UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

INSTITUTO DE INVESTIGACIÓN EN COMUNICACIÓN ÓPTICA

P H D T H E S I S

to obtain the title of

PhD in Applied Sciences

Defended by

José de Jesús Esquivel Gómez

Impact of several microscopic processes in the growth and evolution of Complex Networks

Thesis Advisor: J. Jesús Acosta Elías

defended on January 15, 2016

Jury :

Advisor :	Dr. J. Jesús Acosta Elías
Examiners :	Dr. Raúl E. Balderas Navarro
	Dra. Marcela Mejía Carlos
	Dr. José Salomé Murguía Ibarra
	Dr. Eric Campos Cantón
	Dr. Isaac Campos Cantón
External Examiner :	Dr. Juan Gonzalo Barajas Ramírez

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Dr. J. Jesús Acosta Elías for the continuous support during my PhD study. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Raúl E. Balderas Navarro, Dra. Marcela Mejía Carlos, Dr. José Salomé Murguía Ibarra, Dr. Eric Campos Cantón and Dr. Isaac Campos Cantón, for their valuable comments during my PhD studies.

My sincere thanks to Dr. Raúl E. Balderas Navarro, for their valuable support and friendship.

I would like thank to my external examiner Dr. Juan Gonzalo Barajas Ramírez for the valuable discussions that helped me understand my research area better.

Last but not least, I would like to thank my wife Mary and my children Mauricio and Rodrigo for your patience and support.

Contents

Chapter 1

Abstract

Networks are present in many aspects of our daily lives. For example, Communication Networks as telephone networks, Social Networks as Facebook and Twitter, airline networks, road networks, Internet, and WWW. The networks can be modeled using the tools of the graph theory. For example in a network of papers citations, the vertices are the papers and the edges the citation between them; in a network of web pages, the vertices are the web pages and the edges are the hyperlinks pointing from one page to another; and similarly for friendship networks, epidemic networks, communication networks, etc. The Real networks are commonly termed Complex Networks because have been demonstrated that they have properties more complex than classical random graphs.

One motivation for study networks, is to decipher the local processes that originate a particular behavior between its components and to predict wanted or unwanted effects. For example, it would be important to predict how quickly an epidemic evolves and determinate how the mechanisms of the network can be used to stop or eradicate it.

With the aim of to decipher the local processes that originate the topological and dynamical properties of Complex Networks, in the literature can be found many growth and evolution models. However, at this time do not exists a general model of network growth that, with the incorporation of the appropriate processes, to be able to reproduce the properties found in real-world complex networks. This is due to in the growth and evolution of complex networks exists unknown process that shape the topological and dynamical properties of this class of networks.

In this Thesis, is investigated the impact that some local processes have in the topological properties of Complex Networks. Also are proposed five growth models that reproduce some properties founded in real Complex Networks.

Chapter 2

Introduction

Contents

2.1 Networks

A network in general is any system that admits an abstract mathematical representation as a graph whose nodes (vertices) identify the elements of the system and in which the set of connecting links (edges) represent the presence of a relation or interaction among those elements. [1] Because of this, the mathematical tools used in graph theory are suitable for the study of networks.

2.1.1 Directed and undirected Networks

Networks can be classified as directed or undirected. In directed networks, each link has an origin node and a destination node, the number of incoming links and outgoing links that a node n_i has is called in-degree ($k_{in(i)}$) and out-degree ($k_{out(i)}$) respectively (see Figure 2.1a). In undirected networks, if one node is already connected to other node, then the reverse link is regarded as a new link and it is then said that both nodes have a link and the number of links that a node n_i has is called degree k_i (see Figs. 2.1b).

2.1.2 Some properties of Networks

Some important properties that describe the topology of a network are:

- ***Degree distribution:*** This property describes the probability that a node randomly selected from a network has a certain number of links. In undirected networks, $P(k)$ describes the probability that a node n_i from the network has k links. In directed networks, $P(k_{in})$ and $P(k_{out})$ describe the probability that a node n_i from the network has k_{in} and k_{out} links respectively.

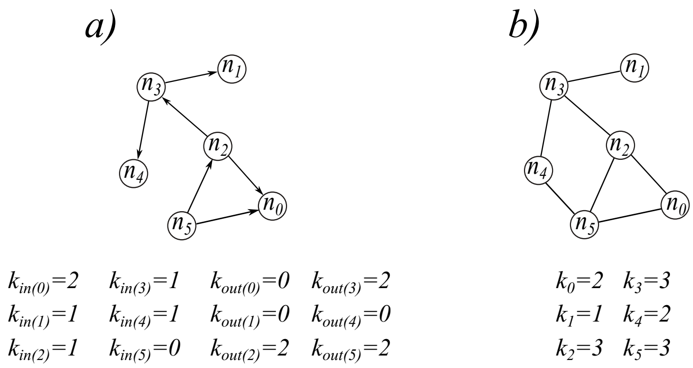

Figure 2.1: In the figure are shown two network examples, where the circles represent the nodes and the lines and arrows to the links. a) A directed network comprising six nodes $n_{0...5}$ with the in-degree and out-degree values for each one node. b) An undirected network comprising six nodes $n_{0...5}$ with the degree values k_i.

- ***Clustering coefficient:*** This property describes the fraction of links among the neighbors nodes of a particular node. The Clustering C_i of a node n_i in the network is obtained as follow:

  - in directed networks

$$C_i = \frac{\varepsilon}{\beta(\beta-1)},$$

  - in undirected networks

$$C_i = \frac{2\varepsilon}{\beta(\beta-1)},$$

  where β is the number of neighbor nodes that n_i has and ε is the number of links among the neighbor nodes of n_i. In order to shed light how the Clustering is calculated see Figs. 2.2a and 2.2b. Finally, Clustering C of the network is the average:

$$C = \frac{1}{N}\sum_{i=0}^{N} C_i$$

  where N is the number of nodes in the network.

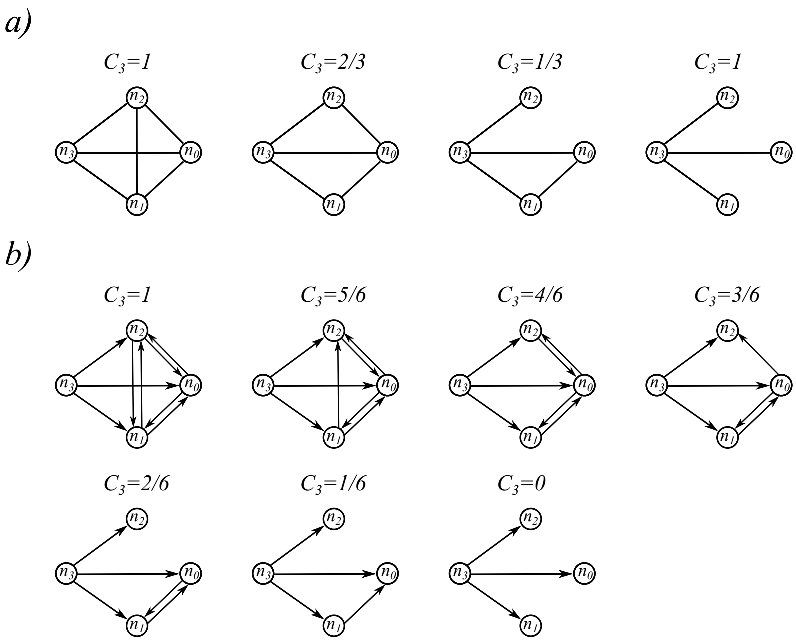

Figure 2.2: Undirected network *a)* and directed network *b)* comprising four nodes $n_{0\ldots3}$, it is shown how the local clustering C_3 of node n_3 decreases as the number of links between his neighbor nodes decreases.

- ***Shortest Path Length:*** The minimum quantity of links for to go from a node n_i to another node n_j in the network is called the shortest path length L_{ij} (see Fig. 2.3). When not exist a possible path between two nodes n_i and n_j, it is said that $L_{ij} = \infty$.
- ***Diameter:*** The greatest L_{ij} from the all possibles L_{ij} in the network is called Diameter of the Network (see Fig. 2.3).

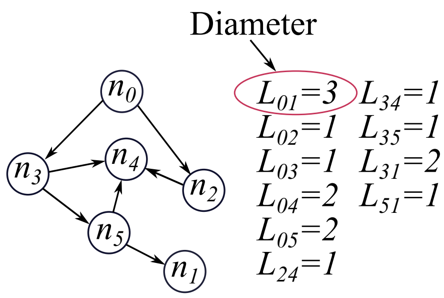

Figure 2.3: Directed network comprising six nodes $n_{0\ldots5}$. In the figure are shown the possible L_{ij} values and the Diameter of the network.

- ***Island Size distribution (I_s):*** A network may consist of several islands (usually called clusters or components), where an island is a set of nodes which is not connected to the rest of the network as depicted in Fig. 2.4. The number of islands with a certain size is described by the Island Size distribution. In a network comprised by islands, the island with the greater quantity of nodes is called giant island.

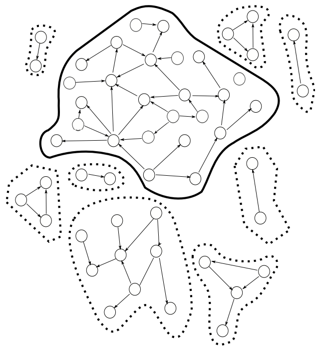

Figure 2.4: Directed network comprising nine islands. In the figure are shown eight islands (enclosed with dotted line) and the giant island (enclosed with solid line).

2.2 Network Study

From 1959, it was assumed that real networks could be modeled as random networks using the well known random graph model proposed by the Hungarian Mathematicians Paul Erdös and Alfréd Rényi [2](ER model). That is, was supposed that real networks could be properly modeled connecting their nodes by links randomly placed. In the ER model it is assumed that initially the network is composed by N isolated nodes and at each time step two different nodes are selected randomly and linked with probability p (with $p > 0$). An important property of networks generated with this model is that have degree distributions that follow a Poisson distribution.

In the late 90's, several publications showed that real-world networks such as small metabolic networks [3], scientific collaborations networks [4] or large informatics networks (*i.e.*, the Internet [5] or WWW [6, 7]) exhibit topological properties different from those found in random networks. For example, the degree distribution $P(k)$ between their nodes decay as a power law $P(k) \sim k^{-\gamma}$ [8], have high clustering coefficient and small diameter, that is, are *small world* networks [9].

This new class of networks are termed complex networks (CN) due to their properties, [10] which suggest that a random model is not suitable for their study. That is, the real networks are more complex than classical random graph [2]. Several examples of real complex networks are shown in Table 2.1.

Table 2.1: List of some real-world complex networks with their corresponding type (d:directed,u:undirected), number of nodes, clustering (C), the average shortest path length (SPL), diameter (D), exponent γ_{in} of the in-degree distribution ($P(k_{in})$) and the exponent γ_{out} of the out-degree distribution ($P(k_{out})$). More examples of complex networks can be found in Refs. [11, 12, 13]

Network	**Type**	**Number of nodes**	**C**	**SPL**	**D**	γ_{in}	γ_{out}
WWW [14]	d	$325,729$	0.087	11.2	46	2.1	2.45
U.S. patents [15]	d	$3,774,324$	0.067	8.24	26	-	-
Internet, AS. [16]	u	$34,761$	0.0485	3.78	10	1.92	1.92
Network of flights between airports of the world. [16]	d	$2,939$	0.25	4.18	14	1.74	1.74
Actor colaboration. [16]	u	$382,219$	0.16	3.7	13	2.13	2.13
User Friendship Youtube. [16]	u	$1,134,890$	0.0062	5.55	24	2.14	2.14
Network of protein interactions. [16]	u	$1,870$	0.055	7.07	19	3.04	3.04
Flickr Social network. [16]	d	$2,302,925$	0.108	5.46	23	1.71	1.71

In the growth of real CN exists local process that shape the topological properties of these networks. For example, in the WWW network links are not static and, at any time, a node (web page) may lose a connection to another node (deleting a hyper-link) and add this same connection to a different node (a rewiring process), new links can appear in the network (links added), also nodes can be dead (deleting nodes) and other unknown local process. These local processes are present in other real networks as Social networks. On the other hand, in other real networks as the paper citation networks the mentioned local processes are not present, that is, the papers not dead, the cites between papers are not rewired (links are static), new cites between old papers can not be appear, but in this type of networks there are other local processes that shape the topological properties of this type of networks. Although many real networks do not share the same local processes, they have very similar topological properties, for example the distribution of links between their nodes follows a power law.

With the discovery of CN the challenge that exists up today aims to develop a general model of network growth that, with the incorporation of the appropriate processes, would be able to reproduce the properties found in real-world CN. In this regard, a lot of models have been proposed so far [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]. Some of them are described in Chapter 3. In this Thesis, is investigated the impact that some local processes have in the topological properties of Complex Networks. In particular, is analyzed the effect that the prohibition of multiple links and its combination with other local process have in network properties as the in-degree distribution, Clustering and shortest-path. Also are present two models capable of generate out-degree distributions that follows a power law, and one model capable of generate island size distributions that decay as a power law.

This thesis is organized as follows. In Section 3 are described some growth models for Complex Networks proposed previously. In Section 4 is developed a growth model that incorporates the multiple links prohibition process. A proposed growth model that incorporates in joint the internal links, rewiring and multiple links prohibition is presented in Section 5. Two growth models capable to generate out-degree distributions that decay as a power law are present in Section 6. In Section 7, a growth model capable to obtain Island Size and In-degree distributions with power-law behavior is present. Finally, Discussion and Conclusions are given in sections 8 and 9 respectively.

On the other hand, it is important to mention that with the models proposed in Sections 4,5 and 6, four research papers were published [27, 28, 29, 30].

Chapter 3

Related work

With the aim of reproducing the properties found in real networks, a lot of models of network generation have been proposed, some of these models are described below.

3.1 Barabási-Albert Model (Preferential Attachment)

In 1999, Barabási and Albert (BA) [17] proposed a growth model for undirected CN. In this model is introduced in first time the preferential attachment concept which assumes that the probability for a node n_i gain new links is directly proportional to the amount of links that n_i has. That is, the BA model is based on the *rich-get-richer* approach.

In the BA model, the growth of the network is by node addition and preferential attachment: initially, there are m_0 nodes and as time evolves a new node is added with $m \leq m_0$ links. The probability $\prod_{BA}$ in which a new node is linked to node n_i in the network, is proportional to the degree k_i of node n_i given by:

$$\prod_{BA}(k_i) = \frac{k_i}{\sum_j k_j}. \tag{3.1}$$

In particular, the BA model is able to obtain degree distributions $P(k)$ that decay as a power law $P(k) \sim k^{-\gamma}$, however it yields a fixed exponent $\gamma = 3$ [17] as depicted in Fig. 3.1. This contrasts to the values of γ found in several real-world CN, which range $1.05 < \gamma < 8.94$ [8, 11, 16, 31].

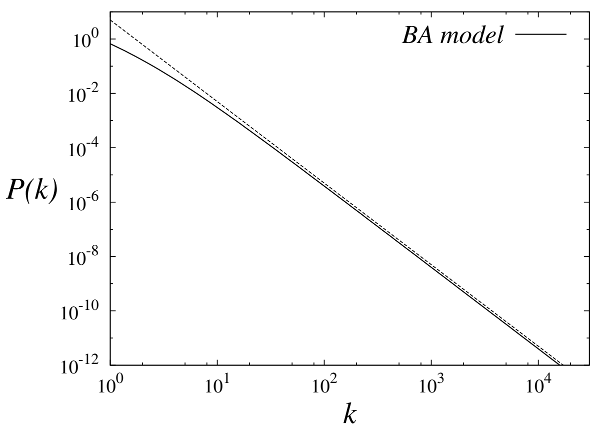

Figure 3.1: In the figure, the solid line represents the degree distribution generated by the BA model and the dashed line is a power law with exponent $\gamma = 3$.

3.2 Initial Attractiveness

In 2000, Dorogovtsev, Mendes and Samukhin [18] proposed an alternative growth model for directed CN. In the rest of the thesis we refer to it as DMS model. The DMS model is mainly based on the BA model with two differences:

1. All the nodes born with a same initial attractiveness A.
2. At each time step a new node and simultaneously m directed links are added to the network. Such links can come from any of the existing nodes (*i.e.*, they may come out from the new node, from old nodes, or even from outside of the network).

Furthermore, in this model the probability $\prod_{DMS}$ that a node n_i in the network gets a link is proportional to both $k_{in(i)}$ and A, as stated in Eq. 3.2.

$$\prod_{DMS}(k_{in(i)}) = \frac{k_{in(i)} + A}{\sum_j (k_{in(j)} + A)}, \tag{3.2}$$

In particular, with this model it is possible to obtain In-degree distributions that follows a power law of the form $P(k_{in}) \sim k_{in}^{-\gamma_{in}}$ with

$$\gamma_{in} = 2 + \frac{A}{m}. \tag{3.3}$$

That is, this model is capable to obtain exponents spanning $2 < \gamma_{in} < \infty$ for the in-degree distribution.

3.3 Nonlinear preferential attachment

In 2000, Krapivsky, Redner, y Leyvraz [19] proposed a growth model for directed CN. This model consider the network growth as follow: at each time step a new node n_{new} with one outgoing link ($m = 1$) is added to the network and links to a node n_i already present in the network with probability

$$\prod(k_i) = \frac{k_i^{\alpha}}{\sum_j (k_j^{\alpha})}, \tag{3.4}$$

where k_i is the sum of $k_{out(i)}$ and $k_{in(i)}$ of n_i.

In particular, in this model is investigated the effect that a nonlinear preferential attachment have in the degree distribution. They found that:

1. For $\alpha < 1$, $P(k)$ follows an exponential distribution.
2. For $\alpha > 1$ nearly all nodes are connected to a same node. When $\alpha > 2$ and considering that the network starts with a single node (n_{root}), all sites are connected to n_{root}, thus the network becomes to be a star graph and $P(k) = \delta_{N-1}(k)$ where N is the number of nodes in the network.

3. For $\alpha = 1$, $P(k)$ decay as a power law with exponent $\gamma = 3$.
4. In the $lim_{\alpha \to 1}$, $P(k)$ decay as a power law with exponent $2 < \gamma < 3$ and $3 < \gamma < \infty$.

That is, they found that the scale-free nature of the network is present only when the preferential attachment is asymptotically linear. In this case the exponent of the degree distribution can be tuned to any value between 2 and ∞.

3.4 Copying

In 2005, Krapivsky and Redner [20] proposed a growth model for directed CN. In this model is introduced in first time the copying process. In this model, the growth of the network is by node addition and copying links. That is, at each time step a new node n_i is added to the network, n_i selects a target node randomly and links to it, as well as to all ancestor nodes (see Fig. 3.2).

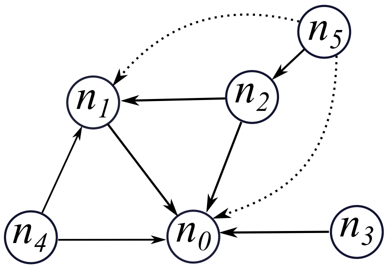

Figure 3.2: Directed network comprising six nodes $n_{0...5}$. The figure shows the the copying process, the node n_5 is added and randomly selects the node n_2 and connects to it and to nodes n_0 and n_1 (dotted arrows).

This model is able to obtain in-degree distribution that follows a power law ($P(k_{in}) \sim k_{in}^{-\gamma}$) with a fixed exponent $\gamma = 2$, and out-degree distribution following a Poisson distribution.

3.5 Accelerated Growth

In real networks as Internet [5], WWW [32], and co-authorship network [33] the average degree increases over the time, that is the number of links in the network increases more rapidly than the number of nodes. This phenomenon is called accelerated growth.

