural, social and technological networks. Our work is based on a previously proposed framework for coevolution of topology and dynamics in networks. Coevolution consists of the coexistence of two dynamical processes on a network, node state change and rewiring of links between nodes, that can take place with different time scales or probabilities. To characterize the emergence of community structure from an initial random configuration, we introduce a statistical quantity corresponding to the product of the modularity change times the average normalized size of the largest network component. This quantity is numerically calculated on the space of parameters representing the actions of connection and disconnection that constitute the adaptive rewiring process. Community structure arises on a connected network for a range of intermediate values of these parameters, denoted as phase II. This region separates a phase I region, where the network remains random and connected, from a phase III region, where the network is fragmented in small components. Our results show that the existence of an adaptive rewiring process is sufficient to induce the formation of communities in networks, even in the absence of node dynamics. We have also established a relation between the emergence of communities and the density of active links, links between nodes in different states in the network. As a main contribution, we have found an analytic solution for the evolution of the density of active links by using a mean field approximation. In addition, based on the general framework, we have proposed a novel coevolution model where an adaptive rewiring process is coupled to a dynamics of opinion formation for the nodes. In this model, nodes represent social agents that have opinions on a discrete-valued scale and they can interact according to a threshold condition, a situation typical in many social systems. We show that our model contains several previous models as special cases. A useful product of this thesis is the elaboration of a computer code for simulations of general models for coevolution of topology and dynamics in networks.

Keywords: Complex Netwoks, Complex Systems, Sociophysics, opinion formation models, non-equilibrium transitions.

Contents

List of Figures

Chapter 1

Introduction

1.1 Complex Networks

A network is a set of elements with relations among them. These elements are called *nodes* or *vertices* and their relations are known as *links* or *edges*. Networks can be used to represent different complex systems . These systems can be natural objects such as the neural network of the nematode worm *Caenorhabditis elegans*[1], food webs[2–4], cellular and metabolic networks[5–8], or they can be artificial or technological structures such as collaborations, co-authorship and citation networks of scientist[9–11], the World-Wide-Web[12], the Internet[13], highway and subway systems[14], power grids[15,16], and telephone call graphs[17]. Due to this generality, different approaches can be employed to study networks and their properties in different scientific areas; for example: in Mathematics graph theory is used to investigate topological properties of regular networks; in Physics, Statistical Mechanics is used to studied critical phenomena and phase transitions in networks[18], and in Social Sciences networks represent the functioning of human societies, focusing on issues of centrality (which individuals are the most connected or have the greatest influence) and connectivity (which individuals are connected through the network).

Physics and other sciences have created in recent years the general concept of complex system to describe a diversity of natural and artificial systems. A *complex system* is a set of interacting elements whose collective behavior cannot be derived from the knowledge of the properties of the isolated elements. The collective behavior is said to emerge from the interactions between the components, without any external influence or design. These systems commonly exhibit two properties: self-organization and emergence. The first one corresponds to the display of organization without the application of an external organizing principle or rule. The second is the manifestation of properties that are not present in the constituents of the system nor can be described by the superposition of their properties. A paradigmatic example of a complex system is the brain. It is well known how a single neuron functions. A single neuron cannot think nor have consciousness by itself, but a network of billions of them forming the brain can give rise to thought, consciousness, and emotions.

Complex systems typically have a heterogeneous structure, a great number of agents capable of interacting with each other at different scales and able to interact with the environment, leading to topological and dynamical evolution of their properties in time. These agents are not necessarily identical; they can be different and interact by different rules. The interactions between agents can be characterized as links forming a complex network, where the topology of the connectivity is not uniform nor trivial. *Complex networks* are a subarea of networks applied to complex systems . The research on natural and technological complex networks has exponentially increased in recent years thanks to the availability of high speed computer power and the access to large databases.

Types of networks

The simplest possible network is one where all the nodes are identical and the links represent the same relation for all linked nodes. However, the connectivity structure of complex networks is more diverse. There can be more that one type of node in the network, and the links can be different to represent more complex relations among the nodes. In this manner, we can have networks where the links have direction, in the sense that a node i has a relation with node j but the node j has no relation with node i. These type of networks are known as directed networks. There can be also relations that are not equal among all nodes but are stronger or weaker depending on the nodes involved. This can be represented by a weighted link, where the strength of the link is quantified. The state of the nodes and their links can also change over time. Sophistication can be added to include these and others variations in networks. Some examples of the different types of networks are shown in Fig.1.1.

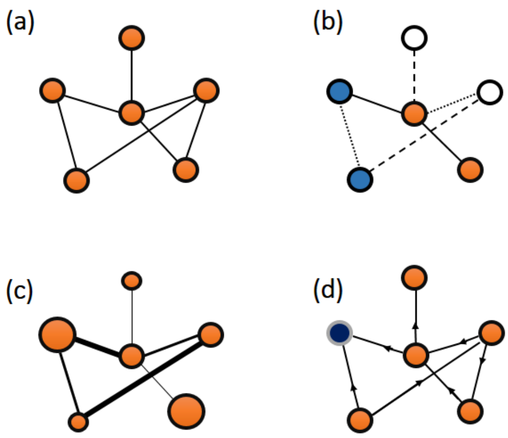


Figure 1.1: Examples of different types of networks: (a) undirected network with only a single type of node and a single type of link; (b) network with a number of discrete nodes and link types; (c) network with varying node and link weights; (d) directed network in which each link has a direction. Figure adapted from Ref. [19]

The variability of complex networks makes it difficult to develop a strict classification. Nonetheless, certain criteria can be used to classify them. The topology (how the nodes are connected) is a criterion that shows a complete spectrum of networks. Ranging from highly regular networks, where all the nodes are connected to the same number of neighbors (e.g. chains, grids, lattices and fully-connected graphs), to the opposite extreme, where the nodes of networks are randomly connected. Erdös and Rényi [20] initiated the systematical study of the properties of random networks. They used probabilistic methods to study the properties as a function of the average number of connections per node. They proposed a method to creating random networks that consisted of creating links between a set of N nodes, prohibiting multiple connections, until having K links. These random networks are denoted by $G^{ER}_{N,K}$. An alternative method to creating random networks, consists of connecting each possible pair of nodes with a probability $0 < p < 1$. These networks are denoted by $G^{ER}_{N,p}$. The structural properties found in $G^{ER}_{N,p}$ networks vary as functions of p. The critical probability $p_c = \frac{1}{N}$, that corresponds to the mean degree (i.e. number of links that a node shares) of the network $\langle k \rangle_c = 1$, shows typical features of a second order transition in the network structure. The statistical properties found in these networks have been rigorously proven [21–23]. In the present thesis, we employ this method to generate random networks.

Real-world networks have been shown to be not random, since they do not possess the same properties as random networks, nor are they regular or uniform. Rather, networks in the real world lie between these two extremes.

1.2 Real-world networks

The investigation of real-world networks has revealed statistical properties common to most of them that lie between those of ordered lattices and completely random networks. This was shown by the works of Watts and Strogatz[24] on small-world networks, and that of Barabási and Albert[25] on scale-free networks, two of the most influential scientific papers of recent times in the area .

1.2.1 Small-world networks

Real-world networks from a variety of contexts are characterized by having, on the average, short paths between any two nodes in the network. This phenomenon is known as the *small-world* effect and is characterized by the quantity L known as the mean geodesic distance between nodes of the network or mean length,

$$L = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geqslant j} d_{i,j}, \tag{1.1}$$

where N is the size of the network and $d_{i,j}$ is the geodesic distance or shortest number of links from node i to node j. A network is called a *small-world* if the quantity L scales logarithmically, or slower, with network size for fixed mean degree. It is often said that a network is *small-world* if the mean length is much smaller than the size of the system, $L \ll N$

The *small-world* phenomenon was already suggested by sociologist Stanley Milgram[26] in his 1960s social experiment. The experiment consisted of sending letters to specific individuals by passing them from person to person. The people involved could only pass the letter to one of their first-name acquaintances. The experiment showed that the documents that arrived to the target individuals took only about six people. This experiment started the idea of the "six degrees of separation", which states that any two people in the world can be connected by just six other people on the average, that would be equivalent to have $L = 6$. This idea has been rigorously investigated in several social networks, such as in the *MSN Microsoft*[27] that involved about 240 million people and 30.000 million conversations, showing an $L = 6.6$; and in the social network *Facebook*[28] that contains 721 million users with around 69 million connections and shows a mean length $L = 4.74$. Many different, unrelated networks actually show the *small-world* effect, such as power grids, the neural network of the worm *C. Elegans*[24], but especially social interaction and collaboration networks. These networks have been widely studied in the literature; for example: collaboration network of film actors[24,29–31], where actors are linked if they have appear in a film together; networks of company executives[32–34], where they are linked if they appear in the same board of directors; networks of coauthorship among scientists[35–40], where scientists are linked if they have coauthored at least one paper; sport leagues, where players are connected if they have played on the same team[41]; coappearance networks[41–43], linked if the individuals are mentioned in the same context; and networks of sexual contacts[44] (Fig.1.2).

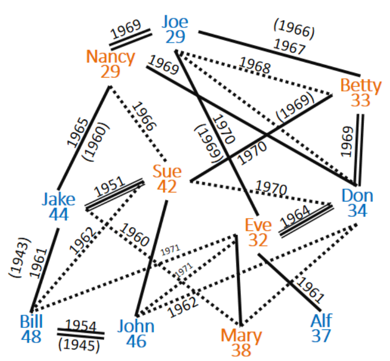


Figure 1.2: Representation of a network of sexual contacts. Population size $N = 2810$ (ages 18 – 74); mean length $L = 6$; Figure adapted from Ref.[45]

A mechanism for the formation of *small-world* networks was proposed by Watts and Strogatz[24]. In this method, we start with N nodes arranged in a circle, where each node is connected to k neighbors. In a clockwise sense, and with a probability $0 < p < 1$, we chose a node and a link to one of its neighbors and we move this link to another node in the ring chosen at random. Duplication of links is avoided. This process is repeated until all nodes have been visited. This method allows one to construct networks ranging from highly regulars networks ($p = 0$) up to completely random ones ($p = 1$).

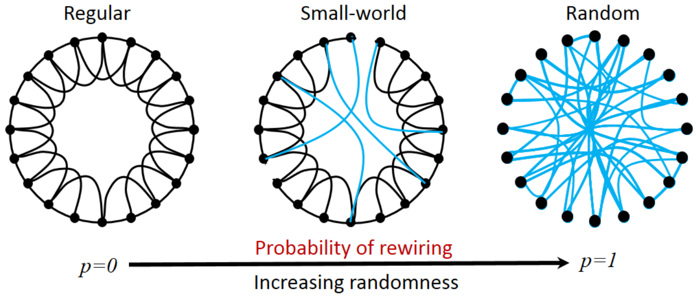


Figure 1.3: Networks obtained by the method of Watts and Strogatz[24] with parameters $N = 20$ and $k = 4$. This method creates networks ranging from highly regular for $p = 0$ and random ones for $p = 1$. The rewired links are in light blue. Small-world networks form for intermediate values of the probability p. Figure adapted from Ref.[24].

Small-world networks also have the property of high clustering coefficient (defined as the fraction of the actual links present in the network compared to all-to-all coupling) like regular networks and at the same time they have characteristic short path lengths between any two nodes like random graphs. The structural properties of the resulting networks can be characterized by the mean geodesic distance, that is a global property, and the clustering coefficient, that is a local property, as a function of the rewiring probability p. This is shown in Fig. 1.4. We can see that the mean geodesic distance $L(p)$ drops quickly as the probability p increases, corresponding to the *small-world* property. Nonetheless, the clustering coefficient $C(p)$ remains near the value 1 as $L(p)$ decreases, which indicates that the transition to the *small-world* regime is pretty much imperceptible at a local level.

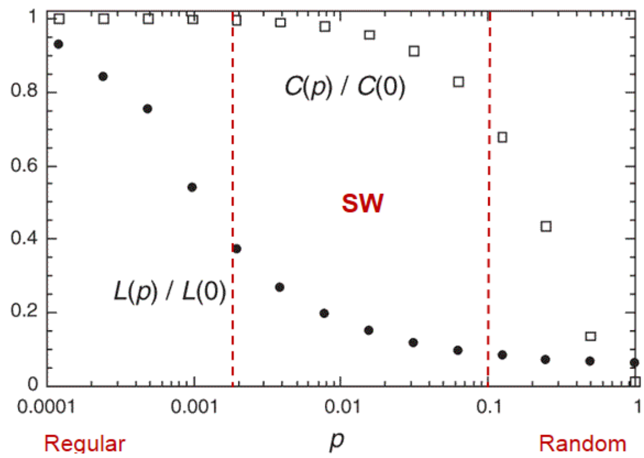


Figure 1.4: Normalized mean geodesic distance and clustering coefficient as functions of the rewiring probability p. The values $L(0)$ and $C(0)$ correspond to the values of these quantities for a regular lattice, respectively. The values obtained are the average over 25 realizations of initial random conditions for the network. Label SW indicates the region where a small-world network appears. network size $N = 100$, with mean degree $\langle k \rangle = 4$. Figure adapted from Ref.[24].

Note that a few random, long-range connections suffice to produce the small-world effect. *Small-world* networks have great interest because they facilitate the transmission of information. For example, they are relevant in the spreading of infectious diseases, with implications for vaccination strategies and the evolution of virulence[46–48].

1.2.2 Scale free networks

It has also been found that many real-world networks possess nodes with different number of links, with a preference of certain nodes over others. This is manifested as a non-uniform distribution of links per node. Very often in real-world networks, many nodes have a few links and few nodes display many links. This can be measured by the degree probability distribution $P(k)$. These distributions are characterized as power laws (scale free) distributions, $P(k) \sim k^{-\gamma}$, with γ between 2 and 3. Thus, the distribution $P(k)$ shows a long tail with values far from the mean value. This drastically differs from the degree distributions of random networks, that are binomials or Poissonian distributions, as seen in Fig.1.5.

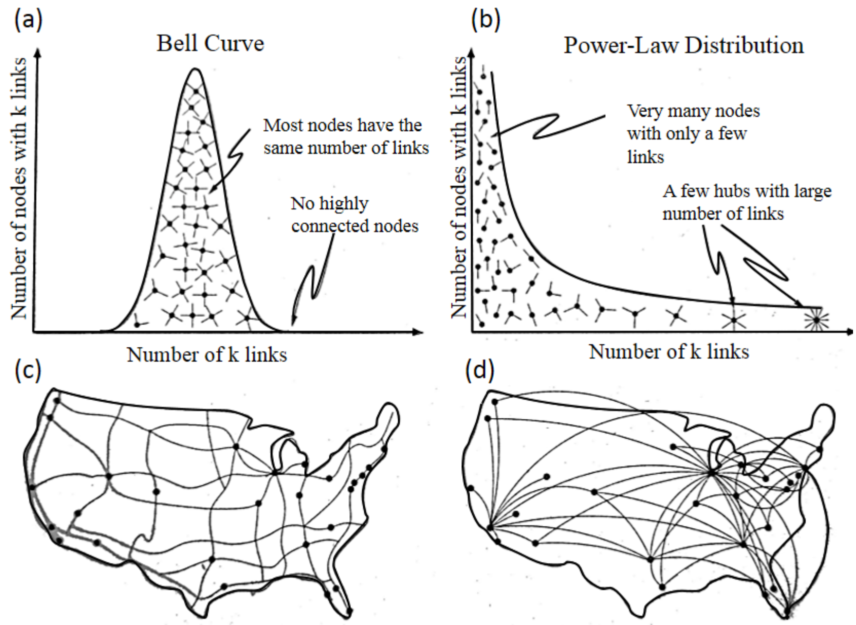


Figure 1.5: Examples of *random* and *scale-free* networks distributions. (a) Degree distribution for a random network. (b) Degree distribution for a scale-free network. (c) Network of highways in the USA shows a Poissonian degree distribution. (d) Network of airports in the USA has a degree distribution that follows a power law.

These networks are known as *scale free networks*, and have been widely studied in the literature[18,25,49–51]. Some examples are shown in Fig. 1.6, where the cumulative degree distribution versus the degree is plotted for several networks. Panels (a), (b), and (c) show power-law degree distributions, as indicated by their straight-line on the doubly logarithmic scales, corresponding to scale-free networks. Panel (d) has a power-law tail, but deviates from a power-law behavior for small degree value. Panel (e) appears to have a truncated power-law degree distribution, or possibly two separated power-law regimes with different powers. Barabási and Albert[25] proposed a mechanism of preferential attachment to explain the emergence of scale-free networks. These networks are formed as the result of the evolution of a network where new nodes entering tend to connect preferentially to the already most connected nodes in the network ("the rich get richer").

These networks have been shown numerically[12,52] and analytically[53,54] to be less susceptible to random error link failures, but more susceptible to targeted attacks. In fact, few nodes dominate the connectivity of the network, forming hubs that the majority of the nodes are attached to. Then, any random failure would most probably affect the nodes with low degree, with negligible effect over the network. In contrast, targeted attacks over the hub nodes could break the network completely. These effects have been studied in relation to the resilience of systems such as the Internet[55], the design of therapeutic drugs[8], and metabolic networks[8,56].

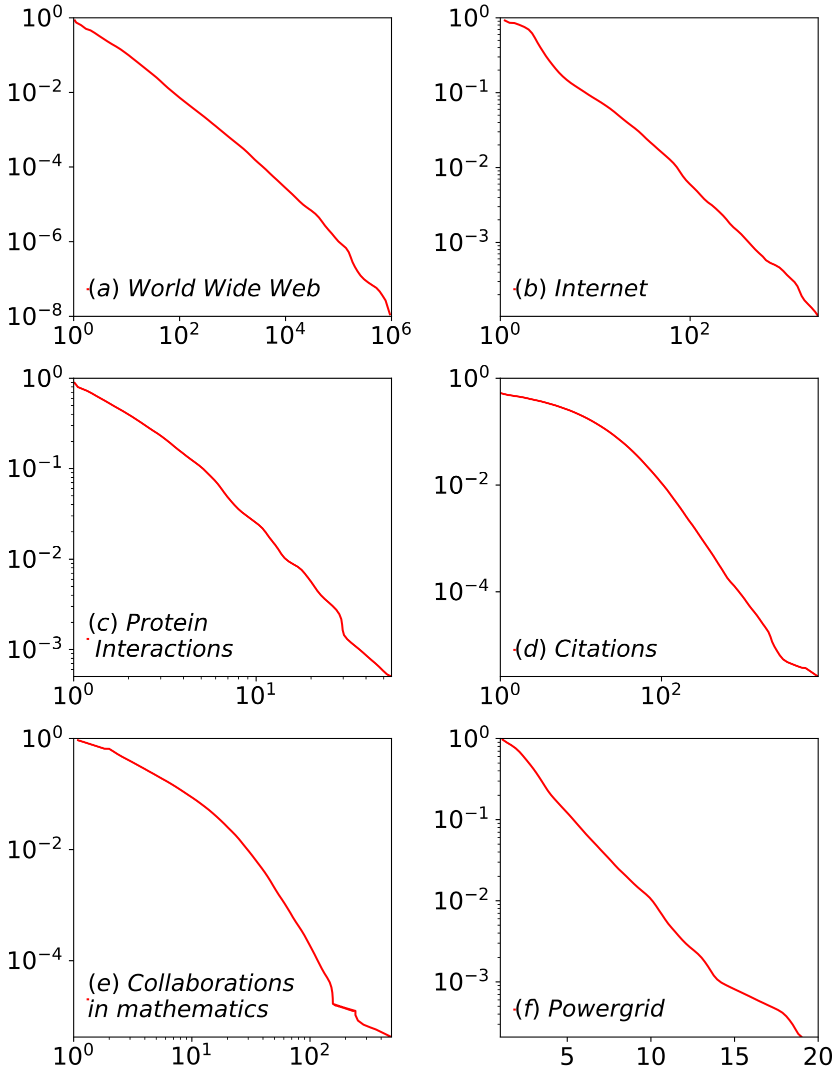


Figure 1.6: Cumulative degree distributions for six different networks. The horizontal axis for each panel is the node degree k and the vertical axis is the cumulative probability distribution of degrees. The networks are: (a) a 300 million nodes subset of the World Wide Web, *circa* 1999 from Ref.[12]; (b) the Internet at the level of autonomous systems, April 1999 from Ref.[57]; (c) the interaction network of proteins in the metabolism of the yeast *S. Cerevisiae*[58];(d) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information[11]; (e) the collaboration network of mathematicians[39]; (f) the power grid of the western United States[24]; Figure adapted from Ref.[19]

1.3 Community structure

Another common topological characteristic of real-world networks is that they show modules and organization in their structures. These modules are groups of nodes that have a high number of links within the groups and low number of links outside them. This feature is known as *community or modular structure* and the groups as *communities*. It has been shown that most social networks have a modular structure[59,60]. Detecting communities and their boundaries allows us to classify nodes accordingly to their roles within the network, instead of just the statistical properties of the network.

