Structure and dynamics of evolving complex networks

Ewan Colman

A Thesis Presented for the Degree of Doctor of Philosophy

Department of Mathematical Sciences

School of Information Systems, Computing and Mathematics

Brunel University

September 2014

Abstract

The analysis of large disordered *complex* networks has recently received enormous attention motivated by both academic and commercial interest.

The most important results in this discipline have come from the analysis of stochastic models which mimic the growth and evolution of real networks as they change over time. The purpose of this thesis is to introduce various novel processes which dictate the development of a network on a small scale, and use techniques learned from statistical physics to derive the dynamical and structural properties of the network on the macroscopic scale.

We introduce each model as a set of mechanisms determining how a network changes over a small period in time, from these rules we derive several topological properties of the network after many iterations, most notably the degree distribution.

1. In the first mechanism, nodes are introduced and linked to older nodes in the network in such a way as to create triangles and maintain a high level of clustering. The mechanism resembles the growth of a citation network and we demonstrate analytically that the mechanism introduced suffices to explain the power-law form commonly found in citation distributions.
2. The second mechanism involves edge rewiring processes - detaching one end of an edge and reattaching it, either to a random node anywhere in the network or to one selected locally.
3. We analyse a variety of processes based around a novel fragmentation mechanism.
4. The final model concerns the problem of finding the electrical resistance across a network. The network grows as a random tree, as it grows the distribution of resistance converges towards a steady state solution. We find an application of the relatively recent concept of a random Fibonacci sequence in deriving the rate of convergence of the mean.

Acknowledgments

I have many people to thank for making the completion of this thesis possible.

For financial support I thank both EPSRC for my scholarship and also Brunel University, Department of Mathematical Sciences for offering it to me. I am exceedingly grateful to Professor Geoff Rodgers for being an inspirational Ph.D supervisor, for his guidance and immense expertise in the subjects I have chosen to research. I would also like to acknowledge the staff at the department of Mathematical Sciences who have all been incredibly helpful and supportive. In particular my second supervisor Dima Savin, head of department and former head, Steve Noble and Tony Rawlins respectively, research director Paresh Date, and the administrative support: Kay Lawrence, Frances Foster and Linda Campbell. I thank Diane Smith for proofreading my work. Finally, I owe many thanks to my family and friends for all the love and encouragement they have given.

Contents

List of publications

Chapter 2

Colman, ER. and Rodgers, GJ., *Complex scale-free networks with tunable power-law exponent and clustering*, Physica A: Statistical Mechanics and its Applications 392 (21): 5501-5510 2013

Colman, ER. and Rodgers, GJ., *Clustered scale-free networks*, Poster for the European Conference on Complex Systems 2013

Chapter 3

Colman, ER. and Rodgers, GJ., *Local rewiring rules for evolving complex networks*, Physica A: Statistical Mechanics and its Applications 419 (0): 80-89 2014

Chapter 4

Colman, ER. and Rodgers, GJ., *Kinetics of node splitting in evolving complex networks*, Physica A: Statistical Mechanics and its Applications 391 (24): 6626-6631 2012

Chapter 5

Colman, ER. and Rodgers, GJ., *The resistance of randomly grown trees*, JOURNAL OF PHYSICS A-MATHEMATICAL AND THEORETICAL 44 (50): 505001-505015 2011

Foreword

I originally chose to study the discipline of complex networks because I believed that it demonstrated how mathematics and mathematical modelling could be used in a wide variety of interesting and novel applications. The breadth of the subject areas that this discipline intersects has by far exceeded my expectations; over the course of my Ph.D I have found myself reading about topics as diverse as linguistics, sociology, ecology, biology, economics ... the list goes on. While my technical abilities have certainly improved, I have equally developed an enormous appreciation of the relevance of this emerging science and its importance in the world beyond the boundaries of academic institutions.

Structure of this thesis

The introduction is divided into two sections. The first is a brief review of complex network literature relevant to the later chapters. It was not my intention to write a comprehensive history of the field, instead I have given a personal perspective on the significant contributions, hopefully imparting some of my curiosity and motivation to study this area. The second part introduces some of the technical background that I consider to be essential to the study of complex networks, again focusing on the aspects that have interested me the most and are relevant to the later chapters. Chapters 2 to 5 detail the four major projects that have led me to find original results. These chapters are mostly independent of each other aside from the central idea, modelling networks which evolve over time, which connects the entire thesis.