The impact of accelerated growth on the in-degree distribution was investigated by Dorogovtsev and Mendes [21], they proposed a growth model for directed CN that incorporates the accelerated growth mechanism in the network evolution. The model is as follow: at each time step a new node n_{new} is added to the network and receives β incoming links from random nodes in the network. In the same time step, $c_0 t^{\theta}$ ($c_0 > 0$ and $0 < \theta < 1$) links are distributed on the network, each one of these links starts in a randomly selected node and finalizes in a node n_j according to the next probability:

$$\prod(k_{in(j)}) = \frac{k_{in(j)} + A}{(k_{in(j)} + A)},$$

which represents the preferential connection. This model is able to obtain in-degree distribution with power law behavior $P(k_{in}) \sim k_{in}^{\gamma}$ with:

$$\gamma = 1 + \frac{1}{1+\theta}.$$

3.6 Rewiring

In several real networks the links are not static. That is, a node n_i connected to other node n_j can disconnect from this and connect to other node n_k. This process is called Rewiring, and is present in real networks as Internet [5] and WWW [32] for example.

The impact that the rewiring have in the degree distribution $P(k)$ was investigated by Albert and Barabási [25], they proposed a model that incorporates jointly the rewiring and addition of links in the network. The model follows the next rules:

1. with probability p, m ($m < m_0$) links are added to the network. For each one link, one end is attached to a randomly selected node and the other end is attached to a node n_i with probability:

$$\prod(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)}. \tag{3.5}$$

2. With probability q, m links are rewired. For each rewire, one link l_{ij} (link that connects to nodes n_i and n_j) selected randomly is removed and a new link $l_{ij'}$ is created accordingly to $\prod(k_{j'})$ (see Eq. 3.5).

3. With probability $1 - p - q$ a new node n_{new} is added to the network and connects to m nodes presents in the network accordingly to Eq. 3.5.

Albert and Barabási found that the exponent γ of the degree distribution $P(k) \sim k^{-\gamma}$ changes with p, q and m, covering a range of exponents from $\gamma = 2$ to ∞. In particular, they found that their model is capable of reproduce the connectivity distribution of movie actors. [25]

3.7 Internal edges and edge removal

Dorogovtsev and Mendes [26], investigated the effect that both addition and dead of links have in the degree distribution. They proposed a growth model consisting in the next rules: at each time step a new node is added to the network, at the same time c links are added between pairs of unconnected nodes n_i and n_j with probability proportional to the product of their degrees $k_i \cdot k_j$. Additionally, c links between old nodes are removed with equal probability.

They found that their model is able to generate degree distributions $P(k)$ with power-law behavior $P(k) \sim k^{-\gamma}$, where

$$\gamma = 2 + \frac{1}{1+2c}.$$

3.8 Aging and Cost

Amaral *et al*. [22], studied the effect that the aging and cost have in $P(k)$. In their model, aging refers to the phenomenon in which the younger nodes have higher probability than old nodes to obtain new links and cost is defined as the the limit of links that each one node can to have. The model evolves following growth and preferential attachment as the BA model, but when a node reaches a certain age or has more than a critical number of links, new links cannot connect to it.

Using numerical simulations, they found that the power law behavior of $P(k)$ becomes to disappear as the nodes age increases rapidly and when the capacity of links for each one node is small.

3.9 Gradual Aging

In 2000 Dorgovtsev and Mendes [34], investigated the effect that the gradual aging of the nodes have in $P(k)$. Their model is by node addition one at each time step t and preferential attachment. They propose that the probability $\prod$ for a node n_i (added at the time step t_i) gain new links should depend of both his degree and his age, as:

$$\prod = \frac{k_i \tau_i^{-\nu}}{\sum_j k_j \tau_j^{-\nu}}. \tag{3.6}$$

where $\nu \geq 0$ and $\tau_i = t - t_i$. They found that the model is able to generate distributions of $P(k)$ with power law behavior only when $\nu < 1$ with exponent $3 < \gamma < \infty$.

Chapter 4

The impact of multiple links prohibition in Directed Complex Networks

In some models as the proposed by Barabási *et.al.* [17, 25], Amaral *et.al.* [22] and Dorogovtsev *et.al.* [18, 21, 26, 34] are allowed multiple links, that is a node n_j could have more than one link from or to a same node. In contrast, several real CN do not have multiple links. For example, in a paper citation network an article in its reference section does not have two identical references, in a friendship network not exist more than once friendship bond between two individuals. The lack of multiple links in such networks suggest that the growth and evolution models of CN should consider such a feature in order to properly describe the topological properties of this class of networks. With this idea, is proposed a new growth model for directed CN based on the DMS model (see Initial Attractiveness in Chapter 3). The proposed model prohibits multiple links between pairs of nodes, and it is designated as multiple links free (MLF) model.

This Chapter is organized as follows. Section 4.1 considers the DMS model by taking into account directed multiple links. Section 4.2 outlines the features of the MLF model. The experiment details and results are shown in Section 4.3. The analytical considerations for the MLF model are presented in Section 4.4. Finally, section 4.5 demonstrates that the MLF model is able to reproduce some topological properties of a real network.

4.1 The DMS model and directed multiple links

As it described in Chapter 3, in DMS model [18] the growth of the network is by node addition and preferential attachment. That is, at each time step t a new node and simultaneously m directed links are added to the network. Such links can come from any of the existing nodes (*i.e.*, they may come out from the new node, from old nodes, or even from outside of the network). Each one link is connected to a node n_i with probability:

$$\prod_{DMS}(k_{in(i)}) = \frac{k_{in(i)} + A}{\sum_j (k_{in(j)} + A)}, \tag{4.1}$$

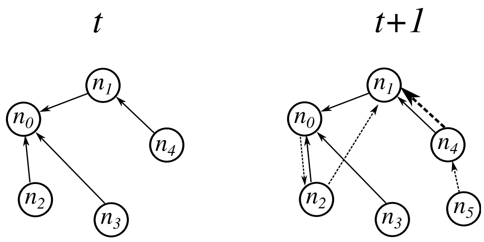

Figure 4.1: An example of a directed network according to the DMS model. This particular network comprises five nodes $n_0, ..., n_4$ at time t. At time step $t+1$, node n_5 is added and simultaneously $m = 4$ directed links (indicated by dashed arrows) are also added. Note, in this case, that a double link between n_4 and n_1 was generated.

where $k_{in(i)}$ and A are the in-degree and initial attractiveness of n_i respectively. A is the same for all the nodes.

It is important to mention that, the DMS model generates multiple links between any pair of nodes, as depicted in Fig. 4.1. Dorogovtsev, *et. al.*, [18] state that for large-scale networks (*i.e.*, $t >> 0$) the probability of emerging multiple links tends to zero. However, although the probability of emerging multiple links in the network tends to zero as $t \to \infty$, the existence of multiple links generated in earlier evolution states of the network remain during the whole life of the network. In contrast, most real CN do not have multiple links. For example, in a paper citation network an article in its reference section does not have two identical references, in a friendship network not exist more than once friendship bond between two individuals. To support this notion, a subset of real networks has been analyzed (see Table 4.1) and in these networks no multiple links were found. More examples of networks with no multiple links can be found in [16].

Table 4.1: List of some real-world directed complex networks with their corresponding number of nodes and links between nodes. No multiple links are found in any of them.

Real-world networks	**Number of nodes**	**Number of directed links**
The WWW at nd.edu domain network. [14]	$325,729$	$1,497,134$
The citation network in the U.S. patents from 1975 to 1999. [15]	$3,774,324$	$16,522,438$

Real-world networks	**Number of nodes**	**Number of directed links**
The Internet topology at the autonomous system level. [35]	39, 280	73, 324
The collaboration network of *Arxiv* Astro Physics category (period January 1993 to April 2003). [36]	18, 772	396, 160
The paper citation network of *Arxiv* High Energy Physics category (period January 1993 to April 2003). [36]	34, 546	421, 578
The paper citation network of *Arxiv* High Energy Physics Theory category (period January 1993 to April 2003). [36]	27, 770	352, 807
The email network of a large European Research Institution (period October 2003 to May 2005). [36]	265, 214	420, 045
The collaboration network of *Arxiv* General Relativity and Quantum Cosmology category (period January 1993 to April 2003). [36]	5, 242	28, 980
The collaboration network of *Arxiv* Condense Matter Physics (period January 1993 to April 2003). [36]	23, 133	18, 693
The collaboration network of *Arxiv* Condense Matter Physics (period January 1993 to April 2003). [36]	23, 133	18, 693
The directed network of flights between airports of the world. [16]	2, 939	30, 501

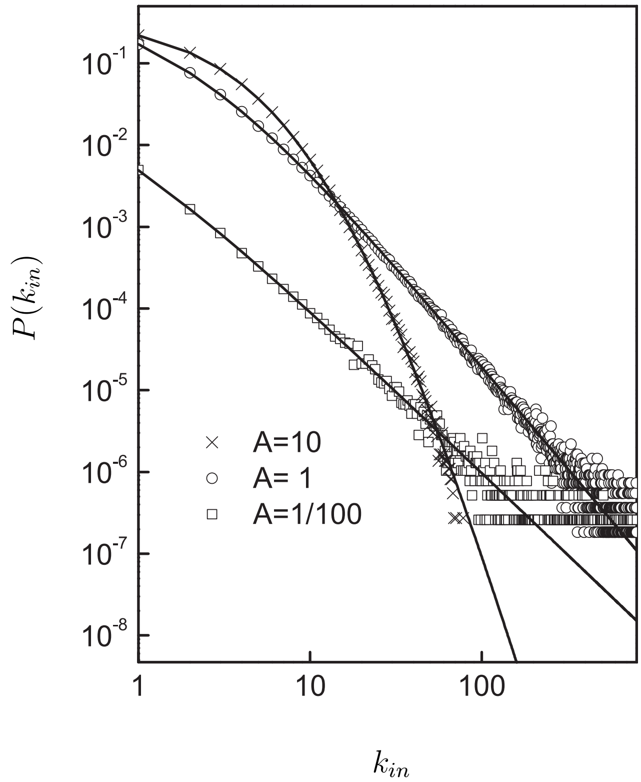

Figure 4.2: Distribution $P(k_{in})$ obtained with $m = 2$ and three different values for A. The symbols $(\times, \bigcirc, \square)$ correspond to three numerical simulations of DMS model considering that links emerge only from every new added node. The line is the analytical solution of DMS model (Eq. 9 in Ref. [18]). Figure shows that the behavior of the DMS model is not affected by the origin of the links added to the network as stated in Ref. [18].

In order to measure the fraction of multiple links generated by DMS model, only links emerging from every new added node to the network are considered. Furthermore, the behavior of the DMS model is not affected by the origin of the links added to the network [18] (see Fig. 4.2).

By using the DMS model, 12 experiments were carried out consisting of two sets of simulations, one with 6 simulations with $m = 2$ and another one with 6 simulations with $m = 8$; both sets used $A = 10, 1, 0.1, 0.01, 0.001$, and 0.0001. In every experiment, the network growth from $t = 2$ up to $t = 10^5$ nodes. And the experiments were performed 10^3 times and then averaged out. In each experiment the amount of directed multiple links q_{dml} and the total number of directed links q_{dl} within the network were measured. With these values the fraction of directed multiple links P_{dml} was calculated as follow:

$$P_{dml} = \frac{q_{dml}}{q_{dl}} \approx \frac{q_{dml}}{mt}. \tag{4.2}$$

Note that m is the same for all the nodes and therefore $q_{dl} \approx mt$.

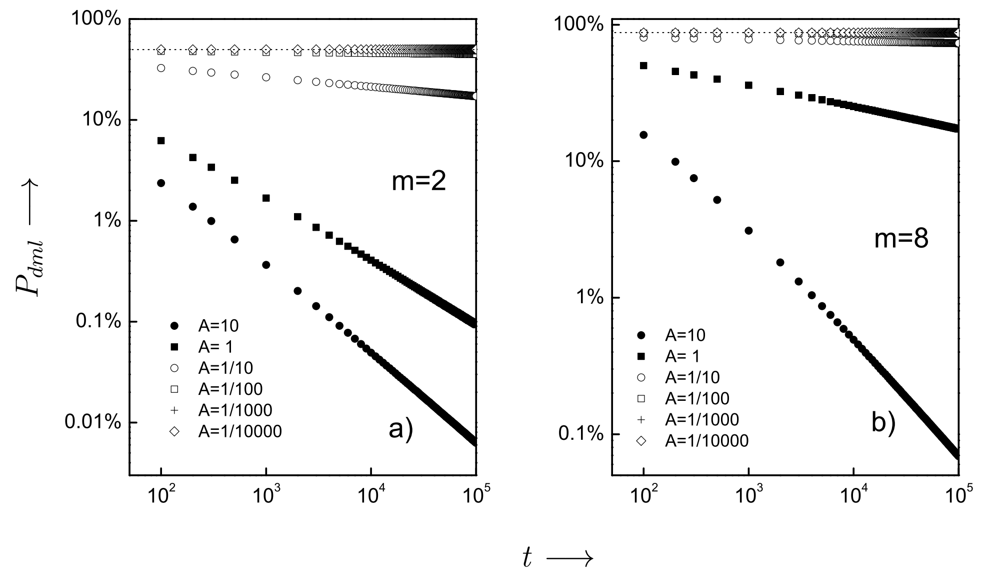

Figure 4.3: Fraction of directed multiple links P_{dml} obtained by using the DMS model for several values of A as a function of time t. (a) with $m = 2$, and (b) with $m = 8$. See text for details.

The values for P_{dml} retrieved from the experiments are shown in Fig. 4.3. It can be see that as $A \to 0$, the value of P_{dml} attains a constant value independent of the size of the network (t). For example, with $m = 2$ (Fig. 4.3a), $P_{dml} \approx 49.9\%$ (dashed line) for $A \approx 0$. In this case, for every two directed links added to the network, one of them is rendered as a directed multiple link. Likewise, for $m = 8$ (Fig. 4.3b), $P_{dml} \approx 87.4\%$ (dashed line) for $A \approx 0$. From eight directed links added to the network, seven are directed multiple links.

From the above, it can infer that when the initial attractiveness A approaches zero, the probability $\prod_{DMS}$ that a node n_i belonging to the network gets a new link is ruled by the number of incoming links $k_{in(i)}$, as stated by Eq. 4.1. Thus, nodes n_j having no incoming links ($k_{in(j)} = 0$) exhibit $\prod_{DMS} \approx 0$. Consequently, every new node has a high probability to link an only one node through its m outgoing links, thus yielding $m - 1$ directed multiple links. Accordingly, for $A \to 0$ the ratio P_{dml} now takes the form:

$$P_{dml}(m) \approx \frac{m-1}{m}, \quad for\ A \to 0. \tag{4.3}$$

To illustrate this process, consider the scenario shown in Fig. 4.4a which describes the growth of a directed network from t_0 up to t_1, with $A = \frac{1}{10000}$ for all the nodes. A new node n_{new} is born at each time step, which is connected through $m = 2$ links and t is divided into two time sub-steps τ_0 and τ_1; each τ_i is used by

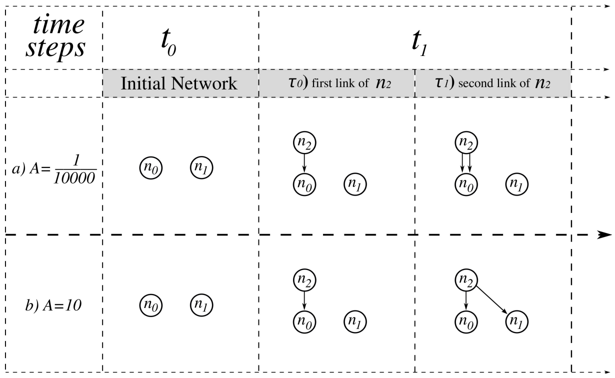

Figure 4.4: a) Growth of a directed network using the DMS model with $A = \frac{1}{10000}$ from t_0 to t_1. b) Growth of a directed network using the DMS model with $A = 10$, from t_0 to t_1.

n_{new} to connect one link. At t_0 the network comprises nodes n_0 and n_1 without incoming links. Therefore, according to Eq. 4.1, $\prod_{DMS}$ takes the value of $\frac{1}{2}$ for each one. At τ_0 of t_1, node n_2 is born and it chooses to link to one of the two nodes already existing in the network. It is assumed that node n_2 links to n_0 using its first link (see τ_0 of t_1 in Fig. 4.4a). At the end of τ_0, the probability $\prod_{DMS}$ to get a second link from n_2 is $\frac{1+0.0001}{1+0.0002} \approx 1$ for n_0, and $\frac{0.0001}{1+0.0002} \approx 0$ for n_1. At τ_1 of t_1, n_2 is linked to n_0 using its second link (see τ_1 of t_1 in Fig. 4.4a), thus giving raise a directed multiple link in the network. At the end of t_1, the probability to get a link from a new node, is $\prod_{DMS} = \frac{2+0.0001}{2+0.0002} \approx 1$ for n_0 and $\frac{0.0001}{2+0.0002} \approx 0$ for n_1. Subsequent new nodes will have a higher probability to link to node n_0 throughout their $m = 2$ outgoing links.

Fig. 4.3 shows that as $A >> 0$ and $t \to \infty$, then $P_{dml} \to 0$. If this would be the case, then the real complex networks would have multiple links in any stage of their evolution. This contrast with the networks listed in Table 4.1 as they have not multiple links. This is further understood by noticing that, as A becomes greater than 1, the difference between subsequents probabilities $\prod_{DMS}$ associated to node n_j decreases, indicating a random process in the network. In order to clarify this feature, consider the scenario shown in Fig. 4.4b with $A = 10$. At t_0 the network comprises nodes n_0 and n_1 without incoming links. Therefore, according to Eq. 4.1, $\prod_{DMS}$ takes value of $\frac{1}{2}$ for both nodes.

At τ_0 of t_1, node n_2 is born and it chooses to link to one of the two nodes already existing in the network. It is assumed that node n_2 links to n_0 using its first link (see τ_0 of t_1 in Fig. 4.4b). At the end of τ_0, the probability $\prod_{DMS}$ to get a second link from n_2 is $\frac{1+10}{1+20} \approx \frac{1}{2}$ for n_0, and $\frac{10}{21} \approx \frac{1}{2}$ for n_1; thus, when comparing to the case outlined in Fig. 4.4a ($A = \frac{1}{10000}$) with $A = 10$, the difference of $\prod_{DMS}$ for nodes n_0 and n_1 decreases and exhibits a random process. At τ_1 of t_1, n_2 is linked to n_1 using its second link (see τ_1 of t_1 in Fig. 4.4b). At the end of t_1, the probability $\prod_{DMS}$ to get a link from a new node, is $\frac{1}{2}$ for both n_0 and n_1. This causes the existence of a smaller probability for directed multiple links emerge.

4.2 The MLF model

The MLF model relies on the DMS mechanism with a different assumption in that each node n_i can have any number of incoming links $k_{in(i)}$, but stemming from different nodes; *i.e.*, it is not allowed that node n_i has more than one incoming link from node n_j. It is worth to mention that in the MLF model the links emerge only from the new nodes added to the network. In essence, the growth in the MLF model is ruled by node addition with preferential attachment: initially, there exists two nodes connected by one directed link, and at each subsequent time step a new node n_{new} is added to the network with $k_{out} = m$ outgoing links to be connected to the nodes already existing in the network. Each time step t is divided into m sub-time steps τ_j ($t \rightarrow \tau_0, \tau_1, ..., \tau_{m-1}$). Additionally, every τ_j is employed by node n_{new} to connect to the network by using its m links. The probability $\prod_{MLF}$ that node n_i belonging to the network gets a link from n_{new} is given by:

$$\prod_{MLF}(k_{in(i)}) = \begin{cases} 0 & if \quad n_i \in V_{new} \\ \dfrac{k_{in(i)} + A}{\sum\limits_{n_j \notin V_{new}} (k_{in(j)} + A)} & if \quad n_i \notin V_{new} \end{cases}, \tag{4.4}$$

where A is the initial attractiveness of n_i and V_{new} is the set of nodes that have received an incoming link from node n_{new}. Such a set is necessary in order to avoid the existence of multiple links.