Communities are groups or nodes that have similar properties or functions within the network. In this manner, we can expect that, for example, communities in the Worldwide Web might correspond to sets of web pages on related topics[61]; communities in social networks might correspond to social units[62]; communities in a citation network might correspond to related papers on a single topic[11]; communities in a metabolic network can suggest that groups of nodes can perform different functions with some degree of independence[63,64]. The existence of communities in the *Human Disease Network* (Fig.1.7) suggests that certain health conditions have the same origin. The study of the *community structure* can give us insight into the dynamics and collective functioning of complex networks.

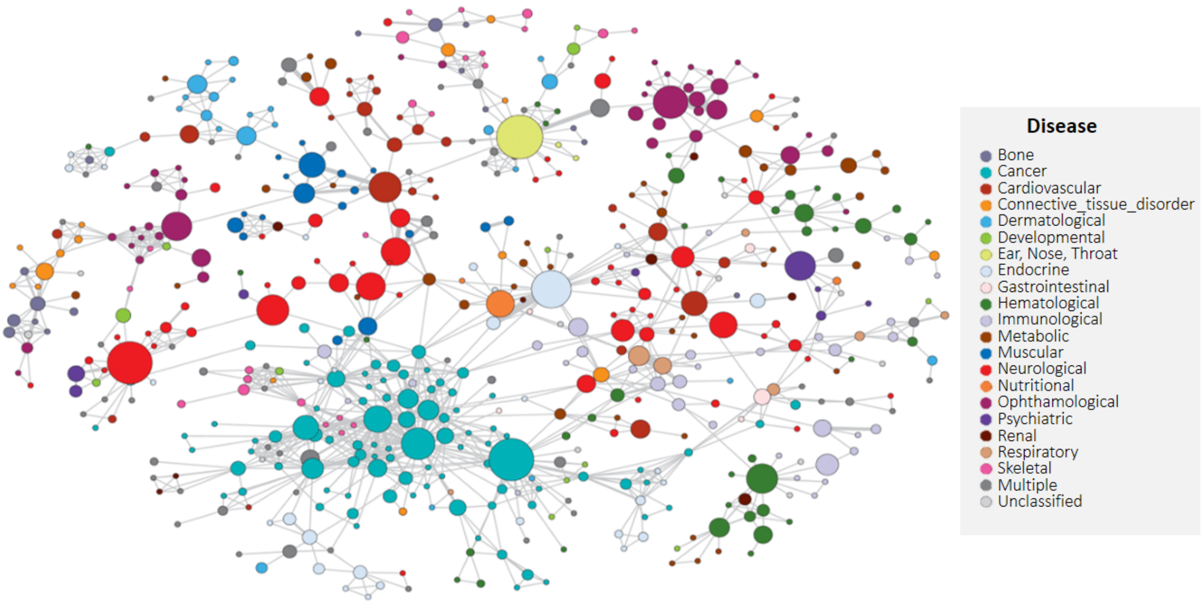


Figure 1.7: The human disease network. It consists of 944 diseases with at least one link to other diseases and 576 diseases form the giant component. This network is *small-world*, *scale-free*, and possesses *community* structure. Figure adapted from Ref.[65]

Two important remarks need to be made about community detection. First, the aim of community detection is to identify the modules in networks mainly based on their topology. Second, even though the idea of community is intuitive, there is not a unique definition for it. The latter remark has lead to the creation of various definitions for the concept of community, and thus different methods to detect them. These definitions can be classified into three main categories[66]: (i) *local definitions* that try to identify communities based on the nodes that are part of subdivisions of a network and its immediate neighbors, disregarding their relation to the rest of the network; (ii) *global definitions* that consider communities as building blocks of the network that can be identified in relation to the other communities of the network; (iii) *definitions based on node similarity* that focus on how similar nodes are to each other according to certain local or global criteria to establish communities. In this thesis, we shall use a *global definition* of communities for their detection according to the Louvain method[67] that will be explain in more detail in the following Sections. This type of method compares a null model with partitions of the original network according to certain criteria to evaluate how good the partitions are.

1.3.1 Importance of communities in real social networks

Empirical knowledge shows that people tend to be divided into groups along their interests, age, social status, occupation, opinions, political views, etc. It has been found that there is some correspondence between these groups and communities in networks[67,68]. A good example can be seen in Fig. 1.8. This figure represents communities found in a mobile phone network from Belgium. Belgium has two main coexisting linguistic communities, so it would be expected that the communities detected in this network should have a high level of linguistic homogeneity. The network contains about 2 million customers along with some demographic information such as sex, age, language, and the postcode of the place where they live. This network was analyzed in Ref.[67] by applying their own community detection method known as the Louvain algorithm. The community detection algorithm shows two clusters with communities that are highly linguistic homogeneous. It was found that in the 36 communities with more than 10.000 individual, more that 85% of their member speak the same language.

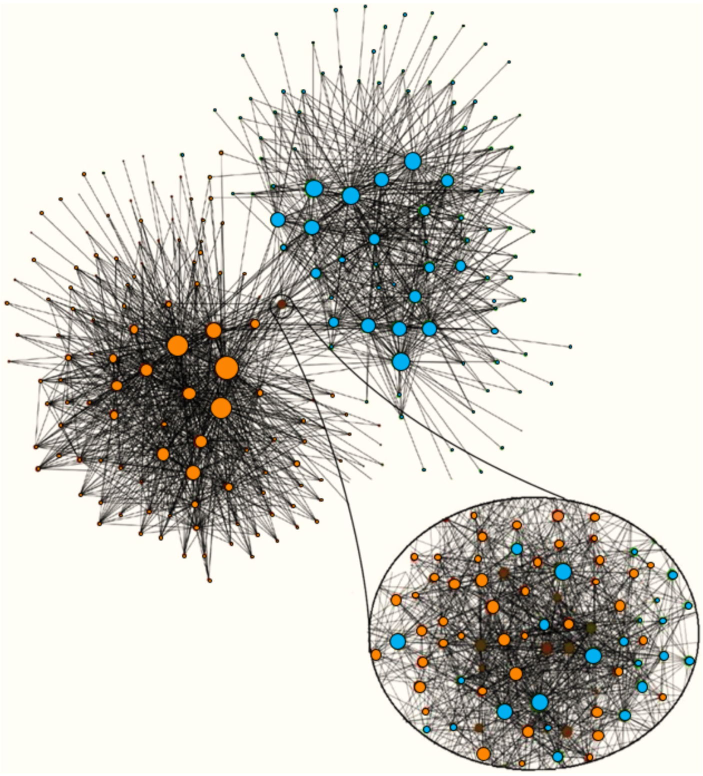


Figure 1.8: Communities in a Belgian mobile phone network. This network contains about 2 million customers. The size of the nodes represents the number of individuals in a community, and its colour represents the main language spoken in each community; being light blue for Dutch and orange for French. Two clusters emerge, each with homogeneous language. Between them there is an intermediate community with less apparent language separation that can be seen in higher resolution in the zoomed section. Figure adapted from Ref.[67].

Similarly, it is expected that detection of communities in other real social networks are related to certain subdivisions based on common features of the individuals that are part of them. Figure. 1.9 shows some examples of real social networks and how these subdivisions have been proposed. For instance, the Zachary's karate club network (Fig.1.9(a)) and the Southern Women Event Participation data set (Fig.1.9(b)) where the communities are expected to be related to friendship and/or acquaintance among members. Other examples are the collaboration network between scientist working at the Santa Fe Institute (Fig.1.9(c)), and the citation patterns in the *Small World* literature (Fig.1.9(d)) the detected communities are related to the research area of their participants. For other real social networks it may be unknown or difficult to find the trait that leads to the formation and detection of communities. For example, the communities found in Lusseau's network for bottle-nose dolphins in Fig.1.9(e) do not correspond to known demographic information of its individuals.

Elucidating the *community structure* in social networks relevant for the demographic identification of network components or modules, as well as for understanding the propagation of opinions and diseases[69].

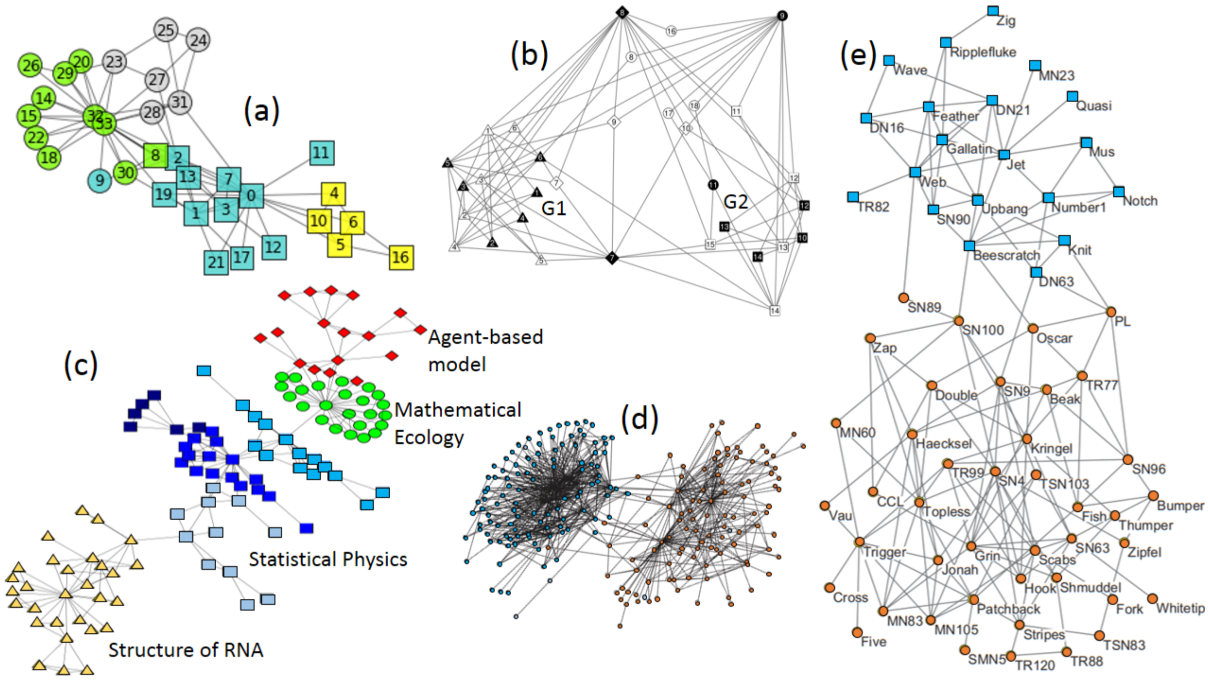


Figure 1.9: Community structure in social networks. (a) Zachary's karate club network, adapted from Ref.[70]. (b) Southern Women Event Participation data set, adapted from Ref.[71]. (c) Collaboration network between scientists working at the Santa Fe Institute, adapted from Ref.[62]. (d) Citation patterns in the small-world research literature, adapted from Ref.[72]. (e) Lusseau's network of bottle-nose dolphins, adapted from Ref.[73]

1.3.2 Detection of communities in networks

Several methods for the detection of communities in networks have been proposed in the literature. They can be classified according to how they work, as follows.

Graph partitioning (Fig.1.10(a)) consists of dividing the nodes of a network into g groups of predefined size such that the links between the groups are minimal. It is necessary to specify beforehand the number of clusters for this method. This may be a problem since it is unusual to know the number of groups or *communities* a priori. Then it is necessary to make some reasonable assumption about the number and sizes of the *communities*.

Hierarchical clustering (Fig.1.10(b)) is used when networks show hierarchical structures; for example the networks can display different levels of grouping such that small groups are included into larger clusters, and these clusters can be included into even larger ones, and so on. For these methods, it is necessary to define a similarity measurement between nodes, such as degree mean geodesic distance between nodes, etc. Then, similarity is measured for all pairs of nodes, no matter if they are connected or not. Finally, groups are made by one of two iterative methods: (1) clusters are merged if they have sufficiently high similarity, or (2) clusters are divided by removing links between nodes with low similarity. The first one is a bottom-up approach and the second one is a top-down approach. These methods create many partitions, so it is necessary to establish certain criteria or condition to select one of the partitions. *Hierarchical clustering* does not require the number nor the sizes of the clusters, but it does not have an intrinsic criterion to discriminate between the partitions created.

Partitional clustering (Fig.1.10(c)) is another method in which the number of clusters must be given at the beginning. This method considers a metric space. Here, each node is a point and distances are defined between points in function of a dissimilarity measurement for nodes. This method divides the point in g clusters such that a cost function based on the distances between points is minimized or maximized. This cost function can be the mean geodesic distance in the clusters, the diameter of the cluster (largest distance between two points), or the largest distance to a centroid in the cluster, among others. This method has the same problem as the *graph partition* method; since it is necessary to know the number of clusters beforehand. Also, this method can be difficult to implement since the selection of a metric space can be natural for some networks, but not for others.

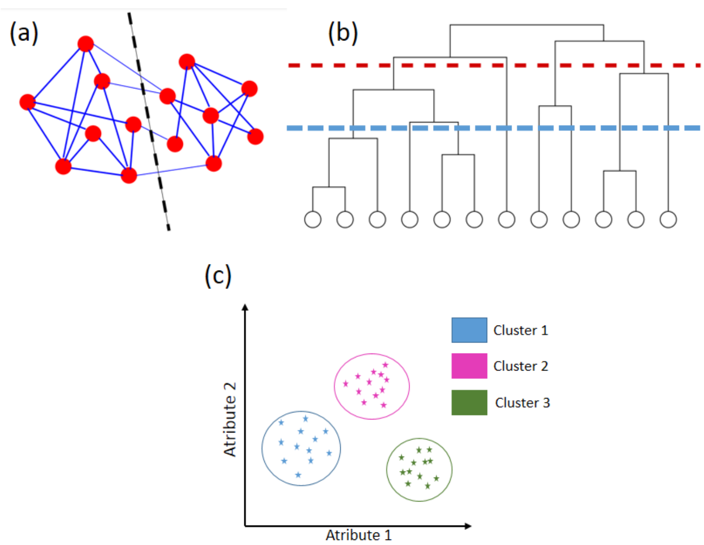


Figure 1.10: Examples of methods for the detection of communities in networks. (a) *Graph partitioning* method on a network with two groups of equal size; the dash lines divides the two communities found. (b) Dendrogram of the *hierarchical clustering* method. Each dashed line denotes a partition of communities found. (c) *Partitional clustering* method on a network with three communities; each point represents a node and each community is established by the optimization of a cost function of their distance.

Divisive algorithms can be used to detect communities based on the elimination of links that connect those communities. For this, it is necessary to establish a property or set or properties that allow their identification. These algorithms are similar to the top-down ones for *hierarchical clustering*, but here the links removed are inter-cluster links and not the links between nodes with low similarity. The inter-cluster links and the links between nodes with low similarity are not necessarily the same.

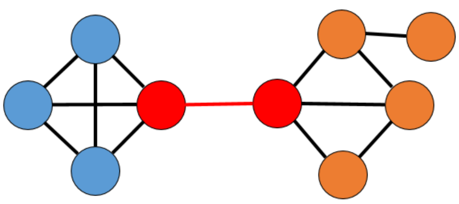
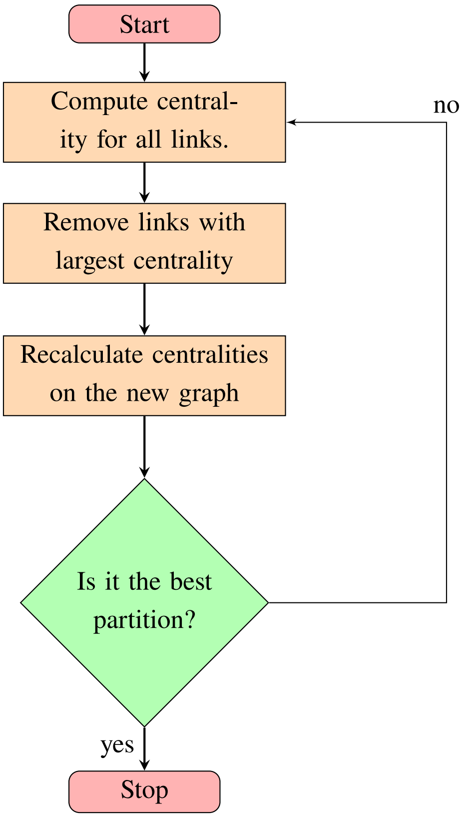


Figure 1.11: Example of link with high edge-betweenness centrality value. In a network with eight nodes and twelve links, the red link that connects the two red nodes shows the higher value for edge-betweenness centrality among all nodes. Figure adapted from Ref. 74.

One of the most popular algorithms was introduced by Girvan and Newman[62,75]. This method selects links according to their values of *edge-betweenness centrality* (Fig.1.11), which is defined as the number of shortest paths that go though a link. The removal of these links will result in the partition of the network into densely connected sub-networks. The Girvan and Newman method consists of the iteration of these steps:



This method gives partitions of the original network, but similarly to *hierarchical clustering*, it does not give a criterion for the best partition. For this, it is necessary to introduce some quantity that can evaluate how good a partition is. Newman and Girvan[75] introduced a quantity that evaluated the quality of a network partition. This quantity is known as *modularity* and it compares the number of links inside the modules of a partition against the expected number of links of random networks with the same size and degree, also known as a *null model*.