Chapter 1

Introduction

1.1 History and Motivation

Academic interest in the dynamics of complex networks has grown in tandem with the role of networked systems in society.

Networks have been studied in one form or another ever since Leonard Euler solved the Konnigsburg bridge problem (see section 1.2). However, the recent ascension of network science, as a topic distinct from any other, such as graph theory, began with a wave of research activity in the late 1990s. In hindsight, it seems unsurprising that large disordered networks began to receive so much attention around this time - access to the Internet was becoming commonplace in developed countries causing unprecedented effects on communication and commerce. The World-Wide Web became an entity in itself, unregulated and self-organised, an example of what we now call a complex network. In the space of just a few years, networks became part of peoples' lives on a scale that could not have been anticipated.

The era saw swathes of programming literate entrepreneurs capitalising on the need for intelligent ways to manage the overload of information. Google achieved this through their search engine, and in particular their clever use of data regarding the hyper-link structure of the World-Wide Web. The concept behind the PageRank algorithm, which was originated by Jon Kleinberg around 1998 and adapted by Larry Page and Sergey Brin while Ph.D candidates, is based on a simple idea: each web-page is ranked proportionally to the number of times we would visit it if we were to click on random hyper-links for a very long time [1, 2]. Each time you do a Google search you are using information about the structure of the entire

network to find the web-page which is, in one respect, the most central. Measures of centrality like this are a widely discussed concept in network science today [3].

As the Internet enabled the collection and sharing of data to an extent not experienced before, opportunities were increasingly found to apply scientific methods to phenomena outside of labs and observatories. In their seminal 1998 paper, Duncan Watts and Steven Strogatz analysed data that had been collected and compiled by film enthusiasts in the Internet Movie Database (imdb.com), and for the first time used a mathematical model to analyse the structure of a social network, in this case the network of collaborations between movie actors (actors are linked if they appeared in at least one movie together) [4]. One of their motivations was to test the idea of 'six degrees of separation', the commonly discussed notion that every person on the planet is connected to every other by a chain of at most six friendship links. One ambitious attempt to prove this had previously been made by the psychologist Stanley Milgram when he sent letters out to random people across the USA, each letter contained instructions asking whoever received it to pass it on to someone they knew who they thought might be closer to the target, the target being one particular person in Boston [5].

Watts and Strogatz saw that the problem could be approached with a greater degree of scientific rigor. They conceived the social network as a purely mathematical object, this allowed them to reference some well known ideas in graph theory and define certain topological properties such as the clustering coefficient (which we discuss in detail later). They found that each actor is linked to every other by an average of 3.65 collaborations but this result is arbitrary compared to the wider impact that the paper had. They demonstrated that the 'small world' property of social networks also exists in power grids and in the neural network of a worm's brain. Clustering and short path lengths have since been repeatedly found in empirically studied networks; so many, in fact, that these properties are now considered universal features that connect many complex networks across nature and society. The paper is currently one of the most cited physics articles of all time.

1.1a Social physics

Of course this wasn't the first attempt by physicists to build a model of human behaviour[1], but it did inspire many people from scientific backgrounds to turn their attention to large scale social systems. The area of physics that lends itself most easily to analysing these systems is statistical mechanics, a field that historically has focused on describing the nature of gases and magnets.

In both examples, gases and magnets, we are faced with the fundamental problem of having too many entities contributing to the system's behaviour. In classical mechanics, if the trajectory of a particle is known then it is possible to predict its position and trajectory at a later time. When two particles are present we can in theory determine their trajectories after they collide. But when 10^{23} particles are involved[2] it becomes impossible to know the initial position of each one, and even if it was known, then the idea that we can compute their positions at any point in the future is also unrealistic.

Remarkably though, because of work founded by James Clerk Maxwell and Ludwig Boltzmann in the late nineteenth century, many results have been found. The 'trick' they discovered is to forget about the precise details of each particle's position and velocity, and instead focus on the amount of energy each particle is carrying. The fact that the total of this energy has an upper bound (i.e. the first law of thermodynamics) means it is possible to mathematically derive the statistical property of how many particles have x amount of energy. It might seem that throwing away