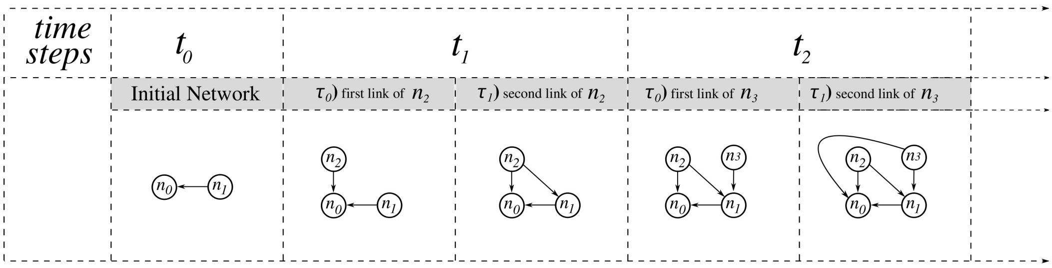

Figure 4.5: Growth of a directed network using the MLF model from t_0 to t_2.

In order to shed light on the behavior of the MLF model, consider the scenario shown in Fig. 4.5 which shows the growth of a directed network from t_0 up to t_2, with $A = 1$ for all the nodes. A new node is born at each time step, which connects through $m = 2$ links and t is divided into two time sub-steps τ_0 and τ_1. At t_0 the network comprises nodes n_0 and n_1. Node n_0 has one incoming link, whereas n_1 has none. Therefore, according to Eq. 5.1, $\prod_{MLF}$ now takes values of $\frac{2}{3}$ and $\frac{1}{3}$ for n_0 and n_1, respectively. At τ_0 of t_1, node n_2 is born and it chooses to link to one of the two nodes already existing in the network. It is assumed that node n_2 links to n_0 using its first link (see τ_0 of t_1 in Fig. 4.5). Subsequently, since n_0 already got an incoming link from n_2, n_0 now belongs to set V_2. At the end of τ_0, the probability $\prod_{MLF}$ to get a second link from n_2 is zero and one for n_0 and n_1, respectively. At τ_1 of t_1, n_2 is linked to n_1 using its second link (see τ_1 of t_1 in Fig. 4.5) and n_1 is added to V_2. At t_2, node n_3 joins the network and it will link in the same way node n_1 did, and so on for the next new nodes joining the network.

4.3 Experiment details and results

Using numerical experiments, was analyzed the impact that different values of A and m have on the in-degree distribution $P(k_{in})$, on the corresponding network clustering and on the shortest path using the proposed MLF model. Twelve experiments were carried out: I) four with $A = 1$ and $m = 1, 2, 8, 32$; II) four with $A = 10^{-2}$ and $m = 1, 2, 8, 32$; III) four with $A = 10^{-4}$ and $m = 1, 2, 8, 32$. In each experiment, the network growth from 2 up to 10^4 nodes, repeated 10^3 times and averaged out.

Fig. 4.6 shows the in-degree distribution $P(k_{in})$, Fig. 4.7 the CDF of clustering (C) and Fig. 4.8 the CDF of shortest path length (SPL) of the network obtained for each experiment. Two different cases can be distinguished and are described as follows:

- *Case 1 ($m = 1$):*
  Figs. 4.6*a*, 4.7*b* and 4.8*c* show the experimental results with $m = 1$. The MLF model produces a $P(k_{in})$ with exponents γ_{MLF}, identical to those predicted by the DMS model (see Eq. 2.3). This is expected since with $m = 1$ the existence of directed multiple links is not possible, and therefore both DMS and MLF models yield similar results. The clustering gives $C = 0$, which is also expected since with $m = 1$ the resulting network is a tree. Finally, the average of the SPL tends to ≈ 1 as $A \to 0$.

- *Case 2 ($m > 1$):*
  Figs. 4.6-4.8 (*d*-*l*) shows the experimental results with $m = 2$, 8 and 32, respectively. Contrary to the case with $m = 1$ presented above, one can see that the MLF model yields $P(k_{in})$ distributions with $\gamma_{MLF} \to 1$, as $A \to 0$ and m attains values larger than one. This contrasts to the lower bound value of $\gamma_{DMS} \approx 2$ (Eq. 2.3). Finally, the clustering C tends to ≈ 0.5 and the SPL tends to ≈ 1, as $A \to 0$.

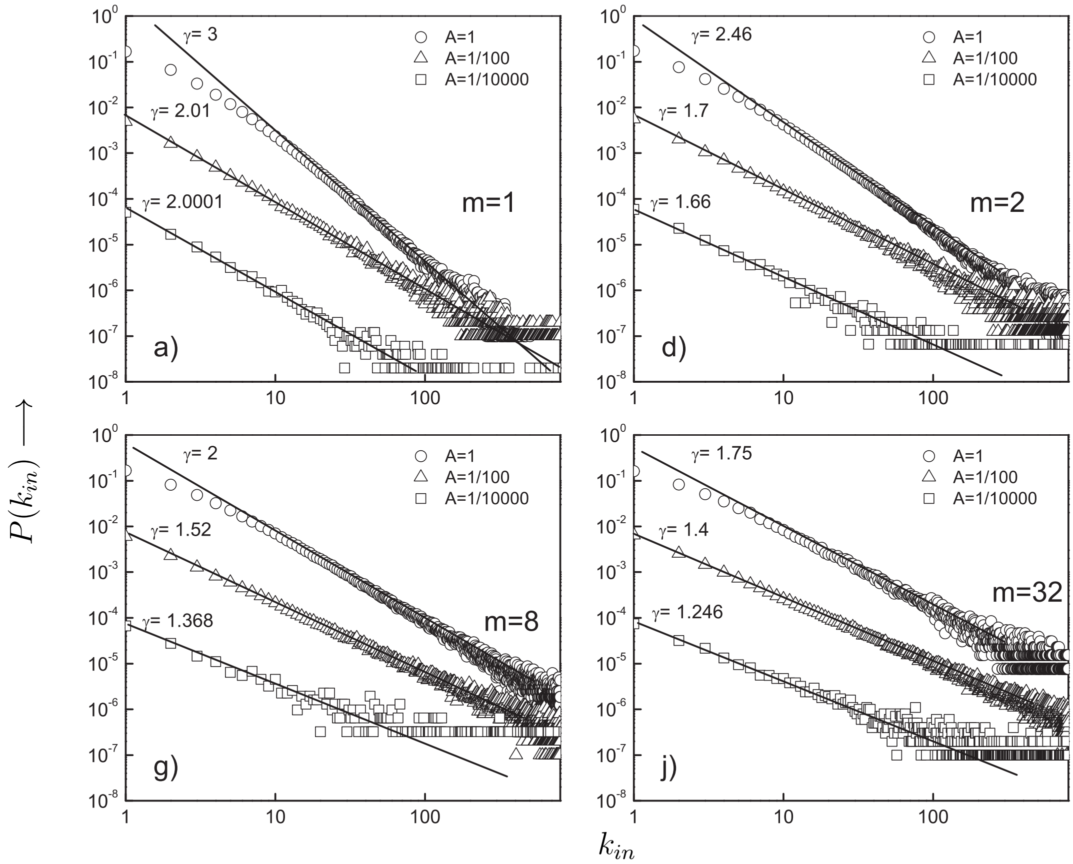

Figure 4.6: a, d, g, j) In-degree distributions $P(k_{in})$ retrieved from numerical experiments with different values of m and A. In the figures, the symbols $(\bigcirc, \triangle, \square)$ represents the In-degree distribution retrieved from the simulations for each value of A and the solid lines are power laws with its respective γ exponent.

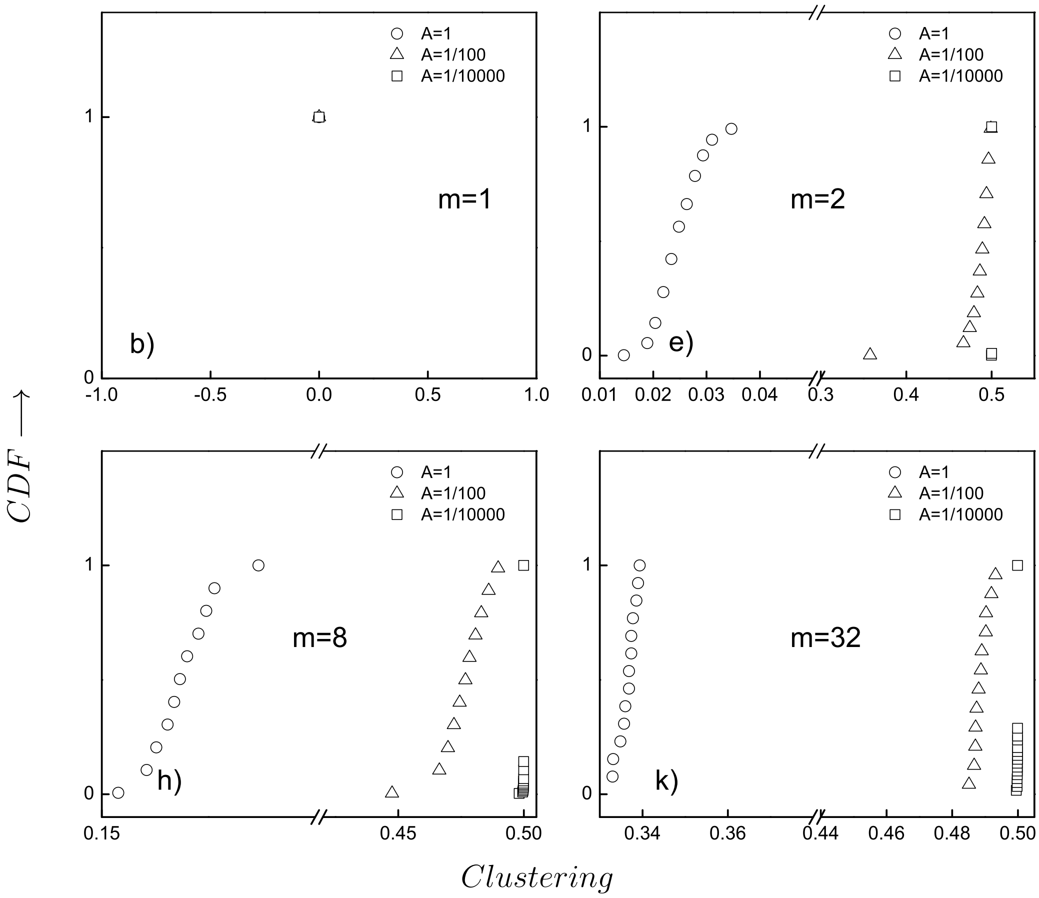

Figure 4.7: b, e, h, k) Cumulative distribution function (CDF) obtained respect to the clustering (C) of the networks.

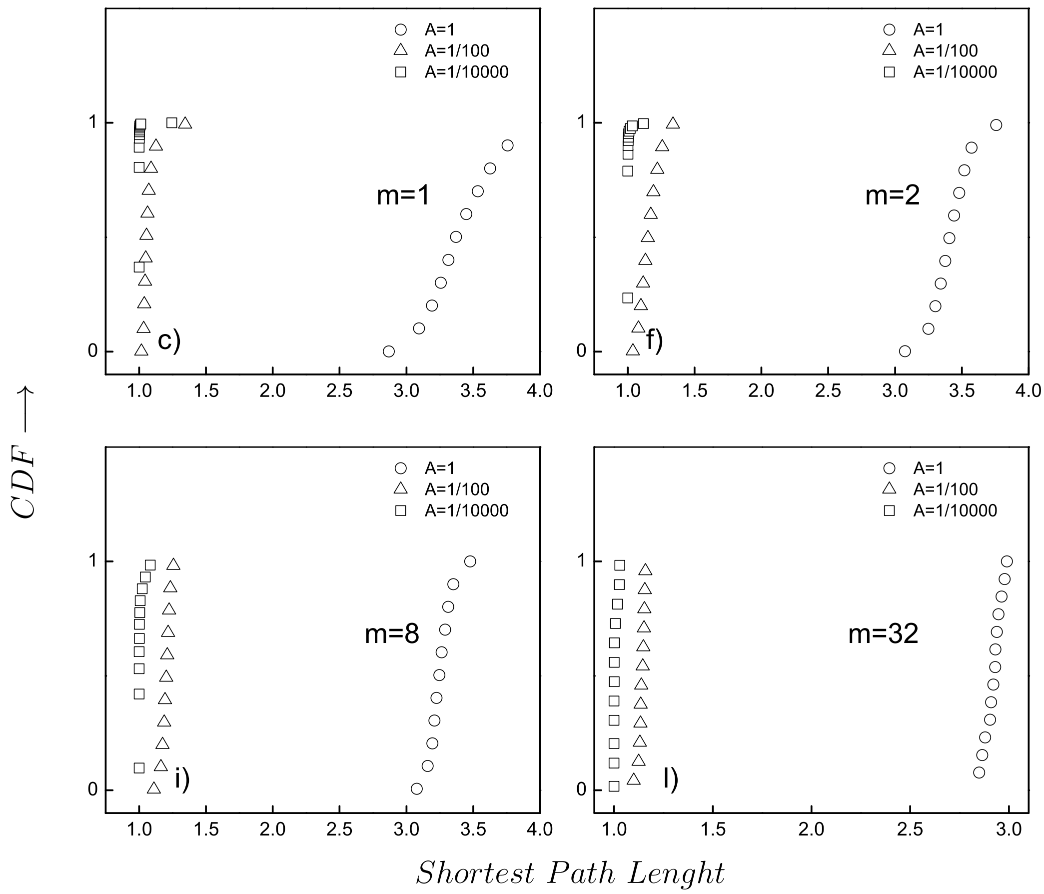

Figure 4.8: c, f, i, l) Cumulative distribution function (CDF) obtained respect to the shortest path lengths (SPL) of the networks.

From the above experimental results, it can infer that as $A \to 0$ and $m >> 1$, the exponent γ_{MLF} of $P(k_{in})$ tends to ≈ 1, the clustering of the network increases and the average length of the shortest-paths in the network decreases.

4.4 Analytical solution of MLF model

In this section, is developed an analytical solution for the MLF model in the limit as initial attractiveness of nodes approaches to zero ($lim_{A\to 0}$).

Recalling that in the DMS model, at each time step a new node is connected to the network through m directed links, Eq. 4.1 can be written as:

$$\prod_{DMS}(k_{in}, t) = \frac{k_{in} + am}{(1+a)mt}, \tag{4.5}$$

where $a = \frac{A}{m}$ according to the procedure outlined in Ref. [18].

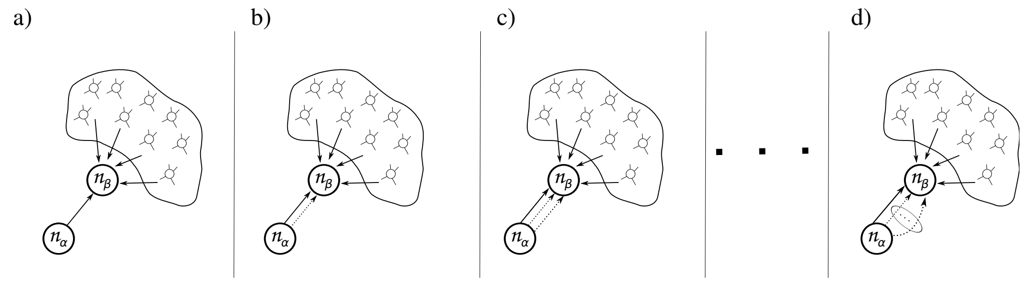

Figure 4.9: Connection of node n_α to n_β through its m outgoing links according to the DMS model in the limit as initial attractiveness of nodes approaches to zero ($lim_{A\to 0}$). See text for details.

From Eq. 4.3 is known that using the DMS model, in $lim_{A\to 0}$, every new added node (n_{new}) has a high probability to connect a single node using its m outgoing links, yielding $m-1$ directed multiple links. This behavior is also denoted by Dorogovtsev *et.al.* [18] (in discussion section). With such a result, it is possible define the probability

$$\Psi_{dml}(x,\varepsilon) \approx \frac{\varepsilon}{x}, \quad in\ lim_{A\to 0}, \tag{4.6}$$

where n_{new} has ε directed multiple links after it has connected to the network using x links, where $x = 1, 2, ..., m$.
To elucidate how ε behaves with respect to x, consider Fig. 4.9. In this figure the node n_α links to node n_β through its m outgoing links. After n_α to connects the first link ($x = 1$) there are no directed multiple links ($\varepsilon = 0$), thus $\Psi_{dml}(x,\varepsilon) = \frac{0}{1}$ (Fig. 4.9 a); after n_α connects the second link ($x = 2$) there is one directed multiple links ($\varepsilon = 1$), thus $\Psi_{dml}(x,\varepsilon) = \frac{1}{2}$ (Fig. 4.9 b); after n_α connects the third link ($x = 3$) there are two directed multiple links ($\varepsilon = 2$), thus $\Psi_{dml}(x,\varepsilon) = \frac{2}{3}$ (Fig. 4.9 c); after n_α connects the $x-th$ link there are $x-1$ directed multiple links ($\varepsilon = x-1$). Therefore, Eq. 4.6 can be written as:

$$\Psi_{dml}(x) \approx \frac{x-1}{x}, \quad in\ lim_{A\to 0}. \tag{4.7}$$

From the above, is possible to define the probability for the $x-th$ link, added by n_{new}, to be a directed non-multiple link as:

$$1 - \Psi_{dml}(x) \approx 1 - \frac{x-1}{x}$$
$$1 - \Psi_{dml}(x) \approx \frac{1}{x}, \quad in\ lim_{A\to 0}. \tag{4.8}$$

Now is defined $P(k_{in(i)}, i, t)$ as being the probability for node n_i to have $k_{in(i)}$ incoming links at time step t. Thus on the average, for an arbitrary node the probability is given by:

$$P(k_{in},t) = \frac{1}{t}\sum_{s=1}^{t} P(k_{in(s)},s,t). \tag{4.9}$$

By taking into account that using the MLF model, at each time step t m non-multiple directed links are added to the network, the temporal evolution for $P(k_{in(i)},i,t)$ is given by the following master equation:

$$P(k_{in(i)},i,t+1) =$$

$$\overbrace{m\left(\overbrace{\prod_{DMS}(k_{in}-1,t)\left(1-\Psi_{dml}(1)\right)}^{first\ link}+\cdots+\overbrace{\prod_{DMS}(k_{in}-1,t)\left(1-\Psi_{dml}(m)\right)}^{m-th\ link}\right)P(k_{in(i)}-1,i,t)}^{p_1}$$

$$+\overbrace{\left[1-m\left(\overbrace{\prod_{DMS}(k_{in},t)\left(1-\Psi_{dml}(1)\right)}^{first\ link}+\cdots+\overbrace{\prod_{DMS}(k_{in},t)\left(1-\Psi_{dml}(m)\right)}^{m-th\ link}\right)\right]P(k_{in(i)},i,t)}^{p_2}. \tag{4.10}$$