By choosing the modularity as a quality function, the community detection problem becomes an optimization problem. The number of possible partitions in a network increases at least exponentially with its size, complicating the problem of modularity optimization as the network size increases. Another problem is that the modularity value of a partition does not have a meaning by itself. It is necessary to compare it with the modularity expected for a random graph of the same size[76], since this one can have high modularity values due to fluctuations[77].

There are algorithms focused on modularity optimization that find good approximations to the real solution, since it is not feasible to measure the modularity for all the possible partitions of a network due to the scaling of the partitions with the size of the network. One of such algorithms was introduced by Blonde *et al.*[67], and it is the one used in the this thesis. In our case, the specific value of the modularity is not important, since we introduce the concept of change of modularity.

1.4 Research problem

As we have seen, the origin of small-world networks was explained by Watts and Strogatz[24] through a process of random rewiring of links that generates long-range connections. A small fraction of long-range interactions is enough to induce the small-world phenomenon. Similarly, a mechanism of preferential attachment was proposed by Barabási and Albert[25] to explain the emergence of scale-free networks, whose probability distribution of links follows a power law. However, although some scenarios have been explored, understanding the mechanisms that lead to the formation of communities in networks remains a problem of much interest[69].

We shall see in Chapter 2 that many natural, social, and technological networks are not static; it often occurs that both the connections between the nodes and their state variables, affect each other and evolve in time. These systems exhibit coevolution of their topology and dynamics.

Our research is motivated by the following question: how does a community structure arise in networks? In this thesis we shall investigate the problem of the formation of communities in the context of coevolutionary dynamics in networks. We also address the question of the influence of the node dynamics on the emergence of community structure in networks.

The research field on complex systems is very recent in Ecuador. In this context, this thesis represents a first contribution by a student to new interdisciplinary research lines at Yachay Tech University, mainly Network Science and Sociophysics. The concepts and techniques developed here can be applied to the analysis of networks in social groups and communities in Ecuador. We have elaborated a computational code for characterizing statistical and topological properties of general networks. This code can be employed by other researchers in our country that are interested in the study of complex systems and computational social science

1.5 General and Specific Objectives

General objective of this thesis

Our main objective is to investigate a process of adaptive rewiring of links, in the context of coevolution dynamics, as a mechanism to explain the emergence of community structure in networks.

Specific objectives of this thesis

1. To investigate the relationship between the formation of communities and the evolution of the number of active links in the network, i.e., links that connect two nodes in different states.
2. By using a mean field approximation, investigate an analytic approach for the evolution of the density of active links in a network.
3. Characterization of the collective properties of a network as a function of the parameters that describe the process of adaptive rewiring; in particular identify rewiring parameter values for which a community structure emerges.
4. To investigate an opinion formation problem in coevolutionary networks, where interacting social agents have opinions on a discrete-valued scale. We address the following questions: Under what conditions do their opinions converge? Can communities emerge in such networks?

5. To develop an efficient computer code for simulations of a general model of coevolution of topology and dynamics in networks.

1.6 Outline of this thesis

This thesis is organized into five chapters. Chapters 1 and 2 correspond to the introductory part, where the main concepts and theoretical basis are reviewed. Chapters 3 and 4 contain our contributions and results; they include our research of an adaptive rewiring process as a mechanism for the formation of community structure in networks and the investigation of an opinion formation model with coevolution dynamics on networks. We end with chapter 5, where conclusions and suggestions for future work are stated.

- Chapter 1: **Introduction**. This chapter presents a short review of the wide topic of complex networks, focusing on the main common properties that have been found in real-world networks: small-world effect, scale-free distribution, and community structure. Our research problem and objectives of this thesis are also presented in this chapter.

- Chapter 2: **Theoretical Framework and Methodology**. This chapter reviews the concept of coevolutionary dynamics in networks. It presents the theoretical framework on which we base our research. Here we describe concepts and statistical quantities that we use to characterize the collective properties of the systems that we investigate in the next chapters. In particular, the tools employed to measure the modularity of networks is discussed.

- Chapter 3: **Adaptive rewiring and the emergence of community structure in networks**. Here we propose and investigate an adaptive rewiring process as a mechanism for the emergence of community structure in networks. We relate the decrease in the number of active links, that connect nodes in different states, to the formation of modular structures. Here we present an analytic mean field approach to the evolution of the density of active links in adaptive rewiring. This chapter contains main results of our research.

- Chapter 4: **A coevolution model with node dynamics**. In this chapter we apply the general framework for coevolution of topology and dynamics to a model of opinion formation. We propose and investigate a model of social dynamics where agents have opinions on a discrete-valued scale and their connections change according to the adaptive rewiring process studied in Chapter 3. We show that several previously proposed models in the literature can be derived as special cases of this model.

- Chapter 5: **Conclusions and Outlook**. In this chapter, we summarize and further discuss the results of this thesis. We also provide an outlook of future research that can be motivated from this work.

Chapter 2

Theoretical Framework and Methodology

2.1 Framework for coevolution of topology and dynamics in networks

Many networks observed in nature and society are not static. Actually, many natural and technological complex systems can be represented as dynamical networks of interacting elements, or nodes, where the connections between the elements and their state variables evolve simultaneously[55,78–81]. The links describing the interactions between nodes can vary their strengths or appear and disappear as the system evolves on diverse timescales. In multiple cases, these variations in the topology of the network arise from a feedback effect of the dynamics of the states of the nodes: the network changes in response to the evolution of those states which in turn determines the modification of the network. Systems that possess this coupling between the topology and states have been denoted as coevolutionary dynamical systems or adaptive networks[78–80]. The collective behaviors emerging in coevolutionary systems depend on the competition between the time scales of these two coexisting processes: the dynamics of states of the nodes and the dynamics of the network connections. Coevolution has been investigated in the context of spatiotemporal dynamical systems, such as neural networks[82,83], coupled map lattices[84,85], motile elements[86], game theory[78,79,87], spin dynamics[88], epidemic propagation[89–92], and models of social dynamics[93–98].

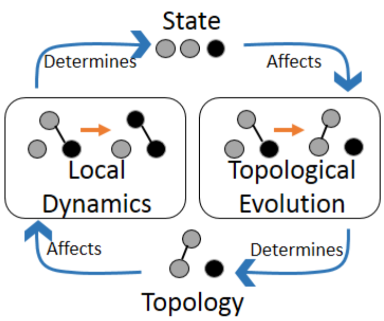


Figure 2.1: Schematic representation of a coevolutionary or adaptive dynamical system. On the one hand, the interaction dynamics of the nodes changes their state variables and affects the connections between nodes (network topology). On the other hand, modifications of the connections in the network influence the states of the nodes. Figure adapted from Ref.[80]

A general framework for the description of coevolution in dynamical systems was proposed by Herrera *et al.* in 2011[99]. In this model, it is assumed that the process by which a node changes its state and the process by which a node change its neighbors, called *adaptive rewiring*, have their own dynamics. Furthermore, each process can occur with its own time scale or probability, independently from the other.

A coevolutionary system can be described by the coexistence of a rewiring process that takes place with a probability P_r, and a process of node state change that occurs with a probability P_c. Then, the premises of the model are:

1. A specific coevolution model is given by a functional relation $f(P_r, P_c) = 0$.
2. A rewiring process consists of two actions: disconnection and connection between nodes.

In this way, the general dynamics of the system can be characterized on the space of the parameters (P_r, P_c). Then, a coevolution model consists of a functional relationship $f(P_r, P_c) = 0$ between the probabilities P_r and P_c that corresponds to a curve on this space. The function $f(P_r, P_c) = 0$ expresses the coupling between the processes of node state change and rewiring of links between nodes. Different models of coevolution can be represented by different functional relations between the probabilities P_r and P_c.

Any rewiring process in a network can be seen as consisting of two basic actions: connection (“attraction") and disconnection (“repulsion") between nodes. It is assumed that either action, disconnection or connection , takes place by some mechanism of comparison of the states of the nodes. Both connecting and disconnecting interactions between nodes, based on some comparison of their states, are often found in social relations, biological systems, and economic dynamics[80,81,93,98]. These actions may represent discrete connection-disconnection events, or to continuous increase-decrease strength of the links, as in weighted networks.

The disconnection action can be characterized by a parameter $d \in [0, 1]$, that measures the probability that two nodes in identical states become disconnected, and such that $1 - d$ is the probability that two nodes in different states disconnect from each other. Similarly, the connection action can be characterized by another parameter $r \in [0, 1]$ that expresses the probability that two nodes in identical states become connected, and such that $1 - r$ is the probability that two nodes in different states connect to each other[99]. Then, any rewiring process subject to disconnection-reconnection actions between nodes can be characterized by a pair of values d and r on the space of parameters (d, r). This rewiring process guarantees the conservation of the total number of links in the network.

In the context of social dynamics, the phenomena of homophily (the tendency to interact between nodes of similar states) and heterophily (the tendency to interact between nodes of different states) can be naturally described on the plane (d, r). Homophily is maximum (and heterophily is minimum) for the values $(d = 0, r = 1)$, while the maximum of heterophily (and the minimum of homophily) corresponds to $(d = 1, r = 0)$. Diverse coexisting degrees of these two phenomena can be characterized as the parameters d and r are varied on the (d, r) plane.

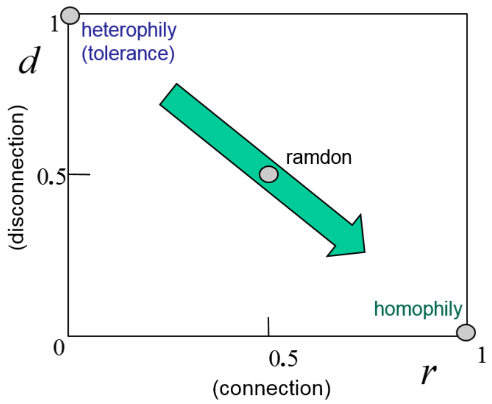
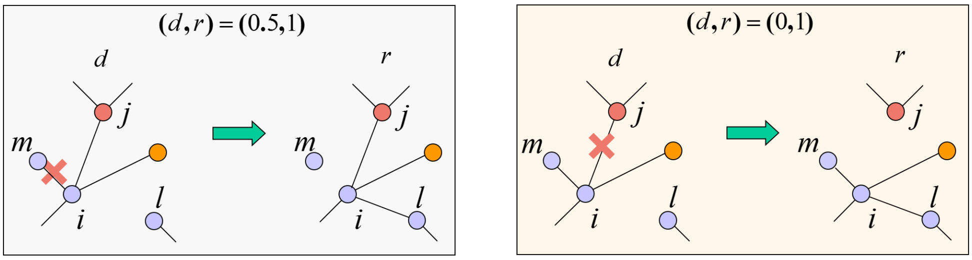


Figure 2.2: The space of parameters (d, r) displays the possible adaptive rewiring processes that can take place on a network in the model of Herrera *et al.*[99]. The point $(d, r) = (0.5, 0.5)$ corresponds to a totally random rewiring, regardless of the states of the nodes. The points $(d, r) = (1, 0)$ and $(d, r) = (0, 1)$ represent the limits of total heterophily (affinity to different ones, tolerance for diversity) and total homophily (affinity for equals, intolerance for diversity), respectively. The arrow signals the transition between these two limits along the diagonal $d = 1 - r$.

Within this general framework, many coevolution models that have appeared in the literature can be characterized as special cases. For example, a functional relation $P_c = 1 - P_r$ and a rewiring process $(d = 0.5, r = 1)$ corresponds to the model in Ref. [94]; the same coupling relation and a rewiring process $(d = 0, r = 1)$ was used in Ref. [95], while the rewirings employed in Refs. [96–98] can be regarded as type $(d = 0, r = 0.5)$. Note that only the rewiring process $(d = 0.5, r = 0.5)$ is completely independent of the states of the nodes.



(a) (b)

Figure 2.3: Examples of adaptive rewiring processes. (a) Rewiring $(d = 0.5, r = 1)$: node i is disconnected from neighbor m chosen at random, and then connected to a node l that shares the same state of i. (b) Rewiring $(d = 0, r = 1)$: node i is disconnected from neighbor j in a different state, and then connected to a node l that shares the same state of i.

As an application of the general framework, Ref. [99] considers a random network of N nodes having mean degree $\langle k \rangle$. Let ν_i be the set of neighbors of node i, possessing k_i elements. The state variable of node i is denoted by g_i. For simplicity, assume that the node state variable is discrete, that is, g_i can take any of G possible options. The states g_i are initially assigned at random with a uniform distribution. Therefore there are, on the average, N/G nodes in each state in the initial random network. Consider a system characterized by some node dynamics with probability P_c and subject to a rewiring process with probability P_r, which may be coupled by a functional relation $f(P_c, P_r) = 0$. Assume that the actions of the rewiring process are described by parameters (d, r).

Then, the coevolution dynamics in this system is given by iterating these three steps [99]:

1. Choose at random a node i such that $k_i > 0$.

2. With probability P_r, apply the rewiring process (d, r): select at random a neighbor $j \in \nu_i$ and a node $l \notin \nu_i$. If the edge (i, j) can be disconnected according to the rule of the disconnection action and the nodes i and l can be connected according to the rule of the connection action, break the edge (i, j) and create the edge (i, l).

3. With probability P_c, apply the node dynamics.

Step 2 describes the rewiring process that allows the acquisition of new connections, while step 3 specifies the process of node state change. It has been verified that the collective behavior of this system is statistically invariant if steps 2 and 3 are interchanged [100].

Many different node state dynamics can be considered in step 3 within this framework for coevolutionary systems. Some examples in the area of social dynamics are: dynamics of opinion formation such as bounded confidence models [101], imitation rules such as voter models [94,102–105], and cultural influence [106].

Note that if $P_r = 0$, we have a static network where only the interaction dynamics among nodes can change their state. Alternatively, if $P_c = 0$, the states of the nodes are kept fixed; only the connections between nodes can change. We shall investigate the case $P_c = 0$ in Chapter 3. As we will show, the presence of an adaptive rewiring process alone is sufficient to induce the formation of community structure in networks. In Chapter 4 we propose and investigate a coevolution model with both coexisting processes, an opinion formation node dynamics and adaptive rewiring, $P_c \neq 0, P_r \neq 0$.

2.2 Characterization of community structure in networks

In order to characterize the presence of community structure in networks we shall use two statistical quantities. The first quantity corresponds to the measure of modularity, based on partitions of the network into non-overlapping communities. The second quantity is the normalized size of the largest component in the network. A similar quantity also used was the normalized size of the largest domain in the network. We call a domain a set of connected nodes that share the same state.

2.2.1 Modularity

Since the definitions of communities may vary; several methods based on modularity optimization have appeared in the literature that take into account different aspects of communities or modular structures[66]. In the present work, we shall employ the well-known algorithm proposed by the Louvain group[67], which displays higher speed and precision than some other methods. This method partitions a network into non-overlapping communities. Then these communities are joined in order to optimize the value of the following parameter introduced by Newman[75]

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i c_j) \tag{2.1}$$

where c_i corresponds to the community that node i was assigned, A_{ij} is the ij-element in the adjacency matrix of the network, that is 1 if nodes i and j are connected, or 0 otherwise. The total number of links is $m = \frac{1}{2}\sum_{ij} A_{ij}$ and k_i is the degree of the edge i.

This parameter compares the value of the fraction of links that falls inside the communities against the expected same value if the links were assigned randomly, considering the sizes of the communities and the degrees of the nodes. The quantity Q can take values in the interval $[-1, 1]$.

2.2.2 Largest component and largest domain

A subgraph or component of a network is a set of connected nodes, regardless of the states of their state variables. We denote by S_m the the average normalized (divided by N) size of the largest component or connected subgraph in the system. This quantity is bounded to the range $[0, 1]$ and provides a measure of the integrity of the network. A value $S_m = 1$ corresponds to a large connected network, while a value $S_m \approx 0$ indicates a fragmented network formed by several separated small components or subgraphs.

In contrast, a domain is defined as a set of connected nodes on a network that share the same state variable. There can be many different domains in a network. A homogeneous collective state, also called an ordered state, occurs when all the elements in a system are in the same state. A disordered or inhomogeneous state takes place when several domains coexist in a system. We characterize the homogeneous or ordered collective state in a system by considering the normalized size (divided by the size of the system) of the largest domain present in the network, denoted by S_g.

In practice, we use an recursive algorithm that searches for the states of the neighbors of a given node, and then does the same search for each one of the neighbors that have the same state, keeping track of the visited nodes. At the end, the algorithm returns the number of nodes that share the same state as the first one. Then we repeat the search again on a node that has not been visited, until all the nodes in the network have been searched. Then, we obtain a set of numbers that correspond to the sizes of the different domains in the network, from which we select the largest one. This value is divided by the size of the network to normalize it, which yields S_g. Then, the limit $S_g \approx 1$ corresponds to one large domain and it characterizes an ordered state. The limit $S_g \approx 0$ is associated with the presence of several domains of small size, and it characterizes a disordered system.

For node dynamics based on imitation of states, such as the voter model, the largest subgraph may coincide with the largest domain in a network, i. e., $S_m = S_g$.

Chapter 3

Adaptive rewiring and the emergence of community structure in networks

3.1 How do communities arise in networks?

As we have seen in the Introduction, many natural and technological systems exhibit network structures characteristic of small worlds, scale free, or communities. In some cases, the characteristic properties of these three complex networks may coexist in a system. Watts and Strogatz[24] proposed a mechanism to explain the origin of small world networks, based on a process of random rewiring of links that creates long-range connections. Even a small fraction of long-range interactions produce a decrease in the mean length of a network while keeping a high clustering coefficient , thus leading to the small-world phenomenon. A mechanism of preferential attachment was provided by Barabási and Albert[25] to explain the emergence of scale-free networks. Preferential attachment is a dynamical process by which new nodes entering a network have greater probability to connect to nodes already having more connections than to connect with nodes possessing few connections. This process creates a probability distribution of links that displays a power law. However, although some scenarios have been proposed, investigating the mechanisms for the formation of communities in networks remains an open problem[107].

In any case, the emergence of characteristic topological structures in a network requires some rewiring process that modifies the properties of the links between nodes. Links can appear and disappear, or their strength can change, as a consequence of a rewiring process. Two kinds of rewiring processes leading to the formation of structures in networks can be distinguished[100]: i) rewirings based on local connectivity properties regardless of the values of the state variables of the nodes, denoted as topological rewirings; and ii) rewirings that depend on the state variables of the nodes, where the link dynamics is coupled to the node state dynamics, called adaptive rewirings. From this point of view both mechanisms, for the origin of small-world networks proposed by Watts and Strogatz, and for the formation of scale-free networks introduced by Barabási and Albert, belong to the class of topological rewiring processes.

In this Chapter, we investigate the emergence of community structure in networks induced by a mechanism of adaptive rewiring, that corresponds to the second class of rewiring processes. We shall show that the dynamics of adaptive rewiring alone is sufficient to explain the emergence of communities. With this aim, we consider the adaptive rewiring process in the general framework for coevolutionary systems from Ref.[99] and described in Chapter 2. Thus, we shall assume no node dynamics; that is $P_c = 0$, such that the states of the nodes do not change in time. Community structure in networks with fixed states for the nodes are commonly observed, for example, in many large cities where neighborhoods representing communities are established according to ethnic features or national origin.