The term labeled as p_1 in Eq. 4.10 is associated to the fact that node n_i is chosen so as to acquire the $first$, *second* or $m-th$ non-multiple incoming link from n_{new}; whereas p_2 describes the situation when n_i is not chosen to acquire the non-multiple incoming link.
Simplifying is obtained,

$$\begin{aligned} P(k_{in(i)},i,t+1) =& m\prod_{DMS}(k_{in}-1,t)\left(\sum_{x=1}^{m}(1-\Psi_{dml}(x))\right)P(k_{in(i)}-1,i,t) \\ &+\left[1-m\prod_{DMS}(k_{in},t)\left(\sum_{x=1}^{m}(1-\Psi_{dml}(x))\right)\right]P(k_{in(i)},i,t). \end{aligned} \tag{4.11}$$

By substituting Eq. 4.5 and defining β as:

$$\beta = \sum_{x=1}^{m}(1-\Psi_{dml}(x)) = \sum_{x=1}^{m}\left(1-\frac{x-1}{x}\right) = \sum_{x=1}^{m}\frac{1}{x}$$

into Eq. 4.11:

$$\begin{aligned} P(k_{in(i)},i,t+1) =& \left[\frac{\beta(k_{in}-1+am)}{(1+a)t}\right]P(k_{in(i)}-1,i,t) \\ &+\left[1-\left(\frac{\beta(k_{in}+am)}{(1+a)t}\right)\right]P(k_{in(i)},i,t), \end{aligned} \tag{4.12}$$

By performing the summatory in Eq. 4.12 from $i = 1$ to t, and by taking account Eq. 4.9 is obtained

- for the left-hand side:

$$\sum_{i=1}^{t} P(k_{in(i)}, i, t+1) = \sum_{i=1}^{t+1} P(k_{in(i)}, i, t+1) - P(k_{in(i)}, t+1, t+1)$$
$$= (t+1)P(k_{in}, t+1) - \delta_{k,0}, \tag{4.13}$$

  where $\delta_{k,0}$ means that node are born with zero in-degree (*i.e.* without incoming links);

- for the right-hand side:

$$\sum_{i=1}^{t} P(k_{in(i)} - 1, i, t) = tP(k_{in} - 1, t) \tag{4.14}$$

$$\sum_{i=1}^{t} P(k_{in(i)}, i, t) = tP(k_{in}, t). \tag{4.15}$$

By inserting Eqs. 4.13, 4.14 and 4.15 into Eq. 4.12, is obtained:

$$(t+1)P(k_{in}, t+1) - \delta_{k,0} = \frac{\beta(k_{in} - 1 + am)}{(1+a)} P(k_{in} - 1, t)$$
$$+ tP(k_{in}, t) - \frac{\beta(k_{in} + am)}{(1+a)} P(k_{in}, t). \tag{4.16}$$

As the number of nodes increases (*i.e.* $t >> 1$), $P(k_{in}, t)$ attains a stationary behavior: $P(k_{in}, t+1) = P(k_{in}, t) = P(k_{in})$; therefore, Eq. 4.16 converts to:

$$P(k_{in}) = \frac{\beta(k_{in} - 1 + am)P(k_{in} - 1) + (1+a)\delta_{k,0}}{1 + a + \beta(k_{in} + am)}. \tag{4.17}$$

Solve the last recurrence equation:

$$P(k_{in}) = \frac{(1+a)\Gamma\left(am + \frac{a+1}{\beta}\right)}{\beta\Gamma(am)} \frac{\Gamma(k_{in} + am)}{\Gamma\left(k_{in} + am + 1 + \frac{a+1}{\beta}\right)} \tag{4.18}$$

$$P(k_{in}) \approx \frac{(1+a)\Gamma\left(am + \frac{a+1}{\beta}\right)}{\beta\Gamma(am)} (k_{in} + am)^{-\left(1 + \frac{a+1}{\beta}\right)}, \tag{4.19}$$

where $\Gamma(\cdot)$ is the Gamma function.

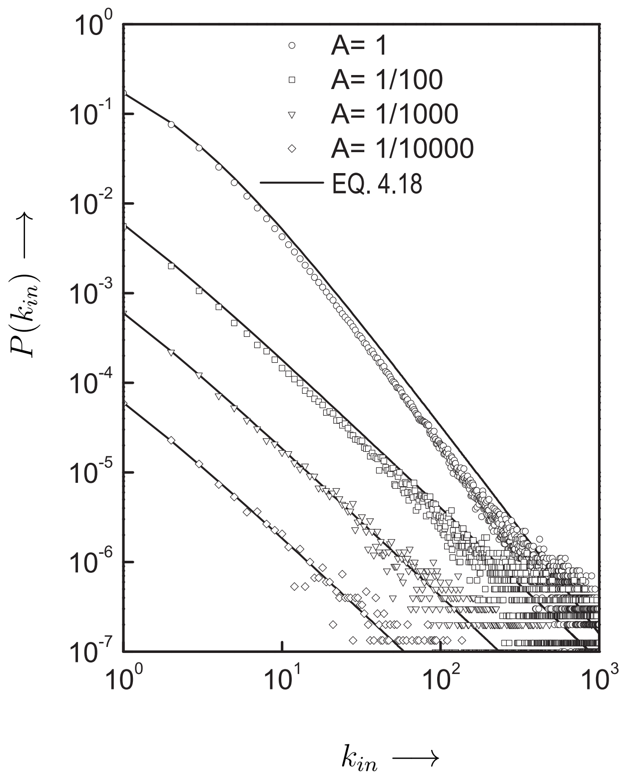

Figure 4.10: a) Experimental results (○, □, ▽ and ◇) and the corresponding fittings using the analytical MLF model (solid line, Eq. 4.18), for $m = 2$ and different values of A. Note a perfect match when $A = \frac{1}{1000}$ and $A = \frac{1}{10000}$. On the contrary, for $A = 1$ and $A = \frac{1}{100}$, the fitting is not good for $k_{in} > 10$.

Thus, is found the scaling exponent γ_{MLF} of the in-degree distribution $P(k_{in})$:

$$\gamma_{MLF} = 1 + \frac{\frac{A}{m} + 1}{\sum_{x=1}^{m} \frac{1}{x}}. \tag{4.20}$$

Finally, in the $lim_{A \to 0}$ the last equation can be writen as: $\gamma_{MLF} = 1 + \frac{1}{\sum_{x=1}^{m} \frac{1}{x}}$, that is $\gamma_{MLF} \approx 1$ as $m >> 1$.

Fig. 4.10 shows four plots of $P(k_{in})$ versus k_{in} using the model outlined in this section (Eq. 4.18) and is compared to the experimental simulations. It is possible to note a close matching when $A = 0.001$ and $A = 0.0001$. This is not the case for larger values of A ($i.e$, $A = 1$). It is pointed that the analytical solution developed is obtained in the limit as initial attractiveness of nodes approaches to zero.

4.5 Using the MLF model to reproduce some topological properties of a real network

To verify that the MLF model is able to reproduce some properties of real complex networks, the network comprising flights between airports of the world ($NFAW$) [16] was chosen. In this network, the airports correspond to the nodes and the flights to the links. This network is formed by 2,939 nodes and 30,501 non-multiple links. [16]

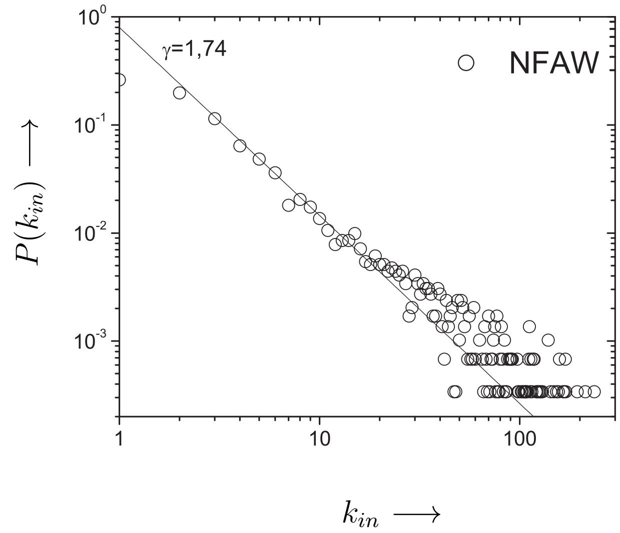

Figure 4.11: In the Figure, ◯ represents In-degree distribution of $NFAW$ network and the solid line a power law function with exponent $\gamma = 1.74$.

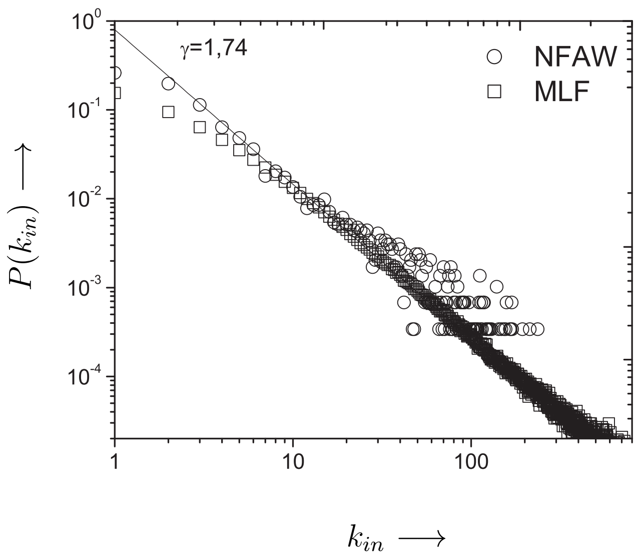

Figure 4.12: Comparison of the In-degree distribution of $NFAW$ network with the obtained from the simulation of MLF model.

Fig. 4.11 shows that the in-degree distribution $P(k_{in})$ of the $NFAW$ network decays as a power-law with exponent $\gamma \approx 1.74$. Furthermore, the Clustering and Diameter of $NFAW$ network are $C = 0.25$ and $D = 14$, respectively.

In order to generate a network with topological properties similar to the properties found in the $NFAW$ network, an experiment that simulates the growth of a directed network from 40 to 2939 nodes using the MLF model with $A = 2$ and $m = 40$ was performed. With such conditions, the in-degree distribution $P(k_{in})$ retrieved from the simulation decays as a power-law with exponent $\gamma \approx 1.74$ close to the exponent of the $NFAW$ network (see Fig. 4.12). In addition, the Clustering and Diameter obtained from the simulation are $C = 0.217$ and $D = 13.85$, respectively. These values are also close to the values of C and D of the $NFAW$ network (see Ref. [16]).

With the previous results, it is possible to deduce that the MLF model is capable of reproduce topological properties of real CN. On the other hand, even though the MLF model is capable of generate a network with topological properties close to the properties of the $NFAW$ network, is not possible to ensure that the local processes incorporated by the MLF model are the only ones involved in the growth and evolution of the $NFAW$. However, it is possible think that the MLF model is a realistic simplification of some of these processes.

In this chapter was proposed a growth model for directed complex networks called MLF model. The MLF model incorporates the prohibition of multiple links between pairs of nodes and the Initial attractiveness, with these characteristics the model is able to generate directed CN with In-degree distribution that decay as a power law $P(k_{in}) \sim k_{in}^{-\gamma_{in}}$ with $1 < \gamma_{in} < \infty$. That is, the model is capable to generate all exponents found in the In-degree distribution of directed CN that are documented [8, 11, 16, 31]. In spite of that, in several real networks exists more local processes than the incorporated by the MLF model. For example in the WWW and Social networks, there may be addition and rewiring of links. For this reason, in the next chapter is developed a growth model that extends the MLF model adding the internal link addition and rewiring processes.

Chapter 5

The impact of local processes and the prohibition of multiple links in the topological properties of directed Complex Networks

Local processes participate in the growth and evolution of CN, which in turn shape the topological and dynamical properties of these networks. For example, in the WWW network, links are not static and, at any time, a node (web page) may loose a connection to another node (deleting a hyper-link) and add this same connection to a different node (a rewiring process), also new links can appear in the network (links added). These local processes are also presents in other CN like the Internet, social networks and collaboration networks.

In this chapter is investigated the impact that the prohibition of multiple links in addition with other local processes have in several topological properties of CN. For that, it is proposed a new growth and evolution model that incorporates the internal links, rewiring and multiple links prohibition because these processes are able to vary the value of the exponent γ of $P(k)$ maintaining the power law behavior (see chapters 3 and 4 for details). In particular, is studied the impact that these processes have in Clustering (C), Shortest path length (SPL) and the In-degree distribution ($P(k_{in})$) of the networks generated with the proposed model.

This chapter is organized as follows: The proposed model is introduced in Section 5.1. A description of the numerical simulations and its results are showed in Section 5.2. Section 5.3 demonstrates that the proposed model is able to reproduce some topological properties of a real network.

5.1 Proposed model

In the proposed model, the network grows by adding nodes and links. To connect the nodes preferential attachment is employed, that is, the probability $\prod$ that a node n_i belonging to the network gets a link from a node n_j is proportional to the sum of the in-degree of n_i and its initial attractiveness A, as follows:

$$\prod(k_{in(i)}) = \begin{cases} 0 & if \quad n_i \in V_j \quad or \quad n_i = n_j \\ \dfrac{k_{in(i)} + A}{\sum\limits_{n_x \notin V_j, n_x \neq n_j} (k_{in(x)} + A)} & if \quad n_i \notin V_j \quad and \quad n_i \neq n_j \end{cases}, \tag{5.1}$$

where V_j is the set of nodes that have received an incoming link from node n_j, as in Ref. [27]. Also, eq. 5.1 prohibits multiple links and loops (a loop is an link that starts and finalizes in the same node).

In the model, it is assumed that initially (at t_0) are m_0 nodes with some links between them and in the following t time-steps either of the three following operations may happen:

- With probability q, m rewiring's happen in the network. For each rewiring, a node n_r is randomly selected. The node n_r should randomly select a neighbor n_s ($n_s \in V_r$) and delete its link to this node (n_s no longer belongs to V_r). Then n_r connects to another node following Eq. 5.1.

- With probability p, m new links are added to the network. For each new link, a node n_r present in the network is randomly selected to be the origin of the new link and the end of the new link is connected to another network's node using Eq. 5.1.

- With probability $1-p-q$, a new node n_{new} is added to the network with $k_{out} = m$ links that must be connected with m different existing nodes accordingly to Eq. 5.1.

In order to show the behavior of the proposed model, consider Fig. 5.1 which shows the growth and evolution of a directed network from t_0 to t_3. Every new node is born with an initial attractiveness $A = 1$ and two outgoing links ($m = 2$). The probabilities $q = \frac{1}{3}$ and $p = \frac{1}{3}$.

At t_0, the network only has three nodes, n_0, n_1 and n_2. Following Eq. (5.1), the probability $\prod$ of obtaining a new incoming link is $\frac{3}{5}$ for n_0 and $\frac{1}{5}$ for n_1 and n_2.

At τ_0 in t_1, a new node n_3 is born and it selects an existing node in the network to connect to. We assume that n_3 employs its first link to connect to n_0 (see τ_0 in t_1 in Fig. 5.1). Now n_0 belongs to the V_3 set. As τ_0 completes, the probability $\prod$ of receiving a second link from n_3 is zero for n_0 and $\frac{1}{2}$ for n_1 and n_2. At τ_1 of t_1, n_3 employs its second link to connect to n_2 (see τ_1 in t_1 in Fig. 5.1) and n_2 now belongs to V_3.

Now assume that there is an addition of $m = 2$ new links at t_2. At τ_0 of t_2, node n_0 is randomly selected to generate a new outgoing link. Following Eq. (5.1), the probability $\prod$ at end of t_1 of receiving the incoming link from n_0 is zero for n_0 (loops are not allowed), $\frac{2}{4}$ for n_2, $\frac{1}{4}$ for n_1 and $\frac{1}{4}$ for n_3. Assume that n_0 connects to n_1

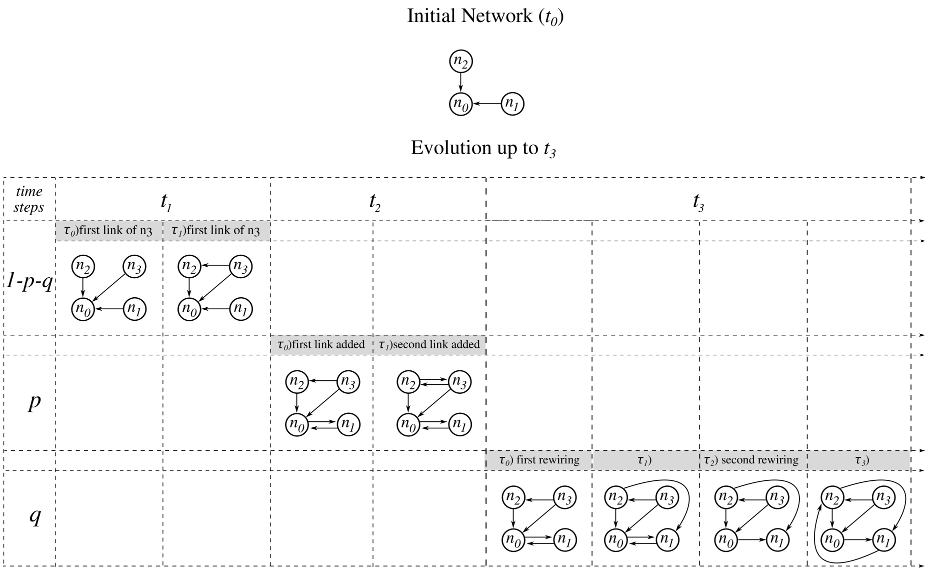

Figure 5.1: Evolution of a directed complex network from t_0 to t_3 using the proposed model.

(see τ_0 in t_2 in Fig. 5.1) and n_1 now belongs to V_0. At τ_1 of t_2, node n_2 is randomly selected to generate a new outgoing link. Following Eq. (5.1), the probability $\prod$ at end of τ_0 in t_1 of receiving the incoming link from n_2 is zero for n_0 and n_2, $\frac{2}{3}$ for n_1 and $\frac{1}{3}$ for n_3. Assume that n_2 connects to n_3 (see τ_1 of t_2 in Fig. 5.1) and n_3 now belongs to V_2.

Now assume that there is $m = 2$ rewiring at t_3. At τ_0 of t_3, node n_2 is randomly selected to perform the rewiring operation. n_2 chooses to disconnect from n_3 and so, n_3 no longer belongs to set V_2. Then, following Eq. (5.1) the probability $\prod$ at the end of τ_0 in t_3 of receiving the incoming link from n_2 is zero for n_0 and n_2, $\frac{2}{3}$ for n_1 and $\frac{1}{3}$ for n_3. Assume that n_2 connects to n_1 (see τ_1 in t_3 in Fig. 5.1) and n_1 now belongs to V_2. At τ_2 of t_3, node n_1 is randomly selected to perform the rewiring operation. n_1 chooses to disconnect from n_0 and so, n_0 no longer belongs to set V_1. Then, following Eq. (5.1) the probability $\prod$ at the end of τ_2 in t_3 of receiving the incoming link from n_1 is $\frac{3}{6}$ for n_0, zero for n_1, $\frac{2}{6}$ for n_2 and $\frac{1}{6}$ for n_3. Assume that n_1 connects to n_2 (see τ_3 in t_3 in Fig. 5.1) and n_2 now belongs to V_1.

5.2 Simulation details

In order to study the effects that the proposed model has in the in-degree distribution $P(k_{in})$, clustering (C) and shortest-path length (SPL) of the generated networks, five experiments were performed using $m = 2$ and $A = 10^{-4}$ and different p and q values. The experiments consisted on running network's growth simulations starting from 2 connected nodes and finishing at 10^4 nodes. Each experiment was repeated 10^4 times and averaged.

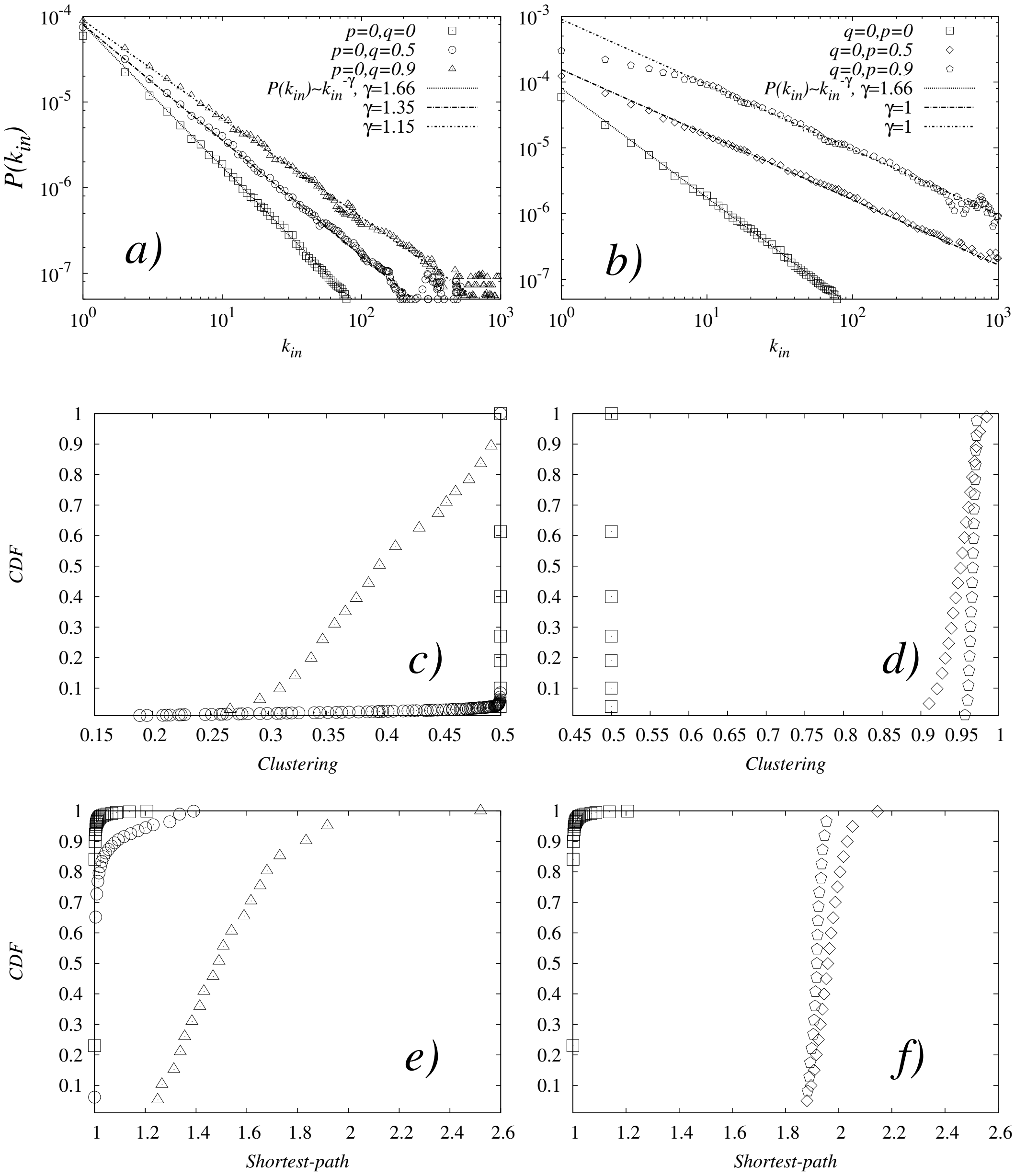

Figure 5.2: a-b) In-degree distribution $P(k_{in})$, c-d) CDF for network Clustering, and e-f) CDF for the Shortest-Path Length of the generated networks. Results produced with $m = 2$, $A = 0.0001$ and different values for p and q.

For the first experiment probabilities $q = 0$ and $p = 0$. With these conditions, $P(k_{in})$ decays as a power-law in the tail with exponent $\gamma \approx 1.66$ (see Fig. 5.2 $a-b$). Additionally, the average length of the SPL in this network is $SPL \approx 1$ (see Fig. 5.2 $c-d$). It can be seen that the clustering value is $C \approx 0.5$ (see Fig. 5.2 $e-f$). These results are the same results previously obtained by $Esquivel\ et.al.$ [27].

In order to measure the impact that the rewiring process has on the $P(k_{in})$, C and SPL, p was fixed to zero and q took values of 0.5 and 0.9. With $q = 0.5$ and $q = 0.9$, $P(k_{in})$ decays as a power-law in the tail with exponent $\gamma \approx 1.35$ and $\gamma \approx 1.15$ respectively (see Fig. 5.2 a). Additionally, the average length of the SPL in this network is $SPL \approx 1$ and $SPL \approx 1.48$ for $q = 0.5$ and $q = 0.9$ respectively (see Fig. 5.2 c). It can be seen that the clustering value is $C \approx 0.5$ and $C \approx 0.39$ for $q = 0.5$ and $q = 0.9$ respectively (see Fig. 5.2 d). That is, as the probability q approximates 1, the γ exponent decreases to ≈ 1, the clustering decreases and the shortest path length increases.

In order to measure the impact that the rewiring process has on the $P(k_{in})$, C and SPL, p was fixed to zero and q took values of 0.5 and 0.9. With $q = 0.5$ and $q = 0.9$, $P(k_{in})$ decays as a power-law in the tail with exponent $\gamma \approx 1.35$ and $\gamma \approx 1.15$ respectively (see Fig. 5.2 a). Additionally, the average length of the SPL in this network is $SPL \approx 1$ and $SPL \approx 1.48$ for $q = 0.5$ and $q = 0.9$ respectively (see Fig. 5.2 e). It can be seen that the clustering value is $C \approx 0.5$ and $C \approx 0.39$ for $q = 0.5$ and $q = 0.9$ respectively (see Fig. 5.2 c). That is, as the probability q approximates 1 the clustering decreases and the shortest path length increases.

From the results in Fig. 5.2a it is possible to generate the hypothesis that as $q \rightarrow 1$, $\gamma \rightarrow 1$. In order to confirm this hypothesis, new experiments were performed with $q = 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.99$ and $p = 0$. The exponent γ in function of q retrieved from the simulations is shown in Fig. 5.3a, as can be seen, the exponent γ approaches to 1.1 as the rewiring probability tends to 1.

Finally, in order to measure the impact that adding links has on the $P(k_{in})$, C and SPL, q was fixed to zero and p took values of 0.5 and 0.9. With $p = 0.5$ and $p = 0.9$, $P(k_{in})$ decays as a power-law in the tail with exponent $\gamma \approx 1$ (see Fig. 5.2 b). That is, the exponent γ is the same for these values of p. However for $p = 0.9$, the $P(k_{in})$ distribution is rescaled with respect to the $P(k_{in})$ obtained with $p = 0.5$, this behavior is probably because as $p \rightarrow 1$, the number of links increases faster than the number of nodes and the network becomes dense. It is important to mention that a similar rescaling behavior is also present in the Barabási-Albert model as the m parameter increases. [17] Additionally, the average length of the SPL in this network is $SPL \approx 1.95$ and $SPL \approx 1.91$ for $p = 0.5$ and $p = 0.9$ respectively (see Fig. 5.2 f). It can also be seen that the clustering value is $C \approx 0.95$ and $C \approx 0.96$ for $p = 0.5$ and $p = 0.9$ respectively (see Fig. 5.2 d).

From Fig. 5.2b it is possible to generate the hypothesis that for $0 < p < 0.5$, $\gamma \rightarrow 1$ and for $0.5 \leq p < 1$, $\gamma \approx 1$. In order to confirm this hypothesis, new experiments were performed with $p = 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.99$ and $q = 0$. The exponent γ in function of p retrieved from the simulations is shown in Fig. 5.3b. This Figure shows that the γ exponent approaches to 1 as the probability p tends to 1.

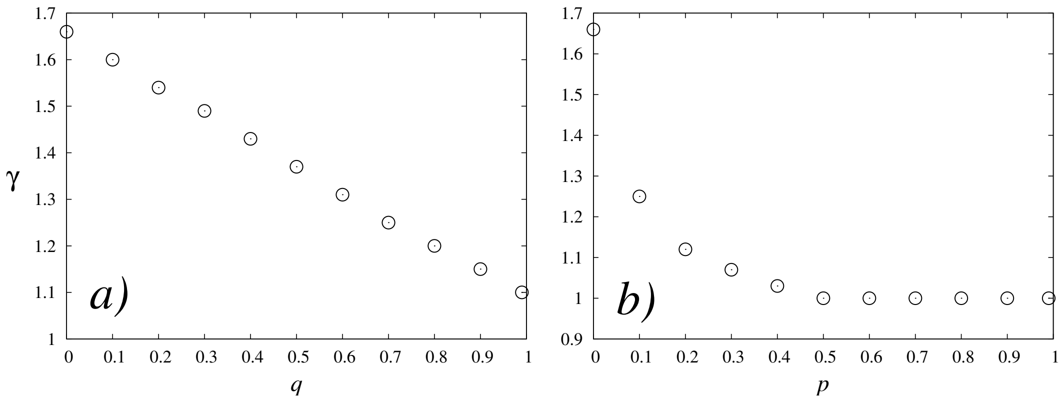

Figure 5.3: a) Values of γ retrieved from the simulations with different values of q and $p = 0$. b) Values of γ retrieved from the simulations with different values of p and $q = 0$.

From these results it is possible to conclude that the proposed model allows to obtain exponent values $\gamma \approx 1$ by increasing the rewiring probability q or the probability of adding links p, without the need to employ large values of m as in the model previously proposed by *Esquivel, et.al.*. [27]

5.3 Using the proposed model to reproduce some topological properties of a real network

To verify that the proposed model is able to reproduce some properties of real complex networks, we chose the trust network from the online social network *Epinions*, [16]. In this network, users of *Epinions* correspond to the nodes and the trust between the users to the links. This network is formed by $75,879$ nodes and $508,837$ non-multiple links and no loops. [16]

Fig. 5.4a) shows the in-degree distribution $P(k_{in})$ of the *Epinions* network that decays as a power-law with $\gamma \approx 1.69$. Furthermore, the Clustering, average length of the SPL and Diameter of *Epinions* network are $C = 0.0657$, $SPL = 4.40$ and $D = 15$, respectively.

In order to generate a network with similar topological properties to the ones found in *Epinions*, an experiment that simulates the growth of a directed network up to $78,879$ nodes was performed using the proposed model with $A = 0.65$, $p = 0.6$, $q = 0.3$ and $m = 4$. With such conditions, the in-degree distribution $P(k_{in})$ retrieved from the simulation decays as a power-law with exponent $\gamma \approx 1.69$ close to the exponent of the *Epinions* network (see Fig. 5.4b)). In addition, the Clustering, the average length of the SPL and Diameter obtained from the simulation are $C = 0.25$, $SPL = 3.51$ and $D = 13$, respectively. Values of SPL and D obtained are also close to the values of SPL and D for the *Epinions* network (see Ref. [16]). However the value of C retrieved from the simulation is very different.

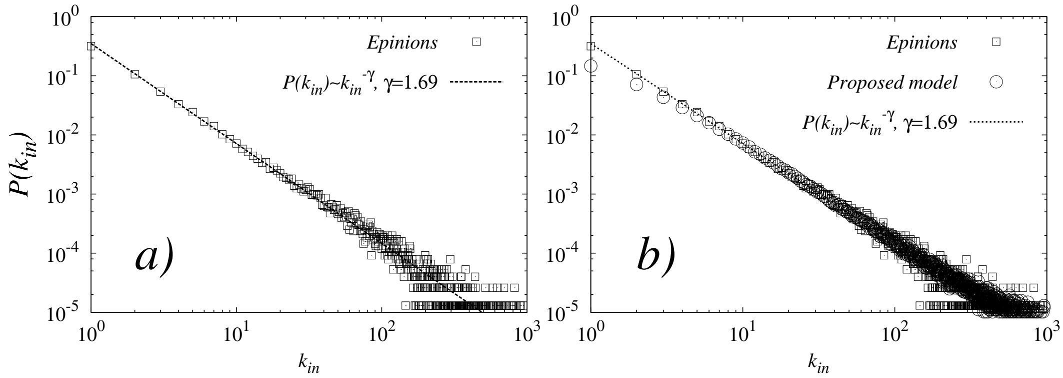

Figure 5.4: *a*) In-degree distribution of *Epinions* network. *b*) Comparison of the In-degree distribution of *Epinions* network with the obtained from the simulation of the proposed model (see text for details).

With the previous results, it is possible to deduce that the MLF model with the incorporation of more local processes is capable of reproduce some topological properties of real CN. On the other hand, even though the comparison between the network generated with the proposed model and the real CN *Epinions* indicates good fits in some properties, can not be ensured that the local processes incorporated by the proposed model are the only ones involved in the growth and evolution of the real *Epinions* network. Also it is not possible to ensure that the frequency of rewiring and addition of links that were define in the simulation are the same as in the evolution of this real network. However, the model can be a realistic simplification of some of these processes and therefore, the network generated with the proposed model has some properties close to those exhibited by the real *Epinions* network.

In the MLF model presented in the previous Chapter and extended in this Chapter is considered that all nodes born with the same amount of outgoing links, that is, is considered that the out-degree is a constant. This consideration is also present in many models proposed previously as the proposed by Barabási and Albert [17]. This contrast with several real networks where the Out-degree distribution follows a power law [11, 16, 36]. In order to approximate this type of out-degree distribution, in the next chapter are developed two growth models capable to generate directed complex networks with out-degree distributions that decay as power law $P(k_{out}) \sim k_{out}^{-\gamma_{out}}$.

Chapter 6

Out-degree distribution in Complex Networks

Among the topological properties of real CN, one of the most studied is the out-degree distribution. This property describes the probability that a node in the network has a particular number of outgoing links.

In the literature, there are many growth models for CN that reproduce some topological properties of real systems. [37] In most of these models, is assumed that all nodes are born with the same amount of outgoing links (*i.e.*, their out-degree is a constant), as in the model proposed by Barabási-Albert [38]. In other models, such as the one proposed by Dorogovtsev *et.al* [18] and the one proposed by Krapivsky and Redner [20], the out-degree distribution decays as an exponential or a poisson distribution, respectively. However, these results differ from the out-degree behavior of several real CN. For example, in metabolic networks [3], the Internet[5], and WWW[7] the out-degree decays as a power-law.

In order to approximate this type of out-degree distribution, some growth models for CN have been proposed. For example, Dorogovtsev *et.al.* [23] and Bollobás *et.al.* [24] have each developed a model capable of producing out-degree distributions that decay as a power-law with exponent $\gamma = 2 + \frac{n_r+n+B}{m}$ and $\gamma = 1 + \frac{1+\delta_{out}(\alpha+\beta)}{\beta+\gamma}$, respectively. Hence in both models the γ exponent is greater than 2.

In this chapter, are present two growth models capable to generate out-degree distributions with power law behavior. In the section 6.1, is present a model capable of generate complex networks with out-degree distribution that follows a power law $P(k_{out}) \sim k^{-\gamma_{out}}$ with $0 < \gamma_{out} < 1$. In the section 6.2, is present a different model capable of generate complex networks with out-degree distribution that follows a power law with $1 < \gamma_{out} < \infty$ and also is demonstrated that the proposed model is able to reproduce the out-degree distribution of the social network of Flickr users [16].

6.1 Model I

In this section is proposed a simple growth model for directed CN which is able to generate out-degree distributions that decay as a power-law with exponent $0 < \gamma_{out} < 1$.

6.1.1 Model details

In the proposed model, the growth of the network is done by adding nodes one at a time. At the beginning, only the node n_0 exists in the network and its out-degree is 0. Then is considered that the out-degree of any new node n_{new} added to the network is determined as follows:

- with probability p where $0 < p < 1$, n_{new} copies the out-degree of a randomly selected node from the network.
- with complementary probability $1 - p$, n_{new} randomly selects an out-degree uniformly distributed from 0 to N. That is, node n_{new} has out-degree $0, 1, 2, \ldots N$.

From the first rule, it is important to note that as the quantity Q_s of nodes with out-degree s increases, the probability that node n_{new} has out-degree s also increases to $\frac{Q_s}{N}$, where N is the total number of nodes in the network. In addition, due to the second rule new nodes may have out-degree of the order N.

6.1.2 Analytical Solution

In order to get an expression for the out-degree distribution generated by the proposed model, the continuum method [17] is used. The following differential equation describes the variation of the quantity Q_s of nodes with out-degree s with respect to the total number N of nodes in the network:

$$\frac{dQ_s(N)}{dN} = p\overbrace{\frac{Q_s(N)}{N}}^{g_1} + \overbrace{(1-p)\frac{1}{N+1}}^{g_2}, \tag{6.1}$$

term g_1 accounts for the situation that a new node copies the out-degree of a randomly selected node in the network. The term g_2 describes the random selection of out-degree for a new node.
Eq. 6.1 can be written in the standard form for a linear differential equation as follows:

$$\frac{dQ_s(N)}{dN} + \left(\frac{-p}{N}\right) Q_s(N) = \frac{1-p}{N+1}, \tag{6.2}$$

multiplying by the integrating factor $e^{-p\int \frac{1}{N}dN} = N^{-p}$, is obtained

$$N^{-p}Q_s(N) = (1-p)\int \frac{N^{-p}}{N+1}dN. \tag{6.3}$$

Since to the integral of Eq. 6.3 is not elementary, the solution retrieved is in terms of the Hypergeometrical Function ${}_2F_1$ [39] as follows:

$$Q_s(N) = {}_2F_1[1, 1-p; 2-p; -N]N + kN^p, \tag{6.4}$$

where k is a constant. To obtain the out-degree distribution $Q_s(N)$, Eq. 6.4 is solved for $s = 1$, $s = 2$, and so on as follows:

- for $Q_1(N)$, it should be considered the initial condition
$$Q_1(2) = \frac{1-p}{2}.$$
This initial condition is due to the fact that at the beginning, the network is formed only by node n_0 with no outgoing links, that is $N = 1$. For this case the quantity $Q_1(1)$ of nodes with out-degree $s = 1$ is zero ($Q_1(1) = 0$). When the node n_1 is added ($N = 2$), the probability for node n_1 to have out-degree $s = 1$ is $\frac{1-p}{2}$. Solving Eq. 6.4 for the initial condition $Q_1(2) = \frac{1-p}{2}$, one gets:

$$\begin{aligned} Q_1(N) =& {}_2F_1[1, 1-p; 2-p; -N]N \\ &+ \left[\frac{1-p}{2} - {}_2F_1[1, 1-p; 2-p; -2](2)\right] N^p 2^{-p}, \end{aligned} \tag{6.5}$$

- for $Q_2(N)$, it should be considered the initial condition
$$Q_2(3) = \frac{1-p}{3}.$$
This initial condition is due to the fact that, before adding node n_2, only nodes n_0 and n_1 are in the network ($N = 2$) and any of them has $s \geq 2$, therefore $Q_2(2) = 0$. When node n_2 is added ($N = 3$), the probability that node n_2 has out-degree $s = 2$ is $\frac{1-p}{3}$. Solving Eq. 6.4 for the initial condition $Q_2(3) = \frac{1-p}{3}$, is obtained:

$$\begin{aligned} Q_2(N) =& {}_2F_1[1, 1-p; 2-p; -N]N \\ &+ \left[\frac{1-p}{3} - {}_2F_1[1, 1-p; 2-p; -3](3)\right] N^p 3^{-p}. \end{aligned} \tag{6.6}$$

From the previous results in Eqs. 6.5 and 6.6, can be deduced that:

$$\begin{aligned} Q_s(N) =& {}_2F_1[1, 1-p; 2-p; -N]N \\ &+ \left[\frac{1-p}{s+1} - (s+1){}_2F_1[1, 1-p; 2-p; -(s+1)]\right] N^p (s+1)^{-p}. \end{aligned} \tag{6.7}$$

Normalizing Eq. 6.7 one gets

$$\begin{aligned} \frac{Q_s(N)}{N} =& {}_2F_1[1, 1-p; 2-p; -N] \\ &+ \left[\frac{1-p}{s+1} - (s+1){}_2F_1[1, 1-p; 2-p; -(s+1)]\right] N^{p-1} (s+1)^{-p}. \end{aligned} \tag{6.8}$$

Eq. 6.8, shows that the exponent γ_{out} of the out-degree distribution obtained with the proposed model is only determined by the probability p. That is, the out-degree distribution obtained decays as a power-law

$$\frac{Q_s}{N} \sim N^{p-1} s^{-p} \quad for \;\; 1 << s << N, \tag{6.9}$$

with exponent $\gamma_{out} = p$.