We shall show that the formation of communities is associated with a decrease in the number of active links in the system, i.e., links that connect two nodes in different states. By using a mean field approach, we have found an analytic solution for the evolution of the density of active links in the network.

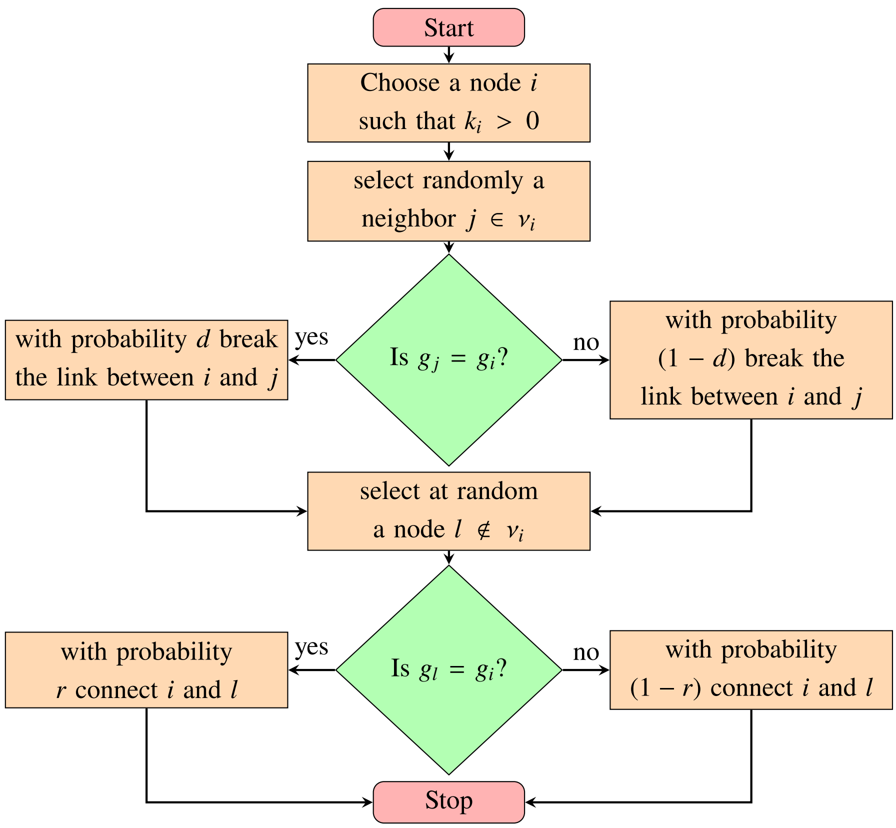
3.2 Adaptive rewiring processes in networks

To investigate the emergence of topological structures such as communities through an adaptive rewiring process, we begin by considering a random network of N nodes denoted by $i = 1, 2, 3, \ldots, N$. Let ν_i be the set of neighbors of node i, possessing k_i elements. We call $\langle k \rangle = (1/N) \sum_{i=1}^{N}$ the mean degree of the network, i.e., $\langle k \rangle$ is the average number of neighbors per node. Then, the average number of links in the network is $\langle k \rangle N/2$.

The state variable of node i is denoted by g_i, which is assumed to be discrete; that is, g_i can take any of G possible values, denoted in the set $\{1, 2, \ldots, G\}$. The states g_i are initially assigned at random with a uniform distribution. Since we assume no node dynamics, the average number of nodes in each state in the network will remain at N/G.

In the general framework of Herrera *et al.*[99], an adaptive rewiring process in a network can be described in terms of the actions of disconnection and connection between nodes. Both actions are based on a criterion for comparison of the states of the nodes. The model assumes conservation of the total number of links in the network; that is, each disconnection action is followed by a reconnection between nodes. These two actions are characterized by the parameters r and d , respectively. The parameter $d \in [0, 1]$ represents the probability that two nodes in identical states become disconnected, so that $1 - d$ is the probability that two nodes in different states disconnect from each other. Similarly, the parameter $r \in [0, 1]$ is the probability that two nodes in identical states become connected, and $1 - r$ is the probability that two nodes in different states connect to each other. Then, a specific adaptive rewiring process is denoted by a pair of values (d, r).

The dynamics of a rewiring process (d, r) on a network is defined by iterating the following algorithm:



A time step corresponds to N iterates this algorithm.

3.3 Analytic mean field approach to adaptive rewiring

A link in the network is called *active* if it connects nodes in different states, while a link connecting nodes in the same state is said *inert*. Let $n(t)$ be the number of active links in the network at time t. Then the average fraction or the density of active links in the network at a given time is the number of active links divided by the average number of links in the network; that is,

$$\rho(t) = n(t)\frac{2\langle k\rangle}{N}. \tag{3.1}$$

Therefore, the fraction of inert links in the network at time t is $(1-\rho(t))$.

In a mean field description, the system is assumed homogeneous. Then, the probability that a given link is active at a time t can be approximated by the average density of links $\rho(t)$. Then, the change Δn in the number of active links in one iteration can be calculated as follows. The probability that an inert link between node i and node $j \in \nu_i$ is randomly selected and becomes active after one update is $d(1-r)$. Since the average fraction of inert links in the network is $(1-\rho)$, then the average increase in the number of active links in one update is $\Delta n_+ = d(1-r)(1-\rho)$. Similarly, the probability that an active link between node i and node $j \in \nu_i$ is randomly selected and becomes inert after one update is $r(1-d)$. Then, since ρ is the average fraction of active links in the network, the average decrease in the number of active links in a time step is $\Delta n_- = r(1-d)\rho$. Thus, the change in the number of active links in one update will be

$$\Delta n = \Delta n_+ - \Delta n_- = d(1-r)(1-\rho) - r(1-d)\rho. \tag{3.2}$$

From Eq. (3.1), we have $n = \frac{\langle k\rangle N}{2}\rho$. Thus, the change in the number of active links can be expressed as

$$\Delta n = \frac{\langle k\rangle N}{2}\Delta\rho. \tag{3.3}$$

Then, we can write

$$\Delta\rho = \frac{2}{N\langle k\rangle}[d(1-r)(1-\rho) - r(1-d)\rho]. \tag{3.4}$$

Then, the change in the average density of active links in a single time interval $\Delta t = 1/N$, in the limit $N \to \infty$, can be expressed by the following differential equation,

$$\frac{d\rho}{dt} = \frac{2}{\langle k\rangle}\left[d(1-r)(1-\rho) - r(1-d)\rho\right]. \tag{3.5}$$

Note that Eq. (3.5) can be seen as a balance equation for the fraction of active links, where the first term on the right hand-side is associated with the increase in the number of active links, while the second term corresponds to the decrease of active links. This competition between creation and destruction of active links is responsible for the emergence of complex topological properties in an dynamical network, such as modular or community structures.

Equation (3.5) can be written as

$$\frac{d\rho}{dt} = \frac{2}{\langle k\rangle}\left[d(1-r) - (d+r-2rd)\rho\right]. \tag{3.6}$$

The solution of Eq. (3.5) is

$$\rho(t) = \rho^* - (\rho^* - \rho_0)e^{-\frac{2}{\langle k\rangle}(d+r-2rd)t}, \tag{3.7}$$

where ρ_0 is the initial density of active links in the network, and

$$\rho^* = \frac{(1-r)d}{r+d-2dr} \tag{3.8}$$

is the stationary or asymptotic solution of Eq. (3.5) for $t \to \infty$.

Note that if the initial density $\rho_0 > \rho^*$, then the density $\rho(t)$ decays to the value ρ^*. Conversely, if $\rho_0 < \rho^*$, the density $\rho(t)$ increases up to the value ρ^*.

3.3.1 Stability analysis of the fixed point solution

Equation (3.5) can be expressed as

$$\frac{d\rho}{dt} = f(\rho^*), \tag{3.9}$$

where

$$f(\rho) = \frac{2}{\langle k \rangle}\left[d(1-r) - (d + r - 2rd)\rho\right] = 0. \tag{3.10}$$

The fixed point or stationary solution $\rho = \rho^*$ of Eq. (3.9) is given by the condition

$$\left.\frac{d\rho}{dt}\right|_{\rho^*} = f(\rho^*) = 0. \tag{3.11}$$

That is,

$$f(\rho^*) = \frac{2}{\langle k \rangle}\left[d(1-r) - (d + r - 2rd)\rho^*\right] = 0, \tag{3.12}$$

which yields

$$\rho^* = \frac{(1-r)d}{r + d - 2dr}. \tag{3.13}$$

The stability of the fixed point solution can be analyzed by considering the behavior of a small perturbation $\Delta\rho$ of it, in the form

$$\rho(t) = \rho^* + \Delta\rho(t). \tag{3.14}$$

Substitution in Eq. (3.9) gives

$$\frac{d}{dt}(\rho^* + \Delta\rho) = f(\rho^* + \Delta\rho). \tag{3.15}$$

By employing a Taylor expansion of the right-hand side, we obtain

$$\cancelto{0}{\frac{d\rho^*}{dt}} + \frac{\Delta\rho}{dt} = \cancelto{0}{f(\rho^*)} + f'(\rho^*)\Delta\rho + O(\Delta\rho^2) \tag{3.16}$$

Then, neglecting second and higher order powers of the small perturbation $\Delta\rho$, we get the equation

$$\frac{\Delta\rho}{dt} = f'(\rho^*)\Delta\rho \tag{3.17}$$

which has the solution

$$\Delta\rho(t) = \Delta\rho(0)e^{f'(\rho^*)t}. \tag{3.18}$$

Therefore, the perturbation $\Delta\rho(t)$ decreases in time if $f'(\rho^*) < 0$, and therefore the fixed point solution ρ^* is *stable*. Alternatively, the fixed point is *unstable* if $f'(\rho^*) > 0$, indicating that the perturbation grows in time.

Then, the fixed point ρ^* is linearly stable if the following condition is fulfilled,

$$\frac{df}{d\rho^*} = \frac{2}{\langle k \rangle}(2rd - d - r) < 0. \tag{3.19}$$

That is,

$$d + r - 2dr > 0. \tag{3.20}$$

Condition Eq. (3.20) for the stability of the fixed point ρ^* is satisfied for $d \in (0,1)$, $r \in (0,1)$. The points $(d,r) = (1,0)$ and $(d,r) = (0,1)$ in a rewiring process correspond to extreme heterophily and extreme homophilly, respectively.

3.3.2 Time evolution of the density of active links

Figure 3.1 shows our mean field analytic solution $\rho(t)$ in Eq. (3.7) (red line) as a function of time for several rewiring processes (d, r). The stationary, asymptotic solution ρ^* is generally reached for short times. For comparison, Fig. 3.1 also displays the time evolution of the fraction of active links calculated from numerical simulations on random networks. The errors are about 10%. Thus, the mean field solution provides a reasonable description of the adaptive rewiring dynamics in the system.

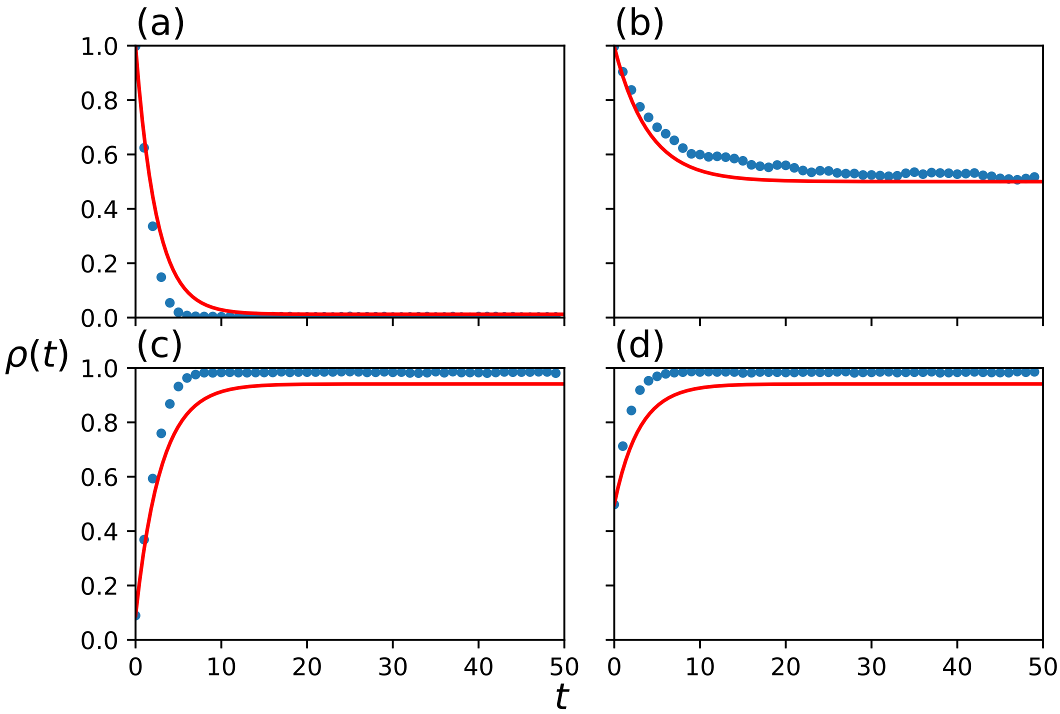


Figure 3.1: Evolution of the density of active links as a function of time for different rewiring processes (d, r) on different initial networks. The red continuous line is the analytic solution Eq. (3.7) and the blue dots correspond to the density of active links calculated from numerical simulations. The simulations were performed on networks with parameters: $N = 3200$, $\langle k \rangle = 4$ and $G = 320$. (a) $(r, d) = (0.9, 0.1)$, initial random network. (b) $(r, d) = (0.5, 0.5)$, initial random network. (c) $(r, d) = (0.2, 0.8)$, initial network with community structure. (d) $(r, d) = (0.2, 0.8)$, initial network with medium density of active links.

3.3.3 Density of active links on rewiring space

Figure 3.2(a) shows the analytic stationary density of active links ρ^* on the space of rewiring parameters (d, r), given by Eq. (3.13)

$$\rho^* = \frac{(1-r)d}{r+d-2dr}. \tag{3.21}$$

For each point (d, r) on this plane, the value ρ^* is displayed by means of a color code indicated by a bar to the right of the figure. Darker colors correspond to smaller values of ρ^*, while lighter colors represent higher values of ρ^*.

The curves of constant density (“isothermal") $\rho^*(d, r) = C =$ constant, are given by

$$d = \frac{Cr}{(1-r) + C(2r-1)}. \tag{3.22}$$

Figure 3.2(b) shows the numerical calculation of the asymptotic density of active links on the space of parameters (d, r), for a random network characterized by $N = 3200$, $\langle k \rangle = 4$, and $G = 320$ possible states per node.

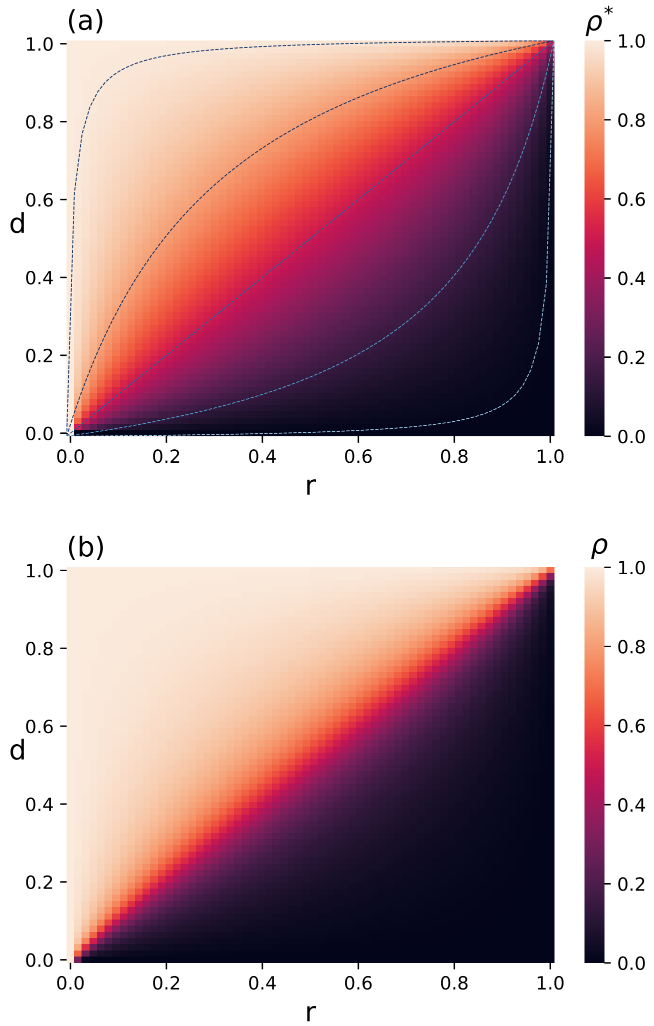


Figure 3.2: (a) Analytic stationary density of active links ρ^* on the space of parameters (r, d). “Isothermal" curves for constant ρ^*, Eq. (3.22), are indicated by dashed lines. b) Asymptotic density of active links on the space (d, r), calculated from numerical simulations on a random network characterized by $N = 3200$, $\langle k \rangle = 4$, and $G = 320$. Each data point shown corresponds to the average over 100 realizations of initial conditions on the network. For both panels, the values of the density are represented as heat maps; the corresponding color code bars are displayed on the right side in each case.

We call $E_{\text{abs}} = |\rho^* - \rho|$ the absolute error between the analytic solution ρ^* and the asymptotic value of the density of active links ρ calculated from numerical simulations. Figure 3.3(a) shows the absolute error $E_{\text{abs}} = |\rho^* - \rho|$ on the space of parameters (d, r). In Fig. 3.3(b), the quantities ρ^* and ρ are plotted as functions of r for values of d along the diagonal $d = 1 - r$.

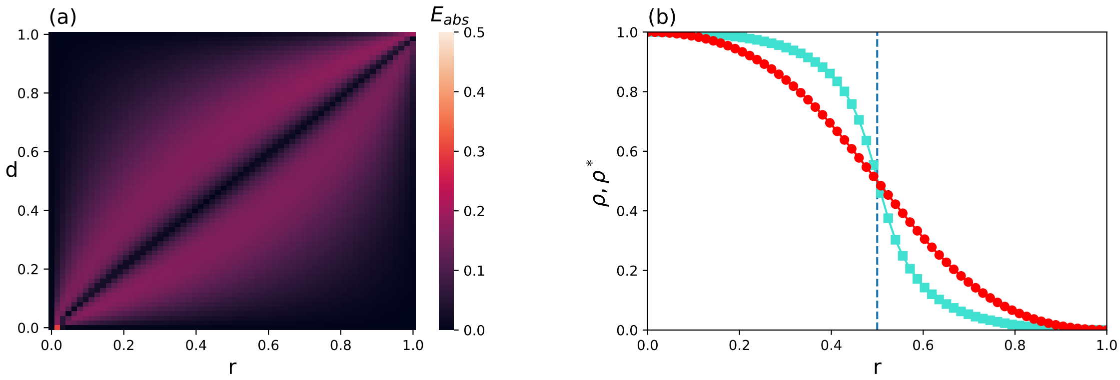


Figure 3.3: Absolute error between the analytic solution ρ^* and the numerical value ρ obtained solution for the asymptotic density of active links.(a) Absolute error $E_{\text{abs}} = |\rho^* - \rho|$ on the space of parameters (d, r). Network parameters are $N = 3200$, $G = 320$, $\langle k \rangle = 4$. (b) Quantities ρ^* (red circles •) and ρ (blue squares ■) as functions of r along the diagonal $d = 1 - r$. Each data point shown corresponds to the average over 100 realizations of initial conditions on the network.