On the other hand, can be deduced that as a consequence of the random out-degree selection by new nodes with probability $1-p$ (second rule of the proposed model), the average out-degree of the nodes grows with the network size. To validate this hypothesis, it was calculated analytically the average out-degree $\bar{s}$ using the following differential equation:

$$\frac{d\bar{s}(N)}{dN} = (1-p)\left[\frac{\frac{N}{2} - \bar{s}(N)}{N+1}\right], \tag{6.10}$$

that describes the increment of the average out-degree $\bar{s}$ with respect to the total number N of nodes in the network. On the right-hand side of Eq. 6.10, the term $\frac{N}{2}$ describes the mean of the random out-degree uniformly selected from 0 to N by a new node. Thus, the term $\frac{N}{2} - \bar{s}(N)$ describes the increment of $\bar{s}$.
Eq. 6.10 can be written in the standard form for a linear differential equation as follows:

$$\frac{d\bar{s}(N)}{dN} + \frac{1-p}{N+1}\bar{s}(N) = \frac{(1-p)N}{2(N+1)}. \tag{6.11}$$

Solving Eq. 6.11 one gets

$$\bar{s}(N) = \frac{N(1-p)-1}{2(2-p)} + \frac{k}{(N+1)^{1-p}}. \tag{6.12}$$

As the total number of nodes in the network increases ($N >> 1$), we can approximate Eq.6.12 as follows:

$$\bar{s}(N) \approx \frac{N(1-p)}{2(2-p)}. \tag{6.13}$$

From Eq. 6.13 it can be seen that, effectively $\bar{s}$ grows proportionally to the network size, that is, in the proposed model the average out-degree of nodes tends to infinity when $N \to \infty$.

6.1.3 Validation of the Analytical Solution

In order to validate the analytical solutions for the out-degree distribution (Eq. 6.7) and average out-degree (Eq. 6.13) of the proposed model, four numerical simulations was performed using $p = 0.1$, $p = 0.3$, $p = 0.6$, and $p = 0.9$. In each simulation, was considered the growth of a directed network from 1 to 10^4 nodes. Figure 6.1 shows that the results of the numerical simulations and the analytical prediction

(Eq. 6.7) for the out-degree distribution fit appropriately. On the other hand, was calculated the average out-degree $\bar{s}$ in each simulation for different network sizes. Figure 6.2 shows that the average out-degree retrieved from the simulations and the analytical prediction (Eq. 6.11) fit also appropriately, that is $\bar{s}$ grows proportionally to the network size as stated by Eq. 6.11. It is important to note that when $p \to 0$ the value of $\bar{s}$ increments rapidly as the network grows ($N >> 1$), this happens because as $p \to 0$ the probability for random out-degree selection by new added nodes increases and the network tends to become dense. This contrasts with some large networks that are sparse where the number of edges is much smaller than the maximum possible and the average out-degree increases slowly as the network grows. [40] In this context, it is important to note that in the proposed model the average out-degree increases slowly as $p \to 1$.

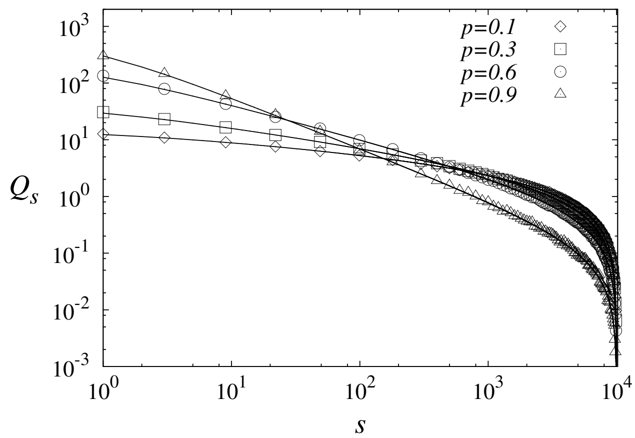

Figure 6.1: Comparison of the out-degree distribution (symbols) retrieved from the simulations and the analytical predictions (lines).

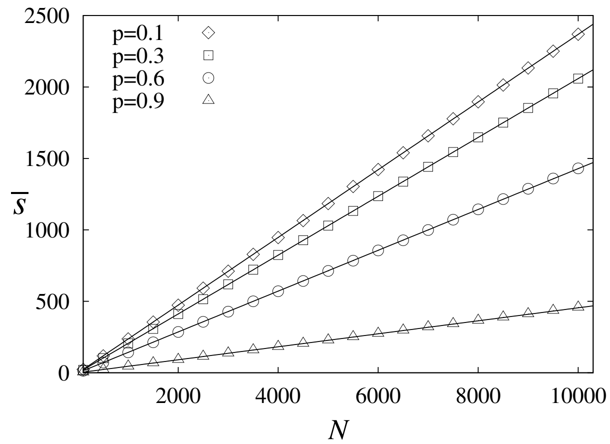

Figure 6.2: Comparison of the Average out-degree $\bar{s}$ retrieved from the simulations for different network sizes and the analytical predictions (lines).

6.2 Model II

At this time, have been proposed several models for growth of CN capable to generate out-degree distributions with power law behavior [23, 24, 29]. These models are not able to produce out-degree distributions with γ_{out} exponents in the range between 1 and 2. However, there are real CN where the γ_{out} exponent value is within this interval. For example, the social network of Flickr users [16], the Any Beat network [41], the online social network Epinions [16] and the network of flights between airports of the world (OpenFlights) [16] where the γ_{out} exponent for the out-degree distribution of these CN is close to 1.74, 1.71, 1.69 and 1.74 respectively.

In this section is introduced a new model for growth of directed CN that allows to obtain out-degree distributions that decay as a power-law with exponents in the range $1 < \gamma_{out} < \infty$. That is, the proposed model is able to generate all exponent values found in documented real CN. [3, 5, 7, 11, 16, 31, 41]

6.2.1 Model details

It has been demonstrated that the growth and evolution of CN is influenced by local processes that shape its topological and dynamical properties [25]. The model proposed in here incorporates two local processes for adding new nodes to the network: a random out-degree selection and a copy of an already present out-degree value. In many large networks the maximum degree of a node is much smaller than the number of nodes [16]. Thus, the proposed model assumes that the probability that a new node n_{new} selects a random out-degree decreases as the network grows. This probability is expressed as $N^{-\alpha}$ where N is the total number of nodes in the network (including n_{new}) and α is a constant greater than 0. In other words, the probability that new nodes have an out-degree close to N tends to zero as $N >> 1$.

In this model, the growth of the network is performed by adding nodes one at a time. At the beginning, only node n_0 is present in the network and its out-degree is 0. Then, the out-degree of any new node n_{new} added to this network is determined as follows:

- With probability $N^{-\alpha}$, n_{new} randomly selects an out-degree uniformly distributed from 0 to $N-1$. That is, n_{new} may have out-degree $0, 1, 2, \ldots, N-1$. It is important to notice that it is possible that n_{new} has an out-degree of the order of $N-1$.

- With complementary probability $1 - N^{-\alpha}$, n_{new} copies the out-degree of a randomly selected node from the network. It is important to notice that as the number Q_s of nodes with out-degree s increases, the probability that n_{new} has out-degree s also increases to $\frac{Q_s}{N-1}$.

6.2.2 Analytical Solution

It is possible to employ the continuum method [17] to obtain the analytical solution for the proposed model. This method is implemented using the following differential equation:

$$\frac{dQ_s(N)}{dN} = \overbrace{N^{-\alpha}\frac{1}{N}}^{g_1} + \overbrace{(1-N^{-\alpha})\frac{Q_s(N)}{N-1}}^{g_2} \tag{6.14}$$

The previous equation describes the variation of the number Q_s of nodes with out-degree s with respect to the total number N of nodes in the network. The term g_1 describes the situation that a new node randomly selects an out-degree value and the term g_2 the situation that a new node copies this value from a randomly selected node in the network.
Eq. 6.14 may be written in the standard form for a linear differential equation:

$$\frac{dQ_s(N)}{dN} + \frac{N^{-\alpha}-1}{N-1}Q_s(N) = \frac{N^{-\alpha}}{N}. \tag{6.15}$$

From Eq. 6.15, it is possible to deduce the integrating factor $I(N) = e^{\int \frac{N^{-\alpha}-1}{N-1}dN}$. Solving for $I(N)$ produces non elementary functions, which complicate the solution of Eq. 6.15. In order to obtain an integrating factor in terms of elementary functions, it is best to simplify Eq. 6.15 as follows:

$$\frac{dQ_s(N)}{dN} + \frac{N^{-\alpha}-1}{N}Q_s(N) = \frac{N^{-\alpha}}{N}. \tag{6.16}$$

This simplification has little implications for large values of N, because $N-1 \approx N$, as $N >> 1$. This allows to employ the following integrating factor: $I_2(N) = e^{\int \frac{N^{-\alpha}-1}{N}dN} = \frac{e^{-\frac{N^{-\alpha}}{\alpha}}}{N}$. Multiplying Eq. 6.16 by $I_2(N)$ produces:

$$\frac{e^{\frac{-N^{-\alpha}}{\alpha}}}{N}Q_s(N) = \int \frac{N^{-(\alpha+1)}e^{\frac{-N^{-\alpha}}{\alpha}}}{N}dN. \tag{6.17}$$

Solving for $Q_s(N)$

$$\frac{e^{\frac{-N^{-\alpha}}{\alpha}}}{N}Q_s(N) = \frac{e^{\frac{-N^{-\alpha}}{\alpha}}}{N} + \int \frac{e^{\frac{-N^{-\alpha}}{\alpha}}}{N^2}dN, \tag{6.18}$$

$$Q_s(N) = 1 + \frac{Ne^{\frac{N^{-\alpha}}{\alpha}}\Gamma\left(\frac{1}{\alpha}, \frac{N^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} + kNe^{\frac{N^{-\alpha}}{\alpha}}, \tag{6.19}$$

where k is a constant and $\Gamma(\cdot)$ is the incomplete Gamma function. In order to obtain the out-degree distribution $Q_s(N)$, it is necessary to solve Eq. 6.19 for $s = 1$, $s = 2$, and so on as follows:

- for $Q_1(N)$, consider the initial condition

$$Q_1(2) = \frac{2^{-\alpha}}{2};$$

this initial condition is due to the fact that, at the beginning the network only has one node, n_0, with no outgoing links ($N = 1$). When the next node, n_1, is added ($N = 2$), the probability that node n_1 has out-degree $s = 1$ is $\frac{2^{-\alpha}}{2}$.

Then, solving Eq. 6.19 for the initial condition $Q_1(2) = \frac{2^{-\alpha}}{2}$ produces:

$$\begin{aligned} Q_1(N) =& 1 + \frac{N e^{\frac{N^{-\alpha}}{\alpha}} \Gamma\left(\frac{1}{\alpha}, \frac{N^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} \\ &+ \left[2^{-(\alpha+1)} - 1 - \frac{2 e^{\frac{2^{-\alpha}}{\alpha}} \Gamma\left(\frac{1}{\alpha}, \frac{2^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} \right] \frac{N e^{\frac{N^{-\alpha}}{\alpha}} e^{\frac{-2^{-\alpha}}{\alpha}}}{2}, \end{aligned} \tag{6.20}$$

- for $Q_2(N)$, consider the initial condition

$$Q_2(3) = \frac{3^{-\alpha}}{3},$$

this initial condition is due to the fact that, before adding node n_2 only n_0 and n_1 exist in the network ($N = 2$) and both have $s < 2$, therefore $Q_2(2) = 0$. When n_2 is added ($N = 3$), the probability that node n_2 has out-degree $s = 2$ is $\frac{3^{-\alpha}}{3}$.

Then, solving Eq. 6.19 with the initial condition $Q_2(3) = \frac{3^{-\alpha}}{3}$, one obtains:

$$\begin{aligned} Q_2(N) =& 1 + \frac{N e^{\frac{N^{-\alpha}}{\alpha}} \Gamma\left(\frac{1}{\alpha}, \frac{N^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} \\ &+ \left[3^{-(\alpha+1)} - 1 - \frac{3 e^{\frac{3^{-\alpha}}{\alpha}} \Gamma\left(\frac{1}{\alpha}, \frac{3^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} \right] \frac{N e^{\frac{N^{-\alpha}}{\alpha}} e^{\frac{-3^{-\alpha}}{\alpha}}}{3}. \end{aligned} \tag{6.21}$$

From the results in Eqs. 6.20 and 6.21, it is possible to deduce that:

$$\begin{aligned} Q_s(N) = 1 + \frac{N e^{\frac{N^{-\alpha}}{\alpha}} \Gamma\left(\frac{1}{\alpha}, \frac{N^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} \\ + \left[(s+1)^{-(\alpha+1)} - 1 - \frac{(s+1) e^{\frac{(s+1)^{-\alpha}}{\alpha}} \Gamma\left(\frac{1}{\alpha}, \frac{(s+1)^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} \right] \frac{N e^{\frac{N^{-\alpha}}{\alpha}} e^{\frac{-(s+1)^{-\alpha}}{\alpha}}}{(s+1)}. \end{aligned} \tag{6.22}$$

Normalizing Eq. 6.22, yields:

$$P_s(N) = \frac{1 + \frac{Ne^{\frac{N^{-\alpha}}{\alpha}}\Gamma\left(\frac{1}{\alpha},\frac{N^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} + \left[(s+1)^{-(\alpha+1)} - 1 - \frac{(s+1)e^{\frac{(s+1)^{-\alpha}}{\alpha}}\Gamma\left(\frac{1}{\alpha},\frac{(s+1)^{-\alpha}}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}}\right]\frac{Ne^{\frac{N^{-\alpha}}{\alpha}}e^{\frac{-(s+1)^{-\alpha}}{\alpha}}}{(s+1)}}{N}. \tag{6.23}$$

Eq. 6.23 describes the out-degree distribution $P_s(N)$ obtained with the proposed model for $1 < s < N$. It can also be noted that, as $s \to N$, Eq. 6.23 predicts that $P_s(N) \approx \frac{1}{N^{\alpha+2}}$. That is $P_s(N)$ decays to 0 rapidly as $s \to N$ and $N >> 1$, therefore the power-law behavior exhibits a cut-off (Figure 6.3a).

In order to obtain the scaling exponent of the out-degree distribution, terms $\Gamma(\cdot)$ into Eq. 6.23 are simplified using:

$$\Gamma(a,x) = \Gamma(a) - \gamma(a,x),$$

where $\gamma(a,x)$ and $\Gamma(a,x)$ are the lower and upper incomplete Gamma functions, respectively. By the following asymptotic property:

$$\gamma(a,x) \to \frac{x^a}{a} \; if \; x \to 0,$$

it is possible to write:

$$\Gamma(a,x) = \Gamma(a) - \frac{x^a}{a} \; if \; x \to 0. \tag{6.24}$$

Using Eq. 6.24 it is possible rewrite the $\Gamma(\cdot)$ terms of Eq. 6.23 as follows:

$$\Gamma\left(\frac{1}{\alpha},\frac{N^{-\alpha}}{\alpha}\right) \to \Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{N} \quad for \;\; N >> 1, \tag{6.25}$$

$$\Gamma\left(\frac{1}{\alpha},\frac{(s+1)^{-\alpha}}{\alpha}\right) \to \Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{s+1} \quad for \;\; s >> 1. \tag{6.26}$$

Substituting Eqs. 6.25 and 6.26 into Eq. 6.23 and considering that $s+1 \approx s$ as $s >> 1$, Eq. 6.23 can be expressed as:

$$\begin{aligned} P_s(N) \approx & \frac{1}{N} + \frac{\left[\Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{N}\right] e^{\frac{N^{-\alpha}}{\alpha}}}{\alpha^{1-\frac{1}{\alpha}}} \\ & + \left[s^{-(\alpha+1)} - 1 - \frac{se^{\frac{s^{-\alpha}}{\alpha}}\left[\Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{s}\right]}{\alpha^{1-\frac{1}{\alpha}}}\right] \frac{e^{\frac{N^{-\alpha}}{\alpha}} e^{\frac{-s^{-\alpha}}{\alpha}}}{s}, \end{aligned} \tag{6.27}$$

$$P_s(N) \approx \frac{1}{N} + \frac{\left[\Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{N}\right] e^{\frac{N^{-\alpha}}{\alpha}}}{\alpha^{1-\frac{1}{\alpha}}} + \left[s^{-(\alpha+2)} e^{\frac{-s^{-\alpha}}{\alpha}} - \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} + s^{-1}\left(1 - e^{\frac{-s^{-\alpha}}{\alpha}}\right)\right] e^{\frac{N^{-\alpha}}{\alpha}}. \tag{6.28}$$

Using the two first terms of the series expansion of $e^{-\frac{s^{-\alpha}}{\alpha}} \approx 1 - \frac{s^{-\alpha}}{\alpha}$ in Eq. 6.28 and simplifying

$$P_s(N) \approx \frac{1}{N} + \frac{\left[\Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{N}\right] e^{\frac{N^{-\alpha}}{\alpha}}}{\alpha^{1-\frac{1}{\alpha}}} - \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} e^{\frac{N^{-\alpha}}{\alpha}} + \left[\frac{1}{s} - \frac{1}{\alpha s^{\alpha+1}} + \frac{1}{\alpha}\right] e^{\frac{N^{-\alpha}}{\alpha}} s^{-(\alpha+1)}; \tag{6.29}$$

for $s >> 1$, $\left[\frac{1}{s} - \frac{1}{\alpha s^{\alpha+1}} + \frac{1}{\alpha}\right] \to \frac{1}{\alpha}$, thus it is possible to rewrite Eq. 6.29 as:

$$P_s(N) \approx \frac{1}{N} + \frac{\left[\Gamma\left(\frac{1}{\alpha}\right) - \frac{\alpha^{1-\frac{1}{\alpha}}}{N}\right] e^{\frac{N^{-\alpha}}{\alpha}}}{\alpha^{1-\frac{1}{\alpha}}} - \frac{\Gamma\left(\frac{1}{\alpha}\right)}{\alpha^{1-\frac{1}{\alpha}}} e^{\frac{N^{-\alpha}}{\alpha}} + \frac{e^{\frac{N^{-\alpha}}{\alpha}}}{\alpha} s^{-(\alpha+1)}. \tag{6.30}$$

Furthermore, in the limit when $N \to \infty$, Eq. 6.30 takes the form

$$P_s \approx \frac{s^{-(\alpha+1)}}{\alpha}. \tag{6.31}$$

Eq. 6.31 shows that the out-degree distribution obtained with the proposed model decays as a power-law $P_s \sim s^{-\gamma_{out}}$ for $1 < s < N$ with scaling exponent $\gamma_{out} = \alpha + 1$.

6.2.3 Validation of the Analytical Solution

To validate the analytical solution of the model as described by Eq. 6.23, four experiments were executed using $\alpha = 0.5, 1, 1.5$ and 2. Each of these experiments simulated the growth of a directed network from $N = 1$ to 10^4 nodes. Figure 6.3b shows that the out-degree distribution produced by these experiments and the analytical predictions by Eq. 6.23 fit appropriately.

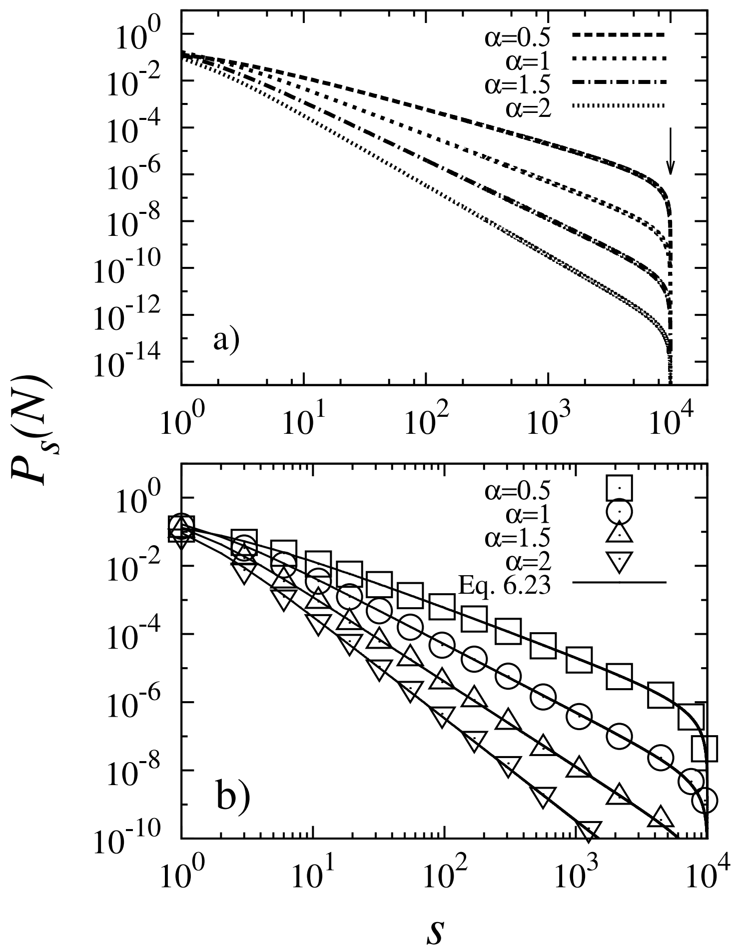

Figure 6.3: a) Analytical solution of the proposed model (Eq. 6.23 in dashed lines) for $N = 10^4$ and different values of α. Notice how the proposed model is able to obtain out-degree distribution P_s that decays as power law. Also, it is possible to note that for values of s close to N the P_s decay rapidly (vertical arrow) and the power law behavior is cut-off. b) Comparison of the out-degree distribution produced by the experiments (symbols ⊡, ⊙, △, ▽) and the analytical prediction in Eq. 6.23 (solid line) for $N = 10^4$ and several values of α.

6.2.4 Comparison with real networks.

To verify that the proposed model is able to reproduce the out-degree distribution of real CN, the social network of Flickr users [16] was selected.

In this network, the users correspond to the nodes and their friendship connections to the links. This network has 2, 302, 925 nodes and 33, 140, 017 links. Figure 6.4a shows that the out-degree distribution of the nodes in the Flickr network decay as a power-law distribution with $\gamma_{out} \approx 1.74$. Figure 6.4b shows that the model proposed by Eq. 6.23 with $\alpha = 0.74$ and $N = 2,302,925$ reproduces appropriately the out-degree distribution of the Flickr network for $s > 1$.

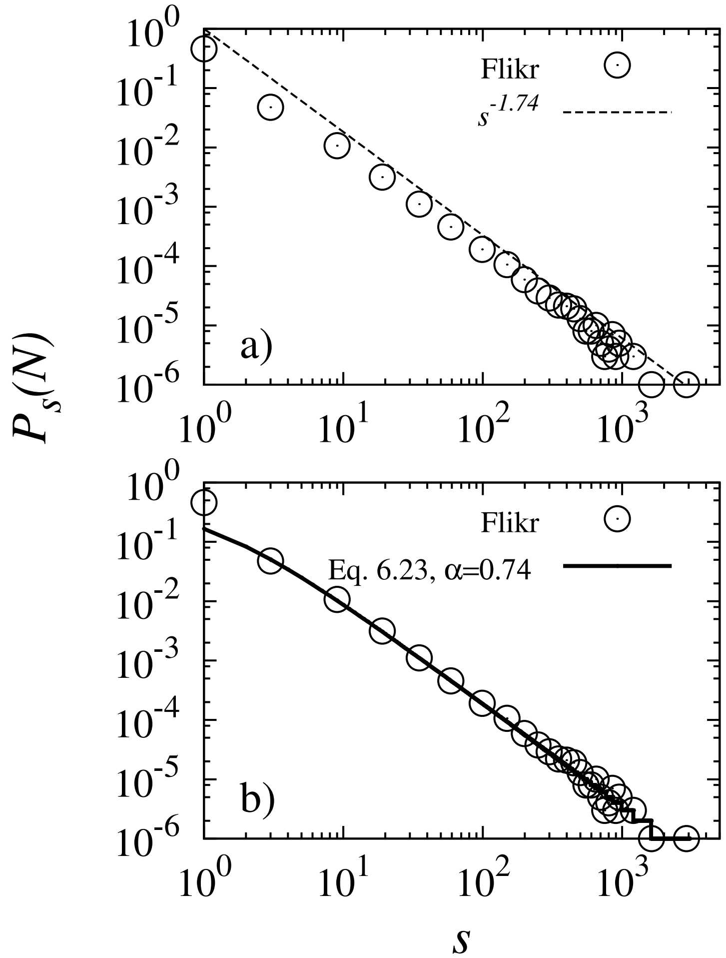

Figure 6.4: a) Out-degree distribution of the Flickr social network. b) Comparison of the out-degree distribution produced by the proposed model (Eq. 6.23) with $\alpha = 0.74$ and $N = 2,302,925$ and the actual out-degree distribution of the Flickr social network.

Although this model produces a good fit with the out-degree distribution of a real network, it is not possible guarantee that the local processes incorporated in this model are the only ones involved in the behavior of the out-degree distribution of the nodes in this network. Unknown processes may help to explain why for $s = 1$, this model does not fit. However, the proposed model provides a simplification of these processes and therefore, reproduces the out-degree distribution of the network.

In many growth models proposed at this time, as the proposed by Barabási *et.al.* [17], Dorogovtsev *et.al.* [18, 21, 34], Krapivsky *et.al.* [19, 20], Amaral *et al.* [22] and Esquivel *et al.* [27] is considered that all nodes in the network form only one component called Giant island, and the new nodes added to the network always connect to it. However, exists real networks as the US patents [15] comprised by a set of islands and its island size distribution follows a power law [42]. In order to approximate this property, a new growth model is developed in the next Chapter.

CHAPTER 7

Islands in Complex Networks

Many growth models have been proposed with the aim of reproducing some topological properties of real Complex Networks (CN) [37], for example the degree distribution, [17] in-degree distribution [20, 27, 28] or out-degree distribution. [24, 29, 30] However, little has been studied about the *Island Size Distribution* (I_s) which describes the number of islands with s nodes. An island is a set of nodes which is not connected to the rest of the network.

Previous models [17, 18, 19, 20, 21, 22, 24, 27, 34] consider that each node added to the network always connects to existing nodes. In other words, all the nodes in these models form a single island, which contains all the nodes of the network. However, in some real networks, as in the U.S. patent's citation network [15], the nodes form more than one island and the I_s follows a power-law: $I_s \sim s^{-\gamma}$ [42].

It is hypothesized that a possible cause for the origination of islands in some real complex networks is that, during network growth some nodes may be born with zero out-going links (*i.e.* patents without references to other patents) and this causes new islands to be generated.

In order to reproduce this property, in this chapter is proposed a CN growth model able to obtain I_s with a power-law behavior.

7.1 Model details

In the proposed model, it is considered that the birth of new islands is governed by the probability Φ considering two cases:

- 1) $\Phi = \frac{1}{N}$, where N is the number of nodes in the network. In this case it is considered that the probability of a new island is born decreases as the number of nodes in the network increases. This idea is mapped from real networks as follows: In a network of cites of scientific papers, when there are few papers (nodes), it is more probable that a new paper does not cite other papers (generating a new island) because it addresses an entirely new scientific theme. Conversely, when the quantity of papers increases, the probability that a new paper addresses an entirely new theme decreases, thus the probability of generating a new island also decreases.

- 2) $\Phi = p$, where $0 < p < 1$. In this case it is considered that the probability of a new island is born remains constant during the whole life of the network.

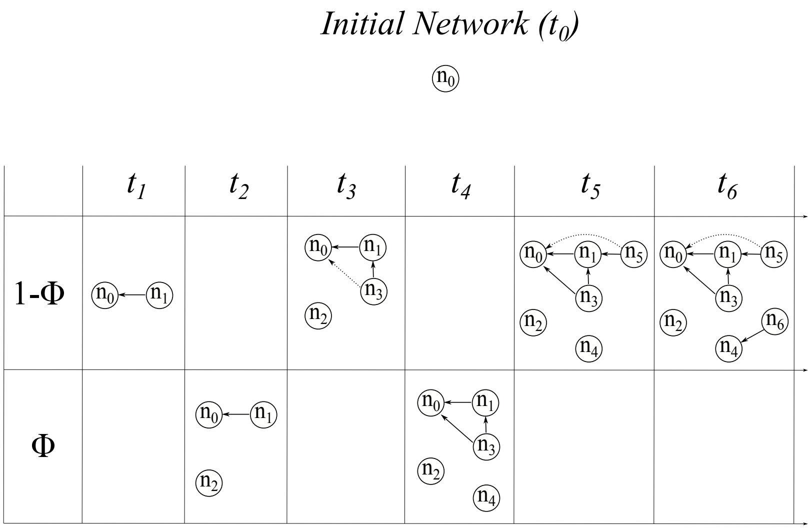

Figure 7.1: Growth of a directed network using the proposed model. At the beginning (t_0), only node n_0 exists in the network. At the next time step (t_1) node n_1 is added to the network and it is assumed that connects to node n_0. In t_2, node n_2 is added and it is assumed that it does not connect to any node, thus a new island is generated. In t_3, n_3 is added and it is assumed that connects to n_1 and n_0 (dashed arrow) because n_0 has an incoming link from n_1. In t_4, n_4 generates a new island as n_2 in t_2. At t_5 and t_6, nodes n_5 and n_6 are added and connect to the network as n_3 did at t_3.

In the model, the growth of the network is performed by adding one node at each time step. At the beginning, only node n_0 exists in the network and for each new node n_{new} added to the network, either one of the following operations is performed:

1. With probability Φ, n_{new} does not connect to any node in the network. That is, n_{new} generates a new island (see Fig. 7.1).
2. With complementary probability $1 - \Phi$, n_{new} randomly selects a node n_r and connects to it, as well as to all nodes that have one incoming link from n_r (see Fig. 7.1).

7.2 Analytical Solution

The continuum method [17] is employed to obtain the analytical solution for the I_s using the following differential equation:

$$\frac{dI_s(N)}{dN} = \overbrace{\Phi\delta_{s,1}}^{g_1} + (1-\Phi)\overbrace{\left(\underbrace{\frac{s-1}{N}I_{s-1}(N)}_{g_2'} - \underbrace{\frac{s}{N}I_s(N)}_{g_2''}\right)}^{g_2}. \tag{7.1}$$

Eq. 7.1 describes the variation of the number I_s of islands with s nodes with respect to the total number N of nodes in the network. Term g_1 describes the birth of a new island; that is, it models the situation that a new node n_{new} does not connect with any node (first rule of this model). The term g_2 depicts the second rule of the model, term g_2' describes the situation that a new node n_{new} randomly selects a node n_r belonging to an island with $s-1$ nodes and connects to it, thus $I_s(N)$ increases. The term g_2'' describes the situation that a new node n_{new} randomly selects a node n_r belonging to an island with s nodes and connects to it, thus $I_s(N)$ decreases.

Eq. 7.1 may also be written in the standard form for a linear differential equation:

$$\frac{dI_s(N)}{dN} + \frac{(1-\Phi)s}{N} I_s(N) = \frac{(1-\Phi)(s-1)}{N} I_{s-1}(N) + \Phi\delta_{s,1}. \tag{7.2}$$

In order to investigate the impact that $\Phi = \frac{1}{N}$ and $\Phi = p$ have in I_s, Eq. 7.2 is solved for each one of them. For $\Phi = \frac{1}{N}$, Eq. 7.2 takes the form:

$$\frac{dI_s(N)}{dN} + \frac{(N-1)s}{N^2} I_s(N) = \frac{(N-1)(s-1)}{N^2} I_{s-1}(N) + \frac{\delta_{s,1}}{N}. \tag{7.3}$$

In order to obtain the $I_s(N)$, Eq. 7.3 is solved for $s = 1$, $s = 2$, and so on. For $s = 1$, Eq. 7.3 takes the form:

$$\frac{dI_1(N)}{dN} + \frac{N-1}{N^2} I_1(N) = \frac{1}{N}; \tag{7.4}$$

solving Eq. 7.4 gives:

$$I_1(N) = 1 - \frac{E_i\left(\frac{1}{N}\right)}{Ne^{\frac{1}{N}}} + \frac{k}{Ne^{\frac{1}{N}}}, \tag{7.5}$$

where k is a constant and $E_i(\cdot)$ is the exponential integral. As $N >> 1$, Eq. 7.5 can be approximated as:

$$I_1(N) \approx 1. \tag{7.6}$$

Solving Eq. 7.3 for the following s values produces:

$$I_s(N) \approx \frac{1}{s}. \tag{7.7}$$

That is, with $\Phi = \frac{1}{N}$ the proposed model is able to produce *Island Size distributions* with a power-law behavior $I_s \sim s^{-\gamma}$ for $1 < s < N$ with fixed exponent $\gamma = 1$.

For $\Phi = p$, Eq. 7.2 takes the form:

$$\frac{dI_s(N)}{dN} + \frac{(1-p)s}{N} I_s(N) = \frac{(1-p)(s-1)}{N} I_{s-1}(N) + p\delta_{s,1}. \tag{7.8}$$

In order to obtain the $I_s(N)$, Eq. 7.8 is solved for $s = 1$, $s = 2$, and so on. For $s = 1$, Eq. 7.8 takes the form:

$$\frac{dI_1(N)}{dN} + \frac{(1-p)}{N} I_1(N) = p. \tag{7.9}$$

Solving Eq. 7.9 gives:

$$I_1(N) = \frac{pN}{(1-p)+1} + \frac{k}{N^{1-p}}, \tag{7.10}$$

where k is a constant. As $N >> 1$, Eq. 7.10 can be approximated as:

$$I_1(N) \approx \frac{pN}{(1-p)+1}. \tag{7.11}$$

Solving Eq. 7.8 for the following s values it is possible to deduce that:

$$I_s(N) \approx \frac{(s-1)!(1-p)^{s-1}pN}{\prod_{x=1}^{s}[x(1-p)+1]}. \tag{7.12}$$

Approximating with the Gamma Function $\Gamma(\cdot)$ is obtained:

$$\begin{aligned} I_s(N) \approx & \frac{\Gamma\left(\frac{1}{1-p}\right) pN}{(1-p)^2} \frac{\Gamma(s)}{\Gamma(s+1+\frac{1}{1-p})} \\ \approx & \frac{\Gamma\left(\frac{1}{1-p}\right) pN}{(1-p)^2} s^{-(1+\frac{1}{1-p})} \quad for\ s >> 1. \end{aligned} \tag{7.13}$$

From Eq. 7.13, when $\Phi = p$ the model is able to produce *Island Size distributions* with a power-law behavior $I_s \sim s^{-\gamma}$ for $1 < s < N$ with exponent $\gamma = 1 + \frac{1}{1-p}$. This allows γ to take values from 2 to ∞ when $\Phi = p$.

In order to obtain the analytical solution for the in-degree distribution generated with the proposed model, the continuum method is used. [17] Hence, the differential equation that describes the in-degree distribution may be written as follows:

$$\begin{aligned} \frac{dQ_i(N)}{dN} = & \overbrace{(1-\Phi)\left(\underbrace{\frac{Q_{i-1}(N)}{N}}_{g_1'} + \underbrace{(i-1)\frac{Q_{i-1}(N)}{N}}_{g_1''}\right)}^{g_1} \\ & \overbrace{-(1-\Phi)\left(\underbrace{\frac{Q_i(N)}{N}}_{g_2'} - \underbrace{i\frac{Q_i(N)}{N}}_{g_2''}\right)}^{g_2} + \overbrace{(1-\Phi)\delta_{i,0}}^{g_3} + \overbrace{\Phi\delta_{i,0}}^{g_4}. \end{aligned} \tag{7.14}$$

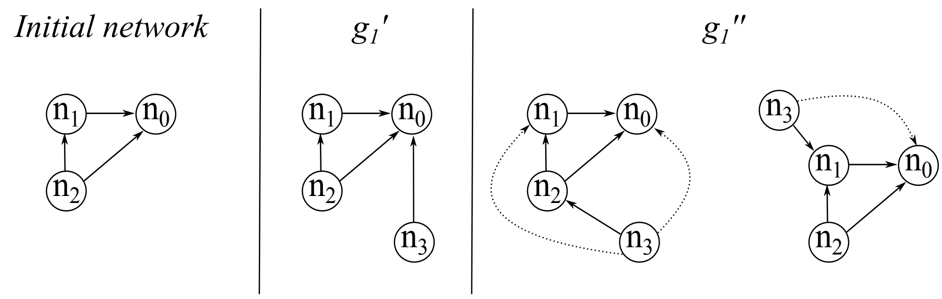

Figure 7.2: Consider a network comprising of three nodes (n_0, n_1, n_2). In this network, the in-neighbors of n_0 are n_1 and n_2. Also the number of nodes with three incoming links is $Q_3 = 0$. There are two possible ways to increase Q_3: 1) A new node n_3 randomly selects node n_0 and connects to it (g'_1 in the figure and Eq. 7.14) thus $Q_3 = 1$; 2) A new node n_3 randomly selects an in-neighbor of n_0 and connects to it (solid line) and to n_0 (dashed line), as stated by g''_1 in Eq. 7.14 and this figure.

Eq. 7.14 describes the variation of the number Q_i of nodes with i incoming links with respect to the number N of nodes in the network. The term g_1 describes how the number of nodes with i incoming links increases, g'_1 describes how a new node n_{new} randomly selects a node n_r with $i-1$ incoming links and connects to it, and g''_1 describes how n_{new} randomly selects an in-neighbor of a node n_j that has $i-1$ incoming links and connects to it (see Fig. 7.2), thus Q_i increases. The term g_2 describes how the number of nodes with i incoming links decreases, terms g'_2 and g''_2 perform similar functions as g'_1 and g''_1. Finally, the terms g_3 and g_4 models the effect of adding a new node with zero incoming links using the second and the first rule of the model.