The analytic model is a mean field approximation that assumes infinite size and complete homogeneity of the density of active links for all values of parameters (d, r). These simplifying assumptions are not achieved in the simulations, especially for some parameters. Thus, one should not expect complete agreement between the mean field model and the simulation for all parameter values. The region of parameters $d \to 1$, $r \to 0$, corresponds to heterophily, where connections between nodes in different states are strongly favored and sustained. There is little probability that nodes in equal states connect. Then, one should expect $\rho \to 1$ in this region, as the simulation actually shows. Similarly, the region $d \to 0$, $r \to 1$ describes homophily, where connections between equals become highly probable, which implies $\rho \to 0$, as seen in the simulation. In either limit, heterophily or homophily, the system reaches a more homogeneous density of active links, as it is assumed in the mean field approach. The mean field approximation gives $\rho^* = 1$ and $\rho^* = 0$ in these two limits respectively, providing a good agreement with the simulations in these regions, as shown in Fig. 3.3(a)-(b).

In the region $d = r$, the actions of disconnection and connection actions between equal nodes have the same probability, as well as these actions have equal probability for nodes in different states. Then one may expect that, on the average, half of the links between should be active, giving a homogeneous density $\rho = 0.5$. Actually, the simulations give values close to $\rho = 0.5$ along the region $r = d$, while the mean field model exactly yields $\rho^* = 0.5$ for $r = d$. Thus, the errors along the line $d = r$ are small, as Fig. 3.3(a) shows.

The analytic model deviates from the simulation for parameter values outside these regions, where the rewiring actions produce a network possessing a less homogeneous distribution of active links.

3.4 Adaptive rewiring and emergence of communities in networks

As we have mentioned, a main objective of the this thesis is to investigate how a process of adaptive rewiring can induce the formation of communities in networks. We assume the rewiring process based on connection-disconnection actions proposed in the general framework for coevolutionary dynamics in networks of Ref.[99].

Homophilic rewiring processes, where $d \to 0$ and $r \to 1$, favor the connections between nodes in similar states, and therefore reduce the number of active links in the network. In finite systems, such processes may lead to the fragmentation of the network. We characterize the integrity of a network by calculating the average normalized (divided by N) size of the largest component or connected subgraph in the system, regardless of the states of the nodes, denoted by S_m. Figure 3.4 shows the quantity S_m numerically calculated for an initial random network on the space of parameters (d, r). We see that, along some critical curve on this plane, S_m exhibits a transition from a regime having a large connected component, for which $S_m \approx 1$, to a fragmented state consisting of small, separated components, for which $S_m \approx 0$.

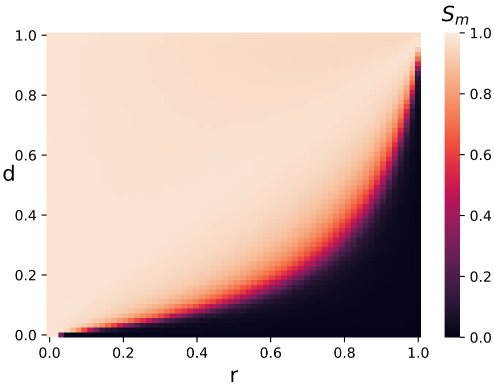


Figure 3.4: Average normalized size S_m of the largest subgraph on the plane (r, d). Network parameters are $N = 3200$, $\langle k \rangle = 4$, $G = 320$. Each data point shown corresponds to the average over 100 realizations of initial random conditions for the network.

We also investigate the presence of modular structure in the network on the plane (d, r). As a measure for the formation of modular structure, we define the quantity $\Delta Q \equiv Q(t \to \infty) - Q(0)$ as the modularity change, where $Q(0)$ is the modularity of the initial random network and $Q(t \to \infty)$ is the stationary modularity of the network for asymptotic times, both calculated through the Lovain's community detection algorithm[67]. Then, a value $\Delta Q = 0$ reflects the subsistence of the initial random topology, while $\Delta Q > 0$ indicates an increase in modularity with respect to the initial random network, independently of the absolute values of Q.

Figure 3.5 displays the modularity change ΔQ, averaged over several realizations of initial conditions for a random network, on the space of parameters (d, r). By comparing Figs. 3.4 and 3.5, we see that parameters for which $\Delta Q > 0$ include values (d, r) where the network is fragmented. The Lovain algorithm considers a network with separated subgraphs as communities, assigning maximum modularity to it. Thus, ΔQ alone cannot effectively distinguish the existence of connected module from a fragmented network; both situations yield $\Delta Q > 0$.

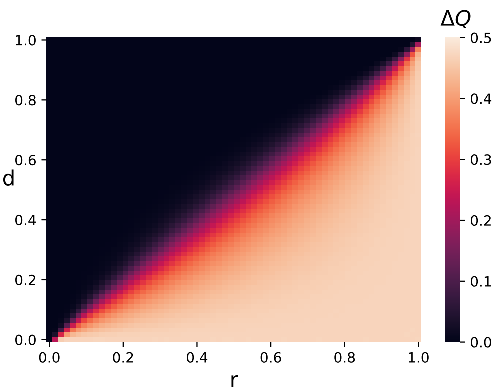


Figure 3.5: Modularity change ΔQ numerically calculated on the plane (d, r). The values of ΔQ are indicated by a color code bar on the right. Network parameters are $N = 3200$, $\langle k \rangle = 4$, $G = 320$. Each value shown corresponds to the average over 100 realizations of initial random conditions for the network.

To elucidate the formation of communities in the network, we plot in Fig. 3.6 the density of active links ρ, the average normalized size of the largest subgraph S_m, and the modularity change ΔQ, as functions of r , numerically calculated for rewiring parameters along the diagonal $d = 1 - r$.

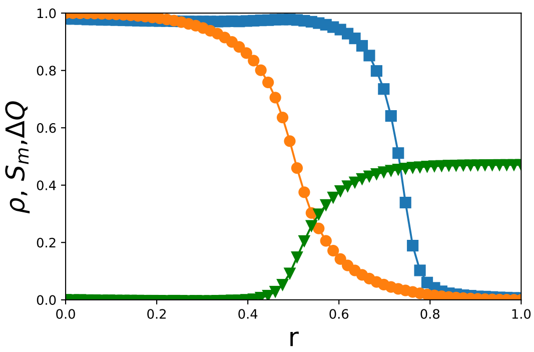


Figure 3.6: Density of active links ρ (orange circles •), average normalized size of the largest subgraph S_m (blue squares ■), and modularity change ΔQ (green triangles ▼), as functions of r , calculated along the diagonal $d = 1 - r$. Network parameters are $N = 3200$, $\langle k \rangle = 4$, $G = 320$. Each point shown corresponds to the average over 100 realizations of initial random conditions for the network.

A fragmented network possesses values $S_m \approx 0$, a maximum increase of modularity ΔQ, and $\rho = 0$. In similar fashion, community structure is characterized by the presence of a large connected component ($S_m \approx 1$), with an increase of modularity with respect to the initial random network ($\Delta Q > 0$), and with a small number of active links ($\rho \ll 1$). This situation occurs for intermediate values of r in Fig. 3.6. Then, Fig. 3.6 suggests a criterion for discerning and characterizing the emergence of community structure. We consider the product $\Delta Q \times S_m$ as a statistical quantity that measures the presence of communities in a network. If $\Delta Q \times S_m \approx 0$, there is no increase in modularity or the network is fragmented. In either case, no community structure exists in the network. In contrast, a large value of the product $\Delta Q \times S_m$ signifies that a community structure has arisen on a connected large subgraph.

Figure 3.7(a) shows the quantity $\Delta Q \times S_m$ calculated on the space of parameters (d, r). The “half-moon" colored region, labeled II, corresponds to values $\Delta Q \times S_m > 0$ that indicate the range of rewiring parameters d and r for which a modular structure emerges in the system. The boundaries of this modularity region can be seen as curves of constant density ρ^* on the plane (d, r), as in Fig. 3.2(a).

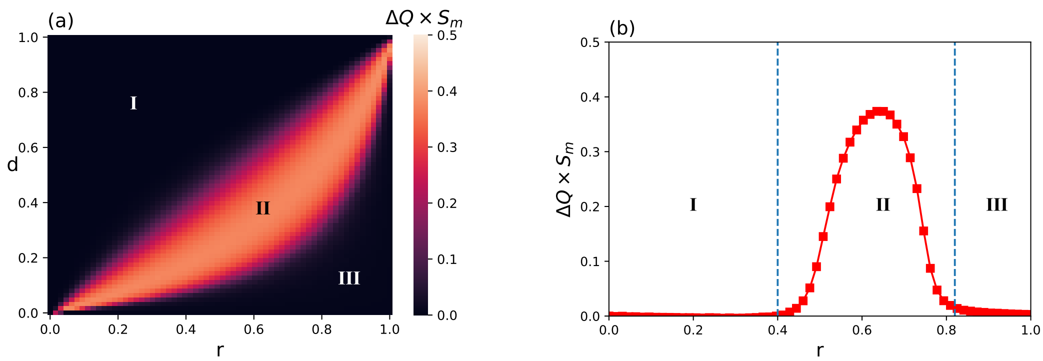


Figure 3.7: (a) Quantity $\Delta Q \times S_m$ calculated on the space of parameters (d, r). The values of $\Delta Q \times S_m$ are indicated by a color code bar on the right. (b) $\Delta Q \times S_m$ as a function of r for parameters along the diagonal $d = 1 - r$. Network parameters in each panel: $N = 3200$, $\langle k \rangle = 4$, $G = 320$. Each point shown corresponds to the average over 100 realizations of initial random conditions for the network.

Three regions, associated with different structures or “phases" of the network, can be distinguished on the space of parameters (d, r) in Fig. 3.7(a): (I) a region where a large random subgraph persists, characterized by $\Delta Q = 0, S_m \approx 1$; (II) a region where communities emerge on a connected graph, corresponding to $\Delta Q > 0, S_m \approx 1$; and (III) a region where the network is fragmented in small separated components; characterized by $\Delta Q > 0, S_m \approx 0$.

The product $\Delta Q \times S_m$ as a function of r along the diagonal $d = 1 - r$ is shown in Fig. 3.7(b). Community structure takes place on region II for intermediate values or r and d . As mentioned in Fig. 2.2, moving along the line $d = 1 - r$ describes a continuous transition between the points of total heterophily or tolerance $(d, r) = (1, 0)$ and total homophily or intolerance $(d, r) = (0, 1)$. Near these points, $\Delta Q \times S_m \approx 0$. No community structure can be sustained close to either situation. Instead, Fig. 3.7(b) revels that communities can arise from an initial random network for intermediate values of the rewiring parameters d and r .

We recall that active links connect nodes in different communities that have distinct states, while inert links connect nodes in the same state within a community. As we mentioned, a community structure is characterized by $S_m \approx 1$ and $\rho \ll 1$, which means a large density of inert links, $1 - \rho$, that belong to communities. Then, we may consider the product $S_m \times (1 - \rho)$ as an indicator of modular structure. Fig. 3.8(a) shows the product $S_m \times (1 - \rho)$ on the plane of parameters (d, r). A colored region where $S_m \times (1 - \rho) \neq 0$ can be seen in Fig. 3.8(a); this region approximately coincides with the “half-moon" region where community structure emerges in Fig. 3.7(a). Figure 3.8(b) shows $S_m \times (1 - \rho)$ as a function of r along the diagonal line $d = 1 - r$. Again, community structure appear for intermediate values of r and d . The quantity $\Delta Q \times S_m$, that includes the modularity change ΔQ, provides a more precise measurement of community structure. However, a valuable insight into the formation of communities can be gained by considering the product $S_m \times (1 - \rho)$ that is numerically simpler to calculate than $S_m \times \Delta Q$.

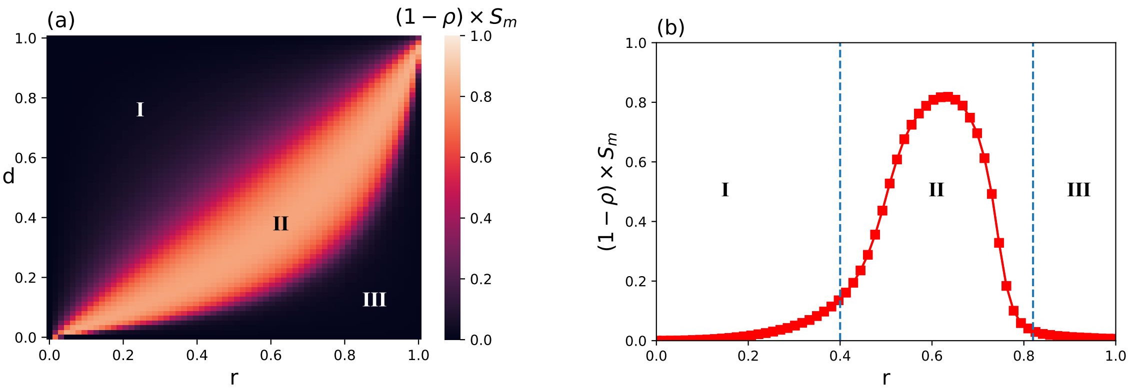


Figure 3.8: (a) Quantity $S_m \times (1 - \rho)$ on the plane (d, r). The values of $S_m \times (1 - \rho)$ are indicated by a color code bar on the right. (b) $S_m \times (1 - \rho)$ as a function of r for parameters along the diagonal $d = 1 - r$. Network parameters in each panel: $N = 3200$, $\langle k \rangle = 4$, $G = 320$. Each point shown corresponds to the average over 100 realizations of initial random conditions for the network. Labels I, II, and III, indicate the three phases of the network.Label I corresponds to the persistence of random network structure, label II corresponds to the emergence of community structure and label III indicates the fragmentation of the network.

Figures 3.7(a) and 3.8(a) reveal the relationship between the density of active links and the emergence of community structure in a network. Since active links connect different domains, when their density is low, the majority of links must lie inside the different domains coexisting on the large connected network. Therefore, there must exist several domains inside which nodes are highly connected, with fewer connections between different domains. This is the main characteristic of a modular or community structure in a network. The formation of communities in the “half moon” region on Fig. 3.7(a) requires the presence of more homophily than heterophily (that is, $r > d$) in the interactions between agents, but not too much homophily (that is, $r < 1$) that would lead to fragmentation of the network.

Figure 3.9 displays the corresponding network configurations resulting from numerical simulations with parameters in regions I, II, and III on the space of rewiring parameters (d, r) in Fig. 3.7(a).

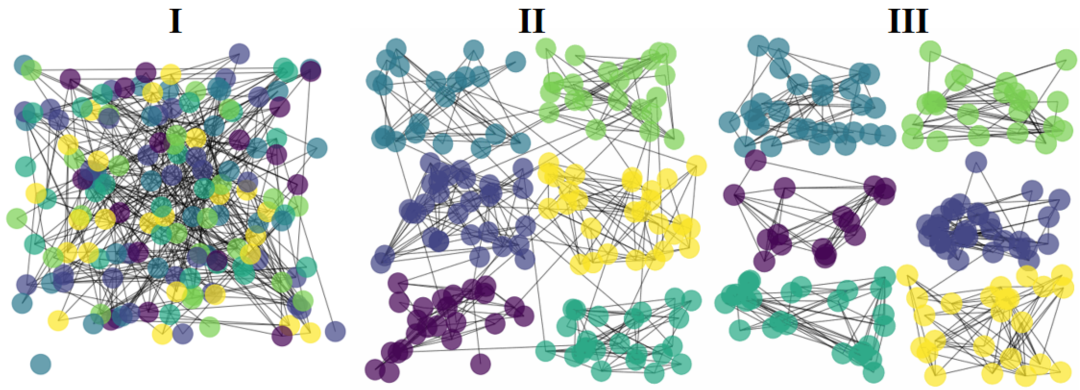


Figure 3.9: Asymptotic network configurations for the three regions of the space of rewiring parameters (d, r) indicated in Fig. 3.7(a). Fixed network parameters: $N = 150$, $G = 6$. (I) connected random network; (II) community structure; (III) fragmented network.

In summary, in this Chapter we have characterized the formation of communities in a network by the statistical quantity $\Delta Q \times S_m$ on the space of parameters (d, r) for connection and disconnection that describe an adaptive rewiring process. We have shown that an adaptive rewiring process, even in the absence of node dynamics, can give rise to a community structure in a network in an intermediate region on the space of parameters (d, r). Furthermore, we have identified the topological phases for the network on this space: a large connected graph, a community structure, and a fragmented state. As we shall see in Chapter 4, this phase diagram on the space of of parameters (d, r) allows one to predict the emergence of communities and the fragmentation of a network, when coevolution with dynamics for the states of nodes is considered.

Chapter 4

A coevolution model with node dynamics

4.1 Coevolution model of discrete opinions with interaction threshold on networks

As an application of the general model for coevolutionary dynamics[99] presented in Chapter 2, in this Chapter we shall investigate a system where both processes, adaptive rewiring and node dynamics, coexist and are coupled on a network. Thus, in contrast with the previous Chapter where there was no dynamics for nodes ($P_c = 0$), here the nodes can interact and change their states as the system evolves.

The node dynamics we shall use are motivated by the following opinion formation problem of wide interest: if the social agents that rate a product or service on a discrete-valued scale interact on a real or virtual network, under what conditions do their opinions converge? Furthermore, can communities emerge from such interactions?

Valued or ranked opinions of individuals are common in social systems. For example, many companies and institutions often ask their costumers to evaluate the quality of a product or the level of satisfaction with a service through some valuation method. Typically, companies use a Likert scale[108] that includes a number of options or answers for a question, from “strongly agree" to “strongly disagree". This system of qualification consists of discrete, valued opinions expressed as integer numbers, $1, 2, \ldots, G$, where 1 (“strongly agree") and G (“strongly disagree") represent the extreme values that the opinion about a subject can take. Companies such as Amazon, Ebay, Mercado Libre, banks, car dealers, mobile phone, fast foods, e-commerce, among others, employ these kinds of opinion surveys to improve their services and profits. The political spectrum on the left–right dimension can also be classified with this scheme; for example: 1 (extreme left), 2 (center-left), 3 (center), 4 (center-right), 5 (extreme right). In many cases, the discrete opinion value is expressed by some symbol or a number of "stars".

Figure 4.1 illustrates two situations where valued opinions of users are considered on different scales.

Various models of discrete-valued opinion dynamics have been employed in the sociodynamics literature[109,110]. Here we use the node interaction dynamics of Ref.[111] that can be considered as a discrete version of the bounded confidence opinion model of Deffuant *et al.*[101].