Eq. 7.14 may be written in the standard form for a linear differential equation:

$$\frac{dQ_i(N)}{dN} + (1-\Phi)\frac{(i+1)Q_i(N)}{N} = (1-\Phi)\frac{iQ_{i-1}(N)}{N} + \delta_{i,0}. \tag{7.15}$$

In order to analyze the impact that $\Phi = \frac{1}{N}$ and $\Phi = p$ have in Q_i, Eq. 7.15 is solved for each one of them. For $\Phi = \frac{1}{N}$, Eq. 7.15 takes the form:

$$\frac{dQ_i(N)}{dN} + \frac{N-1}{N^2}(i+1)Q_i(N) = \frac{N-1}{N^2}iQ_{i-1}(N) + \delta_{i,0}. \tag{7.16}$$

Solving Eq. 7.16 for some i values it is possible to deduce that:

$$Q_i(N) \approx \frac{N+1}{(i+1)(i+2)}. \tag{7.17}$$

That is, with $\Phi = \frac{1}{N}$ the proposed model is able to produce *In-degree distributions* with a power-law behavior $Q_i \sim i^{-\gamma}$ for $1 < i < N$ with fixed exponent $\gamma = 2$. This result was previously obtained by Krapivsky and Redner. [20]

For $\Phi = p$, Eq. 7.15 takes the form:

$$\frac{dQ_i(N)}{dN} + \frac{1-p}{N}(i+1)Q_i(N) = \frac{1-p}{N} i Q_{i-1}(N) + \delta_{i,0}. \tag{7.18}$$

Solving Eq. 7.18 for several i values produces:

$$Q_i(N) \approx \frac{(i)!(1-p)^i N}{\prod_{x=1}^{i+1}[(x+1) - xp]}. \tag{7.19}$$

Approximating with the Gamma Function $\Gamma(\cdot)$ one gets:

$$\begin{aligned} Q_i(N) \approx & \quad \frac{N\Gamma\left(\frac{1}{1-p}\right)}{(p-1)^2} \frac{\Gamma(i+1)}{\Gamma(i+2+\frac{1}{1-p})} \\ \approx & \quad \frac{N\Gamma\left(\frac{1}{1-p}\right)}{(p-1)^2} i^{-(1+\frac{1}{1-p})} \quad for\ i >> 1. \end{aligned} \tag{7.20}$$

Therefore, if $\Phi = p$ the proposed model is able to produce *In-degree distributions* with a power-law behavior $Q_i \sim i^{-\gamma}$ for $1 < i < N$ with exponent $\gamma = 1 + \frac{1}{1-p}$. This allows γ to take values from 2 to ∞ when $\Phi = p$.

7.3 Validation of the Analytical Solution

In order to validate the analytical predictions for I_s (Eq. 7.7, Eq. 7.13) and Q_i (Eq. 7.17, Eq. 7.20), four experiments were performed. The experiments simulated the growth of a directed network from $N = 1$ to 10^4 nodes following the proposed model. Fig. 7.3 shows the comparison of I_s produced by the experiments and the analytical predictions, and it is showed that both fit appropriately. Fig. 7.4 shows the comparison of Q_i produced by the experiments and the analytical predictions, and it is showed that both fit appropriately.

It is important to mention that in this model the case when Φ increases as the number of nodes increases is not considered. This is because when N is large enough, new nodes added to the network would have high probability of not connecting to other nodes, thus generating new islands. Therefore, the resulting network would be composed by a great number of isolated nodes.

Also, it is not considered the situation that a new node can connect to nodes presents in different islands, resulting in the fusion of two or more islands. These case will be included in future work.

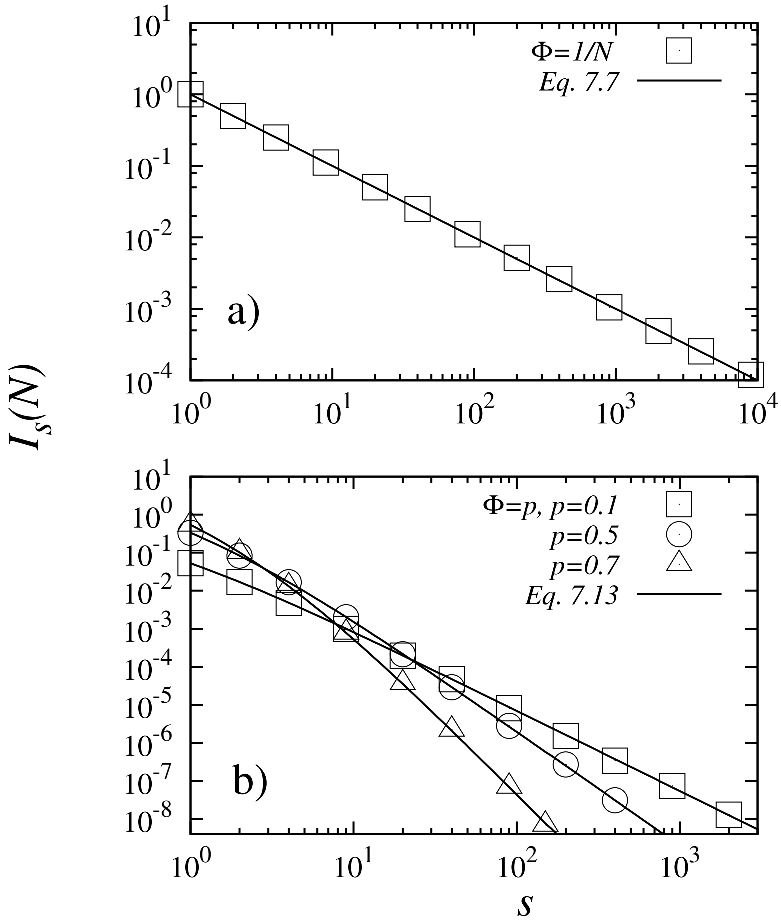

Figure 7.3: Comparison of the I_s obtained experimentally (symbols $\boxdot$ $\odot$ $\triangle$) and the analytical predictions (solid line). *a*) Using $\Phi = \frac{1}{N}$. *b*) Using $\Phi = p$, with $p = 0.1, 0.5$, and 0.7.

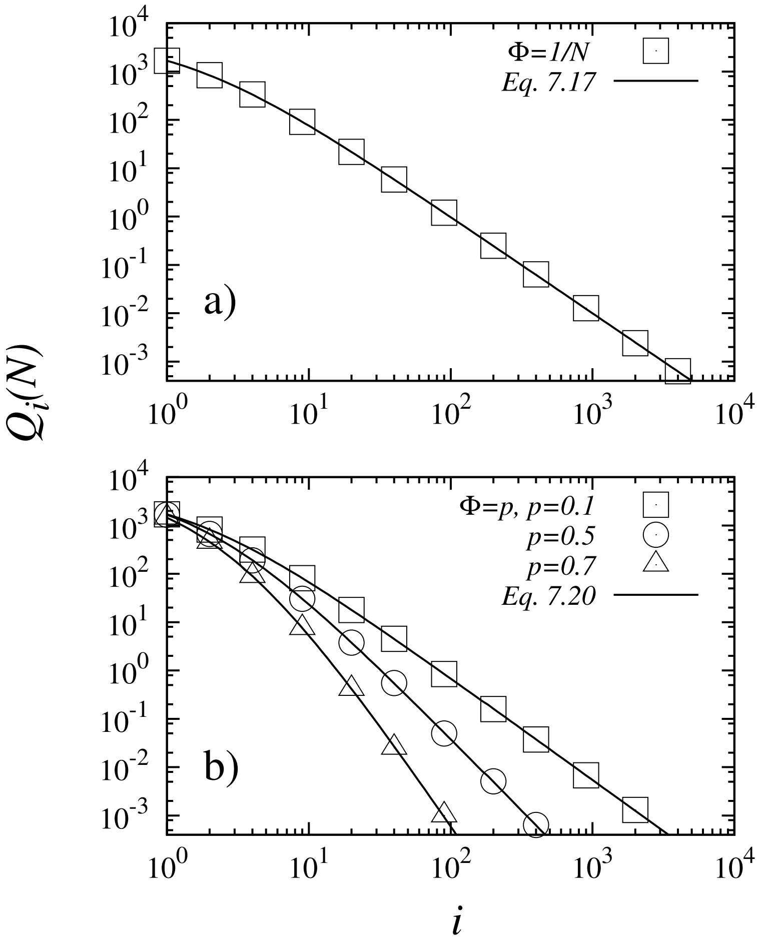

Figure 7.4: Comparison of the Q_i obtained experimentally (symbols ⊡ ⊙ △) and the analytical predictions (solid line). *a*) Using $\Phi = \frac{1}{N}$. *b*) Using $\Phi = p$, with $p = 0.1, 0.5$, and 0.7.

Chapter 8

Discussion

It has been demonstrated that the growth and evolution of CN is influenced by local processes that shape its topological and dynamical properties [25]. In this Thesis, have been proposed new growth models for complex networks that incorporate some local process.

In the Chapter 4 was presented a growth model named MLF, that incorporates the initial attractiveness, prohibition of multiple links and constant out-degree. The model is capable to generate in-degree distributions with power-law behavior with exponent range from 1 to ∞. It was also shown that the model is capable of generate some properties of the real complex network comprising flights between airports of the world [16].

In Chapter 5 was presented a model that incorporates the initial attractiveness, prohibition of multiple links, constant out-degree, addition and rewiring of links with constant probability. That is, this model extends of MLF model. The model is capable of generate in-degree distributions with power-law behavior as MLF model. It was also shown that the model is capable to reproduce some properties of the social network *Epinions* [16].

Two models capable of generate out-degree distributions with power-law behavior were presented in Chapter 6, the models include mainly two process: random out-degree selection and copy of out-degree. The first model, is capable of generate exponents in the range from 0 to 1. The second model is capable of generate exponents in range from 1 to ∞, it also was demonstrated that the second model is capable of reproduce the out-degree distribution of the $Flickr$ social network [16].

In chapter 7 was present a model capable of generating Island size and In-degree distributions with power-law behavior. The model includes three process, the constant born of islands, the born of islands dependent of the time, and the copying of links. The model generates exponents in the range from 2 to ∞ in both Island size and In-degree distributions.

In general, in this thesis has been made comparisons between some topological properties of networks generated with the proposed models and those of real complex networks, and in all the cases were obtained good approximations. Despite of this, it can not be assure that the local processes incorporated by the proposed models are the only ones involved in the growth and evolution of each real network. However, it may be that each corresponding model is a *realistic simplification* of some of these processes and therefore, the generated networks have some properties close to those exhibited by each real network. On the other hand, in this thesis have been studied the impact of some sets of local process separately. However, it would be interesting develop a growth model that incorporate all local process studied in this thesis and investigate their impact in the topological properties of the networks generated.

Chapter 9

Conclusions

In this Thesis have been proposed some growth models for Complex Networks with the aim of study the impact that several local processes have in the topological properties of this class of networks. The proposed models incorporate local processes, such as initial attractiveness, preferential attachment, prohibition of multiple links, addition and rewiring of links, constant out-degree, random out-degree and out-degree copying.

One of the main characteristics of many real complex networks is that do not have multiple links. A contribution of this thesis was the study of the impact that the prohibition of multiple links have in some topological properties of the complex networks. With this aim a growth model was developed, the networks generated with the model showed that the prohibition of multiple links in joint with other local processes may be responsible of the existence of real complex networks with exponents $\gamma < 2$ in its In-degree distribution.

Other important characteristic of many real complex networks is that the out-degree distribution follows a power-law. However many of the proposed growth models at this time consider a constant out-degree, other models generate out-degree distributions with exponential or poisson behavior. In this thesis are proposed two growth models that generate out-degree distribution with power-law behavior, the models incorporate the random out-degree selection and the out-degree copying processes. The first model can generate exponents in the range from 0 to 1. The second model is able to obtain exponents in the range from 1 to ∞. That is, this model is capable to generate all exponents found in the out-degree distributions of the real complex networks.

Also, in this thesis is investigated the island size distribution that describes the probability for an island to have a determined amount of nodes from a network. It has been found that in some real complex networks the Island size distribution follows a power law. In this context, a new model capable of reproduce this property is proposed. The proposed model includes the copy of links, the random and time-dependent born of islands processes. With these local processes the model is capable to generate In-degree and Island Size distributions with power-law behavior with exponent γ tunable from 2 and ∞.

In summary, in this Thesis have been proposed new growth models for complex networks. With the models it is possible to obtain all exponents γ found in the Out-degree and In-degree distribution of real complex networks that are documented. That is, the models are able to generate exponents in the range from 1 to ∞. Also have been demonstrated that some of the models are able to reproduce other properties of complex networks as Clustering, Shortest-path and Diameter.

Bibliography

[1] A. Barrat, M. Barthélemy and A. Vespignani. Dynamical Processes on Complex Networks, Cambridge University Press 2008. (Cited on page 3.)

[2] P. Erdos and A. Renyi ”On Random Graphs I”, Publicationes Mathematicae 6: 290-297 (1959). (Cited on page 6.)

[3] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási. The large-scale organization of metabolic networks, Nature 407, 651-655 (2000). (Cited on pages 6, 41 and 46.)

[4] A.L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert and T. Vicsek. Evolution of the social network of scientific collaborations, Physica A 311, 590-614(2002). (Cited on page 6.)

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology, ACM SIGCOMM, Cambridge, MA, September 1999. (Cited on pages 6, 11, 12, 41 and 46.)

[6] B. A. Huberman and L. A. Adamic. Growth Dynamics of the World Wide Web, Nature 401, 131 (1999). (Cited on page 6.)

[7] R. Albert, H. Jeong, A.L. Barabási. Diameter of the World-Wide Web, Nature 401, 130 (1999). (Cited on pages 6, 41 and 46.)

[8] M.E.J. Newman. Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46, 323-351 (2005). (Cited on pages 6, 9 and 31.)

[9] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks, Nature 393, 440-442 (1998). (Cited on page 6.)

[10] S.N. Dorogovtsev and J.F.F. Mendes. The shortest path to complex networks, arXiv:cond-mat/0404593v4 [cond-mat.stat-mech] (2004). (Cited on page 6.)

[11] R. Albert, A.-L. Barabási. Statistical mechanics of complex networks, Reviews of Modern Physics 74, 47-97 (2002). (Cited on pages 7, 9, 31, 39 and 46.)

[12] S. N. Dorogovtsev and A. V. Goltsev. Critical phenomena in complex networks, Rev. Mod. Phys., Volume 80, 1275-1335, October-December 2008. (Cited on page 7.)

[13] M. E. J. Newman. The structure and function of complex networks, *SIAM Review 45*, 167-256 (2003). (Cited on page 7.)

[14] http://www3.nd.edu/∼networks/resources.htm (Cited on pages 7 and 16.)

[15] B. H. Hall, A. B. Jaffe and M. Trajtenberg (2001). "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498. (Cited on pages 7, 16, 53 and 55.)

[16] J. Kunegis. KONECT - The Koblenz Network Collection, In Proc. Int. Web Observatory Workshop, pages 1343-1350, (2013). (Cited on pages 7, 9, 16, 17, 30, 31, 38, 39, 41, 46, 51 and 63.)

[17] A.L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks, *Physica A* 272, 173-187 (1999). (Cited on pages 8, 9, 15, 37, 39, 42, 47, 53, 55, 56 and 58.)

[18] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Structure of Growing Networks with Preferential Linking, *Phys. Rev. Lett.* **85**, 4633-4636 (2000). (Cited on pages 8, 10, 15, 16, 18, 25, 26, 41, 53 and 55.)

[19] P.L. Krapivsky, S. Redner. Phys. and F. Leyvraz. Connectivity of Growing Random Networks, *Phys. Rev. Lett.* **85** 4629 (2000). (Cited on pages 8, 10, 53 and 55.)

[20] P.L. Krapivsky, and S. Redner. Network growth by copying, Phys. Rev. E **71**, 036118 (2005). (Cited on pages 8, 11, 41, 53, 55 and 59.)

[21] S.N. Dorogovtsev and J.F.F. Mendes. Effect of the accelerating growth of communications networks on their structure, *Phys. Rev. E* **63**, 025101(R) (2001). (Cited on pages 8, 11, 15, 53 and 55.)

[22] L.A.N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks, Proc. Natl. Acad. Sci. U.S.A. 97, 11 149 (2000). (Cited on pages 8, 13, 15, 53 and 55.)

[23] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin. WWW and Internet models from 1955 till our days and the "popularity is attractive" principle. *ArXiv:cond-mat/0009090* (2000). (Cited on pages 8, 41 and 46.)

[24] B. Bollobás, C. Borgs, J. Chayes and O. Riordan. Directed scale-free graphs, *SODA'03* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; pp. 132-139 (2003). (Cited on pages 8, 41, 46 and 55.)

[25] R. Albert and A.L. Barabási. Topology of evolving networks: local events and universality *Phys. Rev. Lett.* **85** 5234 (2000). (Cited on pages 8, 12, 15, 46 and 63.)

[26] S. N. Dorogovtsev and J.F.F. Mendes. Scaling behaviour of developing and decaying networks, Europhys. Lett. 52, 33. (Cited on pages 8, 12 and 15.)

[27] J. Esquivel-Gómez, R. E. Balderas-Navarro, E. Ugalde and J. Acosta-Elías. On the growth of directed complex networks with preferential attachment: Effect

upon the prohibition of multiple links, Int. J. Mod. Phys C. **26**, 6, (2015). (Cited on pages 8, 34, 36, 38, 53 and 55.)

[28] J. Esquivel-Gómez, P. D. Arjona-Villicaña and J. Acosta-Elías. The impact of local processes and the prohibition of multiple links in the topological properties of directed complex networks, Int. J. Mod. Phys C. **26**, 12, (2015). (Cited on pages 8 and 55.)

[29] J. Esquivel-Gómez, E. Stevens-Navarro, U. Pineda-Rico and J. Acosta-Elías. A growth model for directed complex networks with power-law shape in the out-degree distribution, *Sci. Rep.* **5**, 7670; DOI:10.1038/srep07670 (2015). (Cited on pages 8, 46 and 55.)

[30] J. Esquivel-Gómez, P. D. Arjona-Villicaña, E. Stevens-Navarro, U. Pineda-Rico, R. E. Balderas-Navarro and J. Acosta-Elias On a growth model for complex networks capable of producing power-law out-degree distributions with wide range exponents, Sci. Rep. 5, 9067; DOI:10.1038/srep09067 (2015). (Cited on pages 8 and 55.)

[31] H. Ebel, L.-I. Mielsch and S. Bornholdt. Scale-free topology of e-mail networks, *Phys. Rev. E* **66**, 035103(R) (2002). (Cited on pages 9, 31 and 46.)

[32] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajalopagan, R. Stata, A. Tomkins, J. Wiener. Graph structure in the Web, Computer Networks: The International Journal of Computer and Telecommunications Networking, 33 (2000) (Cited on pages 11 and 12.)

[33] A.L. Barabási, H. Jeong, E. Ravasz, Z. Néda, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations, Physica A, 311(3-4), 590-614. (Cited on page 11.)

[34] S. N. Dorogovtsev and J.F.F. Mendes. Evolution of networks with aging of sites, Phys. Rev. E 62, (2000). (Cited on pages 13, 15, 53 and 55.)

[35] The CAIDA AS Relationships Dataset, <20120601> http://www.caida.org/data/active/as-relationships/ (Cited on page 17.)

[36] http://snap.stanford.edu/data/ (Cited on pages 17 and 39.)

[37] A.L. Barabási. Scale-Free Networks: A Decade and Beyond. *Science* **325**, 412 (2009). (Cited on pages 41 and 55.)

[38] A.L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science* **286**, 509-512 (1999). (Cited on page 41.)

[39] E. W. Weisstein. "Hypergeometric Function." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/HypergeometricFunction.html (Cited on page 42.)

[40] J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters,*ACM Transactions on Knowledge Discovery from Data*, Vol. 1, No. 1 (2007). (Cited on page 45.)

[41] M. Fire, R. Puzis and Y. Elovici. Link Prediction in Highly Fractional Data Sets, *Handbook of Computational Approaches to Counterterrorism*, (2012). (Cited on page 46.)

[42] V. Batagelj. Efficient Algorithms for Citation Network Analysis, ArXiv:cs/0309023v1. (Cited on pages 53 and 55.)