Thus, we define a coevolution model of discrete-valued opinions with a threshold for interaction as follows. Consider N agents or nodes forming a network with mean degree $\langle k \rangle$. Let g_i be the state of opinion of agent i, where g_i can take any of the G integer values in the scale $\{1, 2, 3, \ldots, G\}$. Define a threshold for interaction between agents by a parameter U that can take a value in the set $\{1, 2, 3, \ldots, G\}$. The parameter U can be interpreted as the degree of tolerance or openness to interactions that agents possess. For simplicity, we assume homogeneity, i. e.; the threshold value U is the same for all agents in the system. The states g_i are initially assigned at random with a uniform distribution in the set $\{1, 2, 3, \ldots, G\}$. We set a rewiring process characterized by the connection and disconnection parameters (d, r).

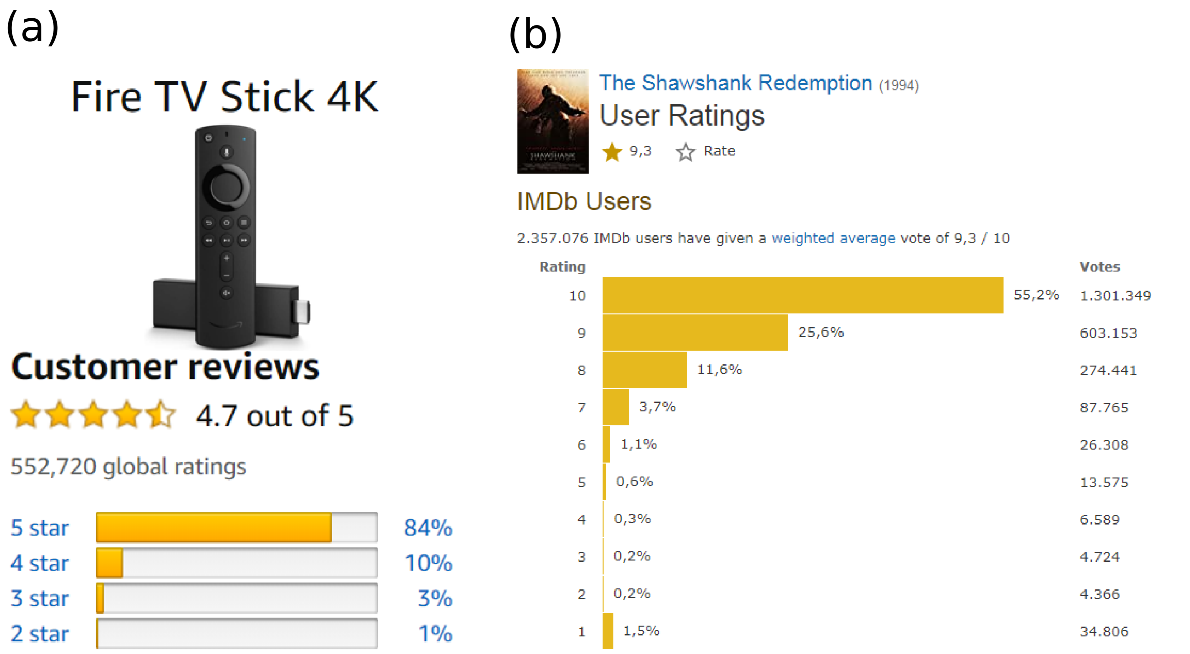


Figure 4.1: (a) Customer reviews of a product on Amazon's platform on a scale from 1 to $G = 5$, with $N = 552,720$ entries. (b) Viewers ratings for the movie *The Shawshank Redemption* (1994) on the Internet Movie Database (IMDb) on a scale from 1 to $G = 10$, with $N = 2,357,076$ entries.

Then, the coevolutionary dynamics of the system consists of iterating the following steps:

1. Choose a node i such that $k_i > 0$.

2. With probability P_r apply rewiring process (d, r): (i) select randomly a neighbor $j \in \nu_i$; if $g_j = g_i$ with probability d break the link between i and j; if $g_j \neq g_i$ with probability $(1 - d)$ break the link between i and j. (ii) select at random a node $l \notin \nu_i$; if $g_l = g_i$, with probability r connect i and l; if $g_l \neq g_i$, with probability $(1 - r)$ connect i and l.

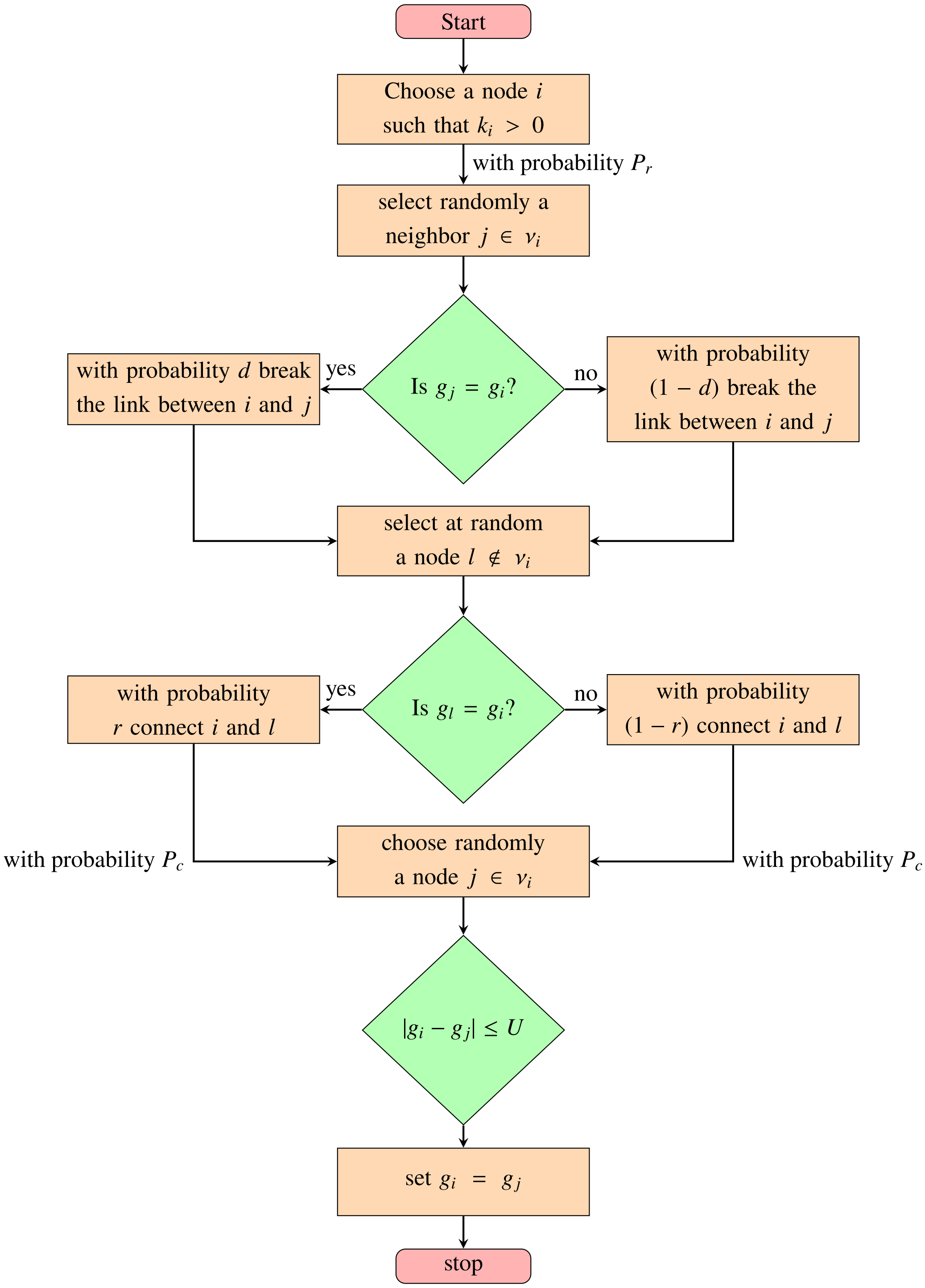
3. With probability $P_c = 1 - P_r$, choose randomly a node $j \in \nu_i$. If If $|g_i - g_j| \leq U$, set $g_i = g_j$. If $|g_i - g_j| > U$, nothing happens.

Step 2 describes the adaptive rewiring process (d, r) that allows the acquisition of new connections. Step 3 specifies the coupling between the processes of rewiring and node dynamics according to the relation $P_c = 1 - P_r$; the process of node state change is such that the states of the connected nodes become similar if the threshold condition is fulfilled. Different coevolution models with this node dynamics can be investigated by considering different rewiring parameters (d, r) or different coupling functions $P_c = f(P_r)$.

Note that this type of node dynamics based on imitation or copying is absorbing. This means that the number of states existing in the system decrease in time. A stationary state is reached in a finite time for a finite system.

The total number of links in our network and the total number of possible opinions G are fixed. In the limit of large system size, the model thus has four parameters: the mean degree $\langle k \rangle$, the mean number of nodes holding a particular opinion N/G, the rewiring probability P_r, and the threshold value U. Although the interaction threshold U is an integer number, it is convenient to express this quantity normalized respect to the number of possible options G . Then, we define the normalized threshold for interaction as $u \equiv U/G$, such that $u \in (0, 1]$. We shall fix $\langle k \rangle$ and N/G and study the collective behavior of the system in terms of the parameters P_r and u.

The algorithm for this process is illustrated in the next flow diagram:



4.2 Network fragmentation

Let us assume the coevolution model of Sec. 4.1 with a rewiring process with parameters $(d, r) = (0.5, 1)$ in region III of Fig. 3.7(a). Rewiring parameters in region III favor the fragmentation of the network. The fragmentation phenomenon is controlled by the time scale of the rewiring process, expressed by the probability P_r. Whereas, the chosen dynamics for the nodes, based on an imitation rule that depends on the threshold value u, tends to increase the number of connected nodes with equal states, promoting the formation of domains. Recall that a domain is a set of connected nodes or subgraph where all members of the subgraph share the same state.

To characterize the collective behavior of the system ensued from the competition between the two processes, we employ, as a statistical order parameter, the average normalized size of the largest domain in the system, denoted by S_g. Figure 4.2 shows the quantity S_g calculated as a function of parameters (P_r, u).

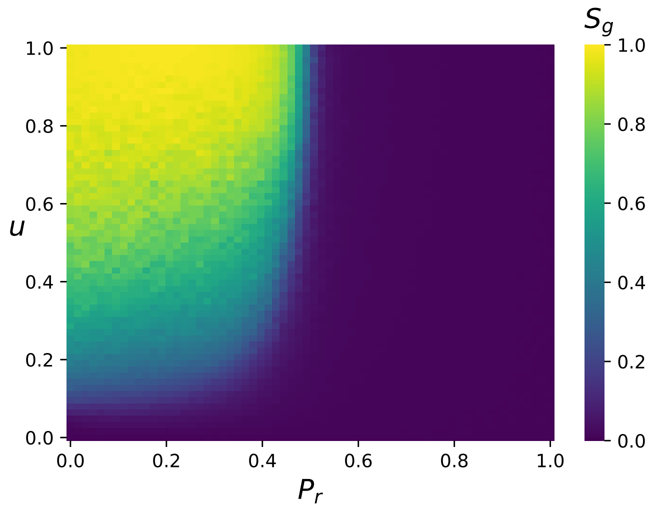


Figure 4.2: Average normalized size S_g of the largest domain on the space of parameters (P_r, u) for the coevolutionary system with $P_c = 1 - P_r$ and discrete-valued opinion node dynamics with $G = 160$. Rewiring parameters: $(d = 0.5, r = 1)$. Network parameters are $N = 1600$, $\langle k \rangle = 4$. The values of S_g are indicated by a color code bar on the right. Each data point shown corresponds to the average over 25 realizations of initial random conditions for the network.

Figure 4.2 displays two main regions on the plane (P_r, u), resulting from the competition between the node dynamics and the rewiring process: (i) a (yellowish) region where $S_g \approx 1$, corresponding to the presence of a large domain or a homogeneous state; and (ii) a (blue) region where $S_g \approx 0$, signaling the existence of a network fragmented in small domains or a disordered state. By varying the parameters P_r and/or u on this plane, the system can exhibit a continuous transition between these two collective states or phases. Fragmentation of the network occurs for large values of the probability P_r; that is, when the rewiring process is faster that the node dynamics that takes place with probability $P_c = 1 - P_r$. Alternatively, a consensus or homogeneous collective state is achieved when the tolerance threshold u of the social agents is large enough and their connections change at a low rate P_r.

The phase diagram of Fig. 4.2 is quite general; it contains, as special cases, various opinion formation models with discrete states previously investigated. For example, the case $u \to 0$ implies that no opinion changes take place. Then, we have the model of adaptive rewiring without node dynamics from Chapter 3, where the domains have small average size $S_g = G/N < 1$. Other cases will be considered next.

4.3 Limiting case $u = 1$

Consider a fixed interaction threshold value $u = 1$ in Fig. 4.2. Then, the node dynamics becomes a complete imitation or copying endeavor between neighbors, regardless of the values of their opinion states. This is just the interaction rule for the well-known voter model where the opinion options are analogous to spin states. Then, the case $u = 1$ in Fig. 4.2 becomes a coevolutionary voter model with G equivalent options[104]. This is actually the coevolution model of Holme and Newman[94]. We obtain this model with $(d, r) = (0.5, 1)$ as a special case by rewriting step 3 in the algorithm of Sec. 4.1 as follows:

3. With probability $P_c = 1 - P_r$, choose randomly a node $j \in \nu_i$ and set $g_i = g_j$.

Figure 4.3(a) shows the average normalized size of the largest domain S_g as a function of the rewiring probability P_r, calculated with the modified step 3, equivalent to the condition $u = 1$.

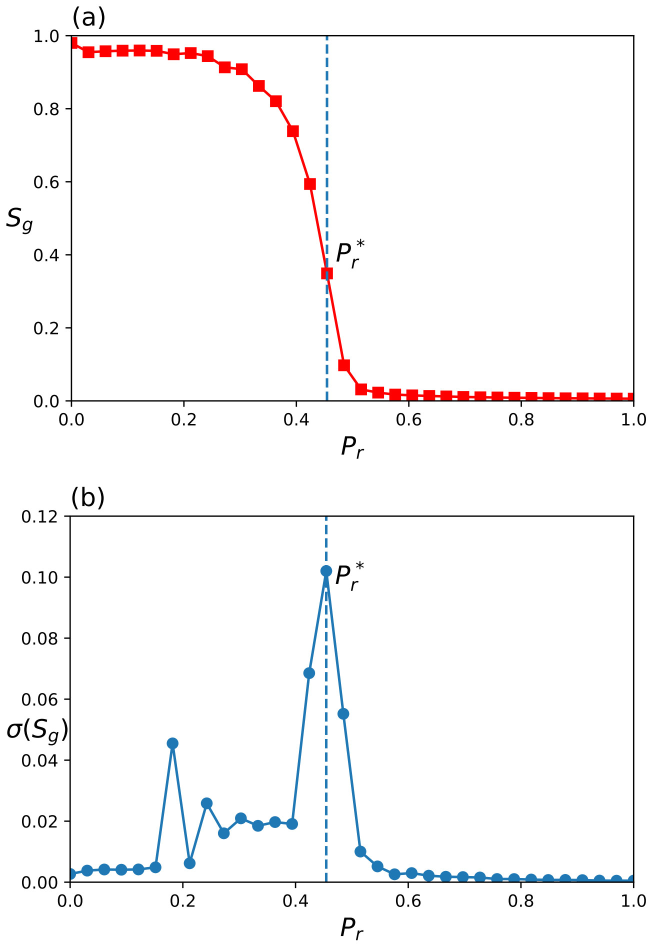


Figure 4.3: (a) Average normalized size of the largest domain S_g as a function of the rewiring probability P_r for the coevolutionary system with $P_c = 1 - P_r$ and discrete-valued opinion node dynamics. Fixed parameters are: $G = 320$, $u = 1$, $(d, r) = (0.5, 1)$. Network parameters are $N = 3200$, $\langle k \rangle = 4$. Each point corresponds to the average over 100 realizations of initial random conditions for the network. (b) Standard deviation $\sigma(S_g)$ about the mean value S_g as a function of the probability P_r. The deviation $\sigma(S_g)$ is calculated over 100 realizations of initial random conditions.

The quantity S_g in Fig.4.3(a) exhibits transition at a critical value P_r^*, from a regime having a large domain for $P_r < P_r^*$, to a fragmented state consisting of small domains for $P_r > P_r^*$, characterized by $S_g \approx 0$. The critical point $P_r^* = 0.46$ for the fragmentation transition is estimated by the value of P_r for which the largest standard deviation or fluctuation of the mean value S_g occurs, as shown in Fig.4.3(b). The transition becomes better defined in the large system limit $N \to \infty$. Holme and Newman[94] found the critical point value $P_r^* = 0.458$ by performing a finite size scaling analysis.

4.4 Limiting case $P_r = 0$

Consider the discrete valued opinion dynamics with interaction threshold on a fixed network. This corresponds to the line $P_r = 0$ along the u variable on the plane in Fig. 4.2, independently from the the parameters (d, r). This model has been actually studied on several fixed networks in Ref.[111]. Here we derive it as a limiting case of our general coevolution model when $P_r = 0$.

Thus, in the algorithm for the coevolutionary dynamics of Sec. 4.1 we delete step 2 and fix $P_c = 1$ in step 3; that is, the node dynamics is always applied on the random network. In this case, no fragmentation of the network takes place. Figure 4.4 shows the average normalized size of the largest domain S_g as a function of the normalized threshold value u for a random network.

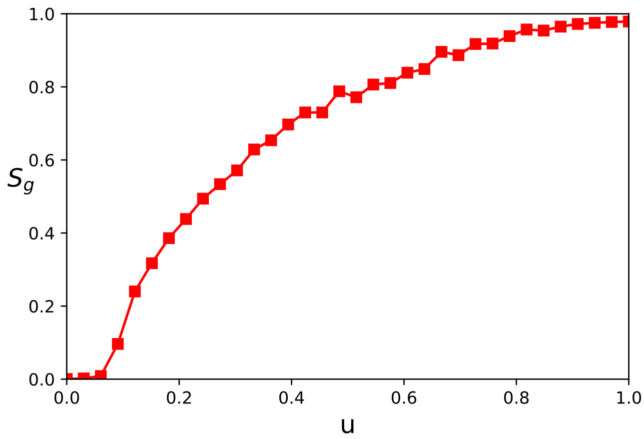


Figure 4.4: Average normalized size of the largest domain S_g as a function of the threshold u for the discrete-valued opinion model with $G = 320$ on a fixed random network ($P_r = 0$). Network parameters are $N = 3200$, $\langle k \rangle = 4$. Each point corresponds to the average over 100 realizations of initial random conditions for the network.

For values of the interaction threshold below a value $u_c = 0.09$ the system reaches a state of disorder, in the sense that a diversity of opinions coexist, characterized by $S_g \approx 0$. For values $u > u_c$ the quantity S_g increases as opinions converge, up to the value $S_g = 1$ that corresponds to a homogeneous or ordered state. The value $u_c = 0.09$ represents a critical value for a continuous non-equilibrium transition between these two collective states or phases in the system. Notice that the simple majority $S_g = 0.5$, where one domain occupies half of the network, is reached in Fig. 4.4 at the value of the interaction threshold $u = 0.25$.

In general, the critical value u_c depends on the topology of the network. For a fully connected network, where each node is linked to all other nodes in the system, it has been shown[111] that $u_c = 0$; that is, no disordered phase appears.

The ordered and disordered phases can be nicely visualized on a fixed two-dimensional network for different values of the interaction threshold u above and below the corresponding critical value, as shown in Fig. 4.5.

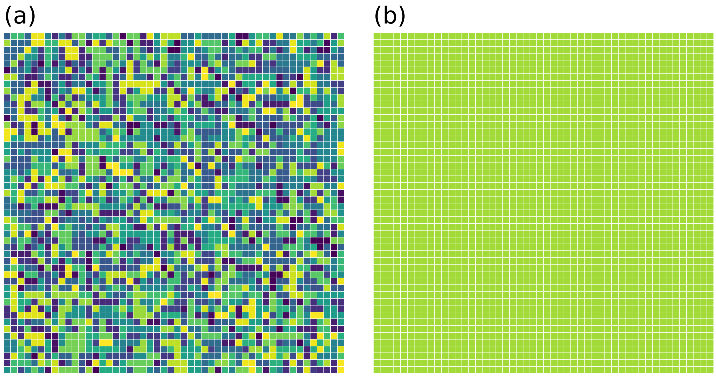


Figure 4.5: Visualization of the resulting patterns for the discrete-valued opinion model with $G = 100$ on a fixed ($P_r = 0$) two-dimensional network (or lattice) with Von Neumann neighborhood and periodic boundary conditions ($k_i = 4, \forall i$), for two different values of the threshold u. Contiguous nodes with the same color belong to the same domain. Network size is $N = 50 \times 50$. The critical value for a two-dimensional lattice, found in Ref.[111], is $u_c = 0.23$ (a) $u < u_c$, disordered phase where several domains coexist ($S_g \approx 0$). (b) $u > u_c$, ordered, homogeneous phase where only one domain emerges ($S_g = 1$).

4.5 Emergence of community structure in coevolutionary networks

We have found in Chapter 3 that rewiring processes corresponding to parameters (d, r) in region II of Fig. 3.7(a) lead to the emergence of community structure in networks. In this Section we investigate the coevolution model of Sec. 4.1 with rewiring parameters $(d, r) = (0.35, 0.65)$, just in region II. The probability P_r regulates the time scale of this rewiring process and also controls the time scale of the node dynamics through the coupling relation $P_c = 1 - P_r$. Thus, by varying P_r, we should expect to promote the formation of communities in the network.

As in Chapter 3, to characterize the emergence of community structure on a connected network, we need two statistical quantities: the average normalized size of the largest connected subgraph, regardless of the states of the nodes, S_m; and the modularity change with respect to the initial random network, ΔQ. We have obtained $S_m \approx 1$ for all values of parameters (u, P_r), indicating that the network always remains connected. The product $S_m \times \Delta Q$, that characterizes the appearance of modular structure, is shown on the space of parameters (u, P_r) in Fig. 4.6.

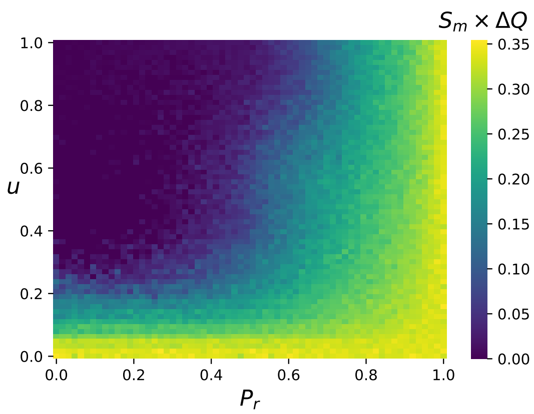


Figure 4.6: Quantity $S_m \times \Delta Q$ on the space of parameters (u, P_r) for the coevolution model with rewiring parameters $(d, r) = (0.35, 0.65)$. The values of $S_m \times \Delta Q$ are indicated by a color code. Initial network parameters are $N = 1600$, $\langle k \rangle = 4$, $G = 160$. Each point is the average over 50 realizations of initial random conditions for the network.

Figure 4.6 exhibits two main regions on the plane (u, P_r): (i) a (yellowish) region where the product $S_m \times \Delta Q > 0$, corresponding to the formation of modular structures on a connected network; (ii) a (blue) region where $S_m \times \Delta Q \rightarrow 0$, indicating that no modular structures have arised from the initial random network. By varying the parameters P_r or u on this space, the system can experience a continuous transition between these two states or phases of the network topology.

Figure 4.7(a) shows the average normalized size of the largest connected subgraph S_m and the product $S_m \times \Delta Q$ as functions of the rewiring probability P_r, with fixed interaction threshold value $u = 1$. Figure 4.7(b) shows S_m and the product $S_m \times \Delta Q$ as functions of the interaction threshold u, with fixed rewiring probability value $P_r = 0.4$.

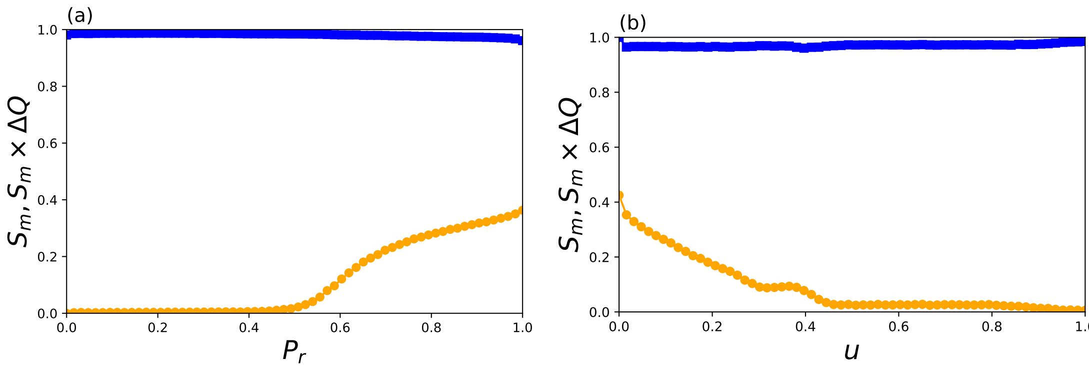


Figure 4.7: (a) S_m (blue squares ■) and $S_m \times \Delta Q$ (right vertical axis, yellow circles •) as functions of the rewiring probability P_r, with fixed value $u = 1$. (b) S_m(blue squares ■) and $S_m \times \Delta Q$ (right vertical axis, yellow circles •) as functions of the interaction threshold u, with fixed value $P_r = 0.4$. Rewiring parameters are $(d, r) = (0.35, 0.65)$. Network parameters are $N = 3200$, $G = 320$, $\langle k \rangle = 4$. In both panels, each point corresponds to the average over 100 realizations of initial random conditions for the network.

Recall from Sec. 4.3 that the case $u = 1$ corresponds to the node dynamics of the Holme-Newman coevolution model[94]. However, now the rewiring parameters $(d, r) = (0.35, 0.65)$ belong to region II where communities emerge and, therefore, no fragmentation of the network occurs. This is reflected by the value $S_m \approx 1$ independently of P_r in Fig. 4.7(a). That said, the quantity $S_m \times \Delta Q$ increases from 0 above the parameter value $P_r = 0.5$, signaling the onset for the formation of community structure in a connected network. Thus, Fig. 4.7(a) reveals that an adaptive rewiring process with connection-disconnection parameters in region II can induce the emergence of communities in a network even in coexistence with an absorbing node dynamics.

Figure 4.7(b) shows a continuous transition from a network with community structure $S_m \times \Delta Q > 0$ to a random network, as the threshold u is increased. These results can be understood as follows. For values $u \rightarrow 0$, nodes do not change their states very often, which is equivalent to having mainly rewiring in the network, analogous to $P_r \rightarrow 0$. Since the rewiring parameters $(d, r) = (0.35, 0.65)$ belong to region II, communities emerge due to the dominant rewiring process. Thus, $S_m \times \Delta Q$ increases for small values of u. For larger values of u, nodes are more likely to copy states from their neighbors, and therefore the system becomes more homogeneous while the network remains random. This leads to the decrease of the product $S_m \times \Delta Q$ as u increases.

Chapter 5

Conclusions & Outlook

As described in the Introduction, the small-world phenomenon in networks has been explained by a mechanism of random rewiring that creates a number of long-range connections[24], while a preferential attachment process gives rise to the scale-free property in networks[25]. In this thesis, we have investigated a process of adaptive rewiring of links as a mechanism to explain the formation of the community structure that is commonly observed in real-world networks.

We have employed the general framework proposed in Ref.[99] for the study of the phenomenon of coevolution in dynamical networks. Coevolution consists of the coexistence of two processes on a network, node state change and rewiring of links between nodes, that can take place with different time scales represented by probabilities P_r and P_c, respectively. A specific coevolutionary model can be expressed by a functional coupling relation $f(P_r, P_c) = 0$. In this framework, the process of adaptive rewiring can be described in terms of two actions: connection and disconnection between nodes, both represented by parameters based on some criteria for comparison of the nodes state variables. In a social context, these actions allow for the description of diverse behaviors such as inclusion-exclusion, homophily-heterophily, and tolerance-intolerance, in terms of probability parameters r and d .

We have shown that the existence of adaptive rewiring is sufficient to induce the formation of community structure in networks, even when the states of the nodes are fixed; i.e., in the absence of node dynamics. A main result of this thesis is that we have successfully characterized the presence of communities in the network by introducing the combined statistical quantity $\Delta Q \times S_m$, corresponding to the product of the modularity change times the average normalized size of the largest subgraph in the system. A $\Delta Q \times S_m > 0$ indicates the formation of a community structure. We have calculated this quantity on the space of the rewiring parameters (d, r). We have shown that community structure arises on a connected network for a range of intermediate values of parameters d and r on this plane. This region, that we have denoted as phase II, separates a region where the network remains random and connected (phase I) from a region where the network is fragmented in small separated components (phase III).

We have also established a relation between the emergence of communities and the density of active links in the network by comparing the quantity $\Delta Q \times S_m$ with the product $S_m \times (1 - \rho)$. Communities are characterized by a low density of active links ρ. A high density of active links is associated with a random network where nodes in different states are randomly connected, while a fragmented network possesses no active links. We consider a main contribution of this thesis the finding of an analytic solution for the evolution of the density of active links in a network, by using a mean field approximation.

The framework of Ref.[99] also allowed us to propose a novel coevolution model where a dynamics of opinion formation for nodes is coupled to an adaptive rewiring process. In this model, the nodes representing social agents have opinions on a discrete-valued scale and they can interact according to a threshold condition. This system of valued opinions is typical of many social, political, and business situations. We showed that our model is quite general and that it contains several previous models as special cases. Our main result has been to show that the adaptive rewiring process still determines the properties of the emerging network in the presence of node dynamics; rewiring parameters values in region II of the (d, r) plane lead to the occurrence of communities, while rewiring parameters in region III of that plane induce the fragmentation of the network. The node dynamics occurring with a probability $P_c = 1 - P_r$ just modulates the outcome determined by the adaptive rewiring process. In this sense, the phase diagram on the space of parameters (d, r) can predict the behavior of the network structure in a coevolutionary dynamical system.

Finally, a useful and important product of this thesis has been the elaboration of our own computer code for simulations of a general model for coevolution of topology and dynamics in networks. This code was run and proven to be efficient on the new High Performance Computing (HPC) cluster of CEDIA.

In all cases considered, we have limited our investigation to the situation where the number of connections in the coevolving network is conserved. This condition is expressed in step 2 of the general coevolution algorithm, where both actions of connection and disconnection always occur. This condition can be generalized by considering different time scales for each of these actions. This will allow for the study of coevolutionary dynamical networks with no conservation of the total number of links.

Other extensions to be investigated in the future include the consideration of variable connection strengths, the inclusion of preferential attachment rules for the connection action, and the influence of different node dynamics, such as differential equations or chaotic elements, on the collective behavior of the system.

Bibliography

[1] T. B. Achacoso and W. S. Yamamoto, *AY's Neuroanatomy of C. Elegans for Computation* (CRC Press, 1991).

[2] J. E. Cohen, F. Briand, and C. M. Newman, *Community food webs: data and theory*, vol. 20 (Springer Science & Business Media, 2012).

[3] R. J. Williams and N. D. Martinez, Nature **404**, 180 (2000).

[4] J. A. Dunne, R. J. Williams, and N. D. Martinez, Proceedings of the National Academy of Sciences **99**, 12917 (2002).

[5] K. W. Kohn, Molecular Biology of the Cell **10**, 2703 (1999).

[6] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, Nature **402**, C47 (1999).

[7] U. S. Bhalla and R. Iyengar, Science **283**, 381 (1999).

[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature **407**, 651 (2000).

[9] M. E. Newman, Proceedings of the National Academy of Sciences **98**, 404 (2001).

[10] P. O. Seglen, Journal of the American Society for Information Science **43**, 628 (1992).

[11] S. Redner, The European Physical Journal B-Condensed Matter and Complex Systems **4**, 131 (1998).

[12] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Computer Networks **33**, 309 (2000).

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos, ACM SIGCOMM Computer Communication Review **29**, 251 (1999).

[14] R. Ding, N. Ujang, H. B. Hamid, M. S. Abd Manan, R. Li, S. S. M. Albadareen, A. Nochian, and J. Wu, Networks and Spatial Economics **19**, 1281 (2019).

[15] A. E. Motter, S. A. Myers, M. Anghel, and T. Nishikawa, Nature Physics **9**, 191 (2013).

[16] G. A. Pagani and M. Aiello, Physica A: Statistical Mechanics and its Applications **392**, 2688 (2013).

[17] J. Abello, A. L. Buchsbaum, and J. R. Westbrook, *A functional approach to external graph algorithms* (1998).

[18] R. Albert and A.-L. Barabási, Reviews of Modern Physics **74**, 47 (2002).

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

[20] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci **5**, 17 (1960).

[21] B. Bollobás and B. Béla, *Random graphs*, 73 (Cambridge University Press, 2001).

[22] S. Janson, T. Luczak, and A. Rucinski, *Random graphs*, vol. 45 (John Wiley & Sons, 2011).

[23] M. Karoński, Journal of Graph Theory **6**, 349 (1982).

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

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

[26] S. Milgram, Psychology Today **2**, 60 (1967).

[27] J. Leskovec and E. Horvitz, IEEE Transactions on Computational Social Systems **1**, 156 (2014).

[28] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, in *Proceedings of the 4th Annual ACM Web Science Conference* (2012), pp. 33–42.

[29] L. A. Adamic, B. A. Huberman, A. Barabási, R. Albert, H. Jeong, and G. Bianconi, Science **287**, 2115 (2000).

[30] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, Proceedings of the National Academy of Sciences **97**, 11149 (2000).

[31] M. E. Newman, S. H. Strogatz, and D. J. Watts, Physical Review E **64**, 026118 (2001).

[32] G. F. Davis and H. R. Greve, American Journal of Sociology **103**, 1 (1997).

[33] G. F. Davis, M. Yoo, and W. E. Baker, Strategic Organization **1**, 301 (2003).

[34] P. Mariolis, Social Science Quarterly **56**, 425 (1975).

[35] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A: Statistical Mechanics and its Applications **311**, 590 (2002).

[36] V. Batagelj and A. Mrvar, Social Networks **22**, 173 (2000).

[37] M. Bordons and I. Gómez, The Web of Knowledge: A festschrift in honor of Eugene Garfield pp. 197–213 (2000).

[38] R. De Castro and J. W. Grossman, The Mathematical Intelligencer **21**, 51 (1999).

[39] J. W. Grossman and P. D. Ion, Congressus Numerantium **108**, 129 (1995).

[40] G. Melin and O. Persson, Scientometrics **36**, 363 (1996).

[41] H. Kautz, B. Selman, and M. Shah, Communications of the ACM **40**, 63 (1997).

[42] S. R. Corman, T. Kuhn, R. D. McPhee, and K. J. Dooley, Human Communication Research **28**, 157 (2002).

[43] L. A. Adamic and E. Adar, Social Networks **25**, 211 (2003).

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

[45] R. Fleischer and C. Hirsch, in *Drawing Graphs* (Springer, 2001), pp. 1–22.

[46] M. Boots and A. Sasaki, Proceedings of the Royal Society of London. Series B: Biological Sciences **266**, 1933 (1999).

[47] M. J. Keeling, Proceedings of the Royal Society of London. Series B: Biological Sciences **266**, 859 (1999).

[48] J. Wallinga, W. J. Edmunds, and M. Kretzschmar, Trends in Microbiology **7**, 372 (1999).

[49] S. N. Dorogovtsev and J. F. Mendes, Advances in Physics **51**, 1079 (2002).

[50] S. N. Dorogovtsev and J. F. Mendes, Physical Review E **63**, 056125 (2001).

[51] S. H. Strogatz, Nature **410**, 268 (2001).

[52] R. Albert, H. Jeong, and A.-L. Barabási, Nature **406**, 378 (2000).

[53] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, Physical Review Letters **85**, 4626 (2000).

[54] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, Physical Review Letters **85**, 5468 (2000).

[55] S. Bornholdt and T. Rohlf, Physical Review Letters **84**, 6114 (2000).

[56] A. Wagner and D. A. Fell, Proceedings of the Royal Society of London. Series B: Biological Sciences **268**, 1803 (2001).

[57] Q. Chen, H. Chang, R. Govindan, and S. Jamin, *The origin of power laws in Internet topologies revisited*, vol. 2 (2002).

[58] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001).

[59] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Structural Analysis in the Social Sciences (Cambridge University Press, 1994).

[60] J. Scott, *Social network analysis: A handbook. 2nd ed, sage publications* (2000).

[61] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, Computer **35**, 66 (2002).

[62] M. Girvan and M. E. Newman, Proceedings of the National Academy of Sciences **99**, 7821 (2002).

[63] P. Holme, M. Huss, and H. Jeong, Bioinformatics **19**, 532 (2003).

[64] R. Guimera and L. A. N. Amaral, Nature **433**, 895 (2005).

[65] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, Proceedings of the National Academy of Sciences **104**, 8685 (2007).

[66] S. Fortunato and C. Castellano, *Community Structure in Graphs* (Springer New York, New York, NY, 2012), pp. 490–512, ISBN 978-1-4614-1800-9, URL `https://doi.org/10.1007/978-1-4614-1800-9_33`.

[67] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Journal of Statistical Mechanics: Theory and Experiment **2008**, P10008 (2008).

[68] M. A. Porter, J.-P. Onnela, and P. J. Mucha, Notices of the AMS **56**, 1082 (2009).

[69] S. Fortunato, Physics Reports **486**, 75 (2010).

[70] L. Donetti and M. A. Munoz, Journal of Statistical Mechanics: Theory and Experiment **2004**, P10012 (2004).

[71] M. J. Barber, Physical Review E **76**, 066102 (2007).

[72] L. Freeman, A Study in the Sociology of Science **1**, 159 (2004).

[73] A. Arenas, A. Fernandez, and S. Gomez, New Journal of Physics **10**, 053039 (2008).

[74] L. Lu and M. Zhang, *Edge Betweenness Centrality* (Springer New York, New York, NY, 2013), pp. 647–648, ISBN 978-1-4419-9863-7, URL `https://doi.org/10.1007/978-1-4419-9863-7_874`.

[75] M. E. Newman and M. Girvan, Physical Review E **69**, 026113 (2004).

[76] J. Reichardt and S. Bornholdt, Physica D: Nonlinear Phenomena **224**, 20 (2006).

[77] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral, Physical Review E **70**, 025101 (2004).

[78] M. G. Zimmermann, V. M. Eguíluz, M. San Miguel, and A. Spadaro, Advances in Complex Systems **3**, 283 (2000).

[79] M. G. Zimmermann, V. M. Eguíluz, and M. San Miguel, Physical Review E **69**, 065102 (2004).

[80] T. Gross and B. Blasius, Journal of the Royal Society Interface **5**, 259 (2008).

[81] T. Gross and H. Sayama, in *Adaptive Networks* (Springer, 2009), pp. 1–8.

[82] J. Ito and K. Kaneko, Neural Networks **13**, 275 (2000).

[83] C. Meisel and T. Gross, Physical Review E **80**, 061917 (2009).

[84] J. Ito and K. Kaneko, Physical Review Letters **88**, 028701 (2001).

[85] P. Gong and C. van Leeuwen, EPL (Europhysics Letters) **67**, 328 (2004).

[86] T. Shibata and K. Kaneko, Physica D: Nonlinear Phenomena **181**, 197 (2003).

[87] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Physics Reports **469**, 93 (2008).

[88] J. Gao, Z. Li, T. Wu, and L. Wang, EPL (Europhysics Letters) **93**, 48003 (2011).

[89] S. Mandrà, S. Fortunato, and C. Castellano, Physical Review E **80**, 056105 (2009).

[90] T. Gross, C. J. D. D'Lima, and B. Blasius, Physical Review Letters **96**, 208701 (2006).

[91] S. Risau-Gusmán and D. H. Zanette, Journal of Theoretical Biology **257**, 52 (2009).

[92] F. Vazquez and D. H. Zanette, Physica D: Nonlinear Phenomena **239**, 1922 (2010).

[93] I. B. Schwartz and L. B. Shaw, Physics **3** (2010).

[94] P. Holme and M. E. Newman, Physical Review E **74**, 056108 (2006).

[95] F. Vazquez, V. M. Eguíluz, and M. San Miguel, Physical Review Letters **100**, 108702 (2008).

[96] D. Centola, J. C. Gonzalez-Avella, V. M. Eguiluz, and M. San Miguel, Journal of Conflict Resolution **51**, 905 (2007).

[97] F. Vazquez, J. C. González-Avella, V. M. Eguíluz, and M. San Miguel, Physical Review E **76**, 046120 (2007).

[98] B. Kozma and A. Barrat, Physical Review E **77**, 016102 (2008).

[99] J. L. Herrera, M. G. Cosenza, K. Tucci, and J. C. González-Avella, EPL (Europhysics Letters) **95**, 58006 (2011).

[100] J. C. González-Avella, M. G. Cosenza, J. L. Herrera, and K. Tucci, EPL (Europhysics Letters) **107**, 28002 (2014).

[101] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Advances in Complex Systems **3**, 87 (2000).

[102] P. Clifford and A. Sudbury, Biometrika **60**, 581 (1973).

[103] R. A. Holley and T. M. Liggett, The Annals of Probability **3**, 643 (1975).

[104] C. Castellano, S. Fortunato, and V. Loreto, Reviews of Modern Physics **81**, 591 (2009).

[105] L. Frachebourg and P. L. Krapivsky, Physical Review E **53**, R3009 (1996).

[106] R. Axelrod, Journal of Conflict Resolution **41**, 203 (1997).

[107] G. Palla, A.-L. Barabási, and T. Vicsek, Nature **446**, 664 (2007).

[108] R. Likert, Archives of Psychology **22**, 1 (1932).

[109] D. Stauffer, A. Sousa, and C. Schulz, Journal of Artificial Societies and Social Simulation **7**, 1 (2004).

[110] A. C. Martins, The European Physical Journal B **93**, 1 (2020).

[111] K. P. Arias, B. Pinilla, and M. G. Cosenza, Anuario Electrónico de Estudios en Comunicación Social "Disertaciones" **13**, 81 (2020).

Glossary of Symbols and Terms

G Number of possible states for a node in a network. 17, 22, 33, 34, 37, 51

N Number of nodes that are part of a network. 2–4, 17, 18, 22, 27, 33, 51

P_c Probability for the node dynamic process to happen. 16, 17, 35, 41

P_r Probability for the rewiring process to happen. xv, 16, 17, 34–37, 39–41

S_g Normalized size of the largest domain in a network. xv, 18, 36–38

S_m Normalized size of the largest component in a network. xv, 18, 28, 29, 39, 40

ΔQ Modularity change. xv, 28–30, 39

$\langle k \rangle$ Average number of neighbors that a node has in a network. 17, 22, 33, 34

ρ Density of active links. xv, 23, 26, 27, 29, 41

ρ^* Analytic stationary density of active links. xv, 23–27, 29

d Probability that two nodes in identical states become disconnected. 16, 22, 26, 29, 30, 34, 35, 41

r Probability that two nodes in identical states become connected. 16, 22, 26, 27, 29, 30, 34, 35, 41

u Normalized threshold value for the discrete-valued opinion model. xv, xvi, 34, 36, 38–40, 51

active link Link that connects two nodes with different states. xi, xv, 12, 13, 22, 23, 25–27, 29, 30, 41

clustering coefficient Fraction of the actual links present in the network compared to all-to-all coupling. xv, 4, 21

community Groups of nodes that have a high number of links within the groups and low number of links outside them. xi, 7–10, 12, 18, 21–23, 27–31, 33, 39–42

complex system Set of interacting elements whose collective behavior cannot be derived from the knowledge of the properties of the isolated element. 1, 12, 15

degree Number of links that a node shares. 2, 4–7, 9, 17, 18, 22, 33, 34, 50

domain Set of connected nodes on a network that share the same state variable. xv, 18, 19, 30, 36–39

geodesic distance Shortest number of links that separates one node from another in a network. xv, 3, 4, 9, 10

graph Visual representation of a network. 1, 2, 4, 11

inert link Link that connect two nodes with the same states. 23, 30

modularity quantity that evaluated the quality of a network partition by comparing the number of links inside the modules of a partition against the expected number of links of a null model. xi, 11, 13, 18, 28–30, 39, 41

network Set of elements (known as nodes or vertices) with relations (also know as links or edges) among them. xi, xv, 1–13, 15–18, 21–23, 25, 27–31, 33, 36–42, 49–51

null model Random networks with the same size and degree as a network with certain structure. 7, 11

subgraph Set of connected nodes in a network also known as a component of a network. xv, 18, 19, 27–30, 36, 39–41

Appendix A

Computer code used for the coevolution of topology and dynamics in networks

We have developed our own computer code for the coevolution of topology and dynamics in networks used in this work. This code was written in the *C* programming language. The following program creates an Erdös–Rényi random network with N nodes, whose states are uniformly and randomly assigned from the G possible options. Then, the program applies the coevolution dynamics according to the five parameters P_r, P_c, r, d and u. All the parameters, along with the size of the system, the number of possible states, the number of realizations and the number of iterations of the dynamic, are defined in the code. The outputs of the program are four files. The first two files correspond to the link representation of the network and to the states of the nodes before the dynamics is applied. The last two files correspond to the link representation of the network and the states after the dynamics is applied. All the statistical quantities calculated in this thesis were performed over those four files.

```c
#include <math.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <pthread.h>
#include <stdbool.h>
#include<unistd.h>
#include <gsl/gsl_rng.h>
#define MAX_THREADS 64

/*This function returns a random int between 0 and RAND_MAX*/
unsigned long int rand_int(void *ptr){
    gsl_rng *rng_r=(gsl_rng*)ptr;
    return gsl_rng_get(rng_r);
}

/*This structures contains all the variables of the system*/
struct Inputs{
    int N;
    int time;
    int k;
    int G;
    int runs;
    int size;
    float d;
    float r;
```

```
    float U;
    float Pr;
    float Pc;
    unsigned int thread_id;

};

/*This function saves a file with the all the links of the network, the char name[] is the name of the file*/
void print_links(int *matrix, int N,char name[]){
    FILE *fptr;
    fptr=fopen(name,"w");
    fprintf(fptr,"%d\n",N);
    for (int i=0;i<N;i++){
        for (int j=i+1;j<N;j++){
            if (*(matrix+i*N+j)==1){
                fprintf(fptr,"%d %d\n",i,j);
            }
        }
    }
    fclose(fptr);
}


/*this function is does the rewiring process of disconnection and reconnection*/
void rewire(float d, float r,int *matrix, double *list_states, int N, int node,void *ptr_rng_r){
    gsl_rng *rng_r=(gsl_rng*)ptr_rng_r;

    /*Here we find all the neighbors of our node*/
    int nodes_same_state=0;
    int nodes_different_state=0;
    int index_same_state[N];
    int index_different_states[N];

    int disconnect_index=-1;
    int reconnect_index=-1;

    for (int i=0;i<N;i++){
        if (*(matrix+node*N+i)==1){
            if (*(list_states+node)==*(list_states+i)){
                index_same_state[nodes_same_state]=i;
                nodes_same_state++;

            }
            else{
                index_different_states[nodes_different_state]=i;
                nodes_different_state++;

            }
        }
    }

    /*Here we choose the neighbor that will be disconnected  from our node*/
    double d_rand=(double)rand_int(rng_r)/gsl_rng_max(rng_r);

    if (d_rand<=d){
        if (nodes_same_state>0){
```

```
            int temp=rand_int(rng_r)%nodes_same_state;
            disconnect_index=index_same_state[temp];
        }
    }
    else{
        if (nodes_different_state>0){
            int temp=rand_int(rng_r)%nodes_different_state;
            disconnect_index=index_different_states[temp];
        }
    }

    /*Here we find all the nodes in the network that are not connected to our node*/
    int nodes_same_state_disconnected=0;
    int nodes_different_state_disconnected=0;
    int index_same_state_disconnected[N];
    int index_different_states_disconnected[N];

    for (int i=0;i<N;i++){
        if (*(matrix+node*N+i)==0 &&node!=i){
            if (*(list_states+node)==*(list_states+i)){
                index_same_state_disconnected[nodes_same_state_disconnected]=i;
                nodes_same_state_disconnected++;
            }
            else{
                index_different_states_disconnected[nodes_different_state_disconnected]=i;
                nodes_different_state_disconnected++;
            }
        }
    }

    /*Here we choose the node to connect to */
    double r_rand=(double)rand_int(rng_r)/gsl_rng_max(rng_r);

    if(r_rand<=r){
        if (nodes_same_state_disconnected>0){
            int temp_node=index_same_state_disconnected[rand_int(rng_r)%
    nodes_same_state_disconnected];
                reconnect_index=temp_node;
        }
    }
    else{
        if (nodes_different_state_disconnected>0){
            int temp_node= index_different_states_disconnected[rand_int(rng_r)%
    nodes_different_state_disconnected];
                reconnect_index=temp_node;
        }
    }

    /*In order to maintain the number of links constant we do the connection and reconnection
    together only if it is possible*/
    if (reconnect_index!=-1 && disconnect_index!=-1){
            *(matrix+node*N+disconnect_index)=0;
            *(matrix+disconnect_index*N+node)=0;
            *(matrix+node*N+reconnect_index)=1;
            *(matrix+reconnect_index*N+node)=1;
    }
}
```

```c
/*this function saves a file with the states of the nodes. The first value corresponds to the node 1, the second to the node 2 and so on*/
void print_states(double *list_states,int N,char name[]){
    FILE *fptr;
    fptr=fopen(name,"w");
    for (int i=0;i<N;i++){
        fprintf(fptr, "%lf ",*(list_states+i));
    }
    fclose(fptr);
}


/*This function saves the adjacency matrix of the network*/
void print_matrix(int *matrix, int N,char name[],char mode[]){

    FILE *fptr;
    fptr=fopen(name,mode);

    for (int i=0;i<N;i++){
        for (int j=0;j<N;j++){
      fprintf(fptr,"%d ",*(matrix+i*N+j));
        }
        fprintf(fptr,"\n");
    }

    fclose(fptr);
}


/*This function is in charge of the node dynamic, in this case a discrete-valued opinions model*/
void node_dynnamics(float U,int *matrix, double *list_states, int N, int node,void *ptr_rng_r){
    gsl_rng *rng_r=(gsl_rng*)ptr_rng_r;

    int number_neighbors=0;
    int index_neighbors[N];

    for (int i=0;i<N;i++){
        if (*(matrix+node*N+i)==1){
            index_neighbors[number_neighbors]=i;
            number_neighbors++;
        }
    }
    /*Here we compare the states of the nodes and apply the node dynamic*/
    if (number_neighbors>0){
        int node_2=rand_int(rng_r)%number_neighbors;
        if(fabs((double)(*(list_states+node)-*(list_states+index_neighbors[node_2]))) <=U ){
            *(list_states+node)=*(list_states+index_neighbors[node_2]);
        }
    }

}

/*This function does the time evolution of the network*/
void evolution(float Pc, float Pr, float U,float d, float r, int *matrix, double *list_states, int N,int n_evol,void *ptr_rng_r){
    gsl_rng *rng_r=(gsl_rng*)ptr_rng_r;
```

```

    for (int t_step=0;t_step<n_evol;t_step++){

        int node=rand_int(rng_r)%N;
        /*Here we check if the choose node has neighbors for the dynamic*/
        int n_neighbors=0;
        for (int i=0;i<N;i++){
            if (*(matrix+node*N+i)==1){
                n_neighbors++;
            }
        }

        if (n_neighbors>0){

            double temp_Pr=(double)rand_int(rng_r)/gsl_rng_max(rng_r);
            if (temp_Pr<=Pr){
                rewire(d,r, matrix,list_states, N,node, rng_r);
            }
            double temp_Pc=(double)rand_int(rng_r)/gsl_rng_max(rng_r);
            if (temp_Pc<=Pc){
                node_dynnamics(U,matrix, list_states, N, node,rng_r);
            }
        }
    }
}

/*this function calculates and the average degree of the network*/
float average_k(int *matrix,int N){
    int sum_k=0;
    for (int i=0;i<N;i++){
        for (int j=0;j<N;j++){
            if(*(matrix+i*N+j)==1){
                sum_k++;
            }
        }
    }
    return (float)sum_k/N;
}


/*this function does all the runs with the fixed variables*/
void section(void *ptr){
    struct Inputs *inputs=(struct Inputs*)ptr;

    int N=(*inputs).N;
    int time=(*inputs).time;
    int k=(*inputs).k;
    int G=(*inputs).G;
    int runs=(*inputs).runs;
    int size=(*inputs).size;
    float d=(*inputs).d;
    float r=(*inputs).r;
    float U=(*inputs).U;
    float Pr=(*inputs).Pr;
    float Pc=(*inputs).Pc;
    unsigned int thread_id=(*inputs).thread_id;
    const gsl_rng_type * rng_T;
    gsl_rng *rng_r;
```

```
    gsl_rng_env_setup();
    rng_T = gsl_rng_rand48;
    rng_r = gsl_rng_alloc(rng_T);
    gsl_rng_set(rng_r,thread_id);

    int T_steps=N*time;

    /*here all the tuns are done*/
    for (int run=0;run<runs;run++){

            char info_1[20];
            sprintf(info_1,"Pr_%d_run%d",thread_id,run);
            /*here we created the matrix for the network and the list for the states*/
            int *matrix;
            double *list_states;

            /*Here we calculate the probability to created a Erdos Reyni random network as function of p*/
            double p=(double)k/(N-1);

            matrix=(int*)calloc(N*N,sizeof(int));
            list_states=(double*)calloc(N,sizeof(double));

            /*Here we give the nodes a discrete state chose randomly  from the G possible options*/
            for (int i=0;i<N;i++){
                    *(list_states+i)=(double)(rand_int(rng_r)%G)/G;
                }

            char name_states_initial[50]="states_i_";
            char name_states_final[50]="states_f_";

            strcat(name_states_initial,info_1 );
            strcat(name_states_final,info_1 );

            /*In this part we fill the adjacent matrix depending of p*/
            for (int i=0;i<N;i++){
                for (int j=i+1;j<N;j++){
                    double r=(double)rand_int(rng_r)/gsl_rng_max(rng_r);
                    if (r<=p){
                        *(matrix+i*N+j)=1;
                        *(matrix+j*N+i)=1;
                    }
                    else{
                        *(matrix+i*N+j)=0;
                        *(matrix+j*N+i)=0;
                    }
                }
            }
            char name_links_initial[50]="links_i_MD_";
            char name_links_final[50]="links_f_MD_";

            strcat(name_links_final,info_1);
            strcat(name_links_initial,info_1);

            /*Here we save the network and states before the evolution*/
            print_links(matrix, N, name_links_initial);
            print_states(list_states,N,name_states_initial);
```

```
            /*Here we do the time evolution of the network*/
            evolution(Pc,Pr,U,d,r, matrix,list_states, N,T_steps,rng_r);

            /*Here we save the network and states after the evolution*/
            print_links(matrix, N, name_links_final);
            print_states(list_states,N,name_states_final);


            free(matrix);
            free(list_states);
        }

    gsl_rng_free (rng_r);
    printf("Thread id: %d out of %d threads done\n",thread_id,MAX_THREADS);

}

/*The main function is in charge the parallelize the code so it can be run with multiple
    threads*/
void main(){

    /*size and definition correspond to the number of data points wanted*/
    struct Inputs inputs[MAX_THREADS];
    int size=MAX_THREADS;
    float definition=(float)1/size;

    /*Here there are the parameter that are going to be send to each thread*/
    /*Here, the variables that must change can be set to different values for example Pr and
    Pc in this case*/
    for (int thread=0;thread<MAX_THREADS;thread++){
        inputs[thread].N=3200;
        inputs[thread].time=10000;
        inputs[thread].k=4;
        inputs[thread].G=320;
        inputs[thread].runs=100;
        inputs[thread].size=size;
        inputs[thread].d=0.35;
        inputs[thread].r=0.65;
        inputs[thread].U=1.0;
        inputs[thread].thread_id=(unsigned int)thread;
        inputs[thread].Pr=definition*thread;
        inputs[thread].Pc=1-definition*thread;
    }

    /*Here, all the threads are created and run*/
    pthread_t thread_ids[MAX_THREADS];
    for (int thread=0;thread<MAX_THREADS;thread++){
        pthread_create(&thread_ids[thread],NULL,(void*)&section,(void*)&inputs[thread]);

    }

    /*Here, all the threads are joined*/
    for (int thread=0;thread<MAX_THREADS;thread++){
        pthread_join(thread_ids[thread],NULL);
    }

}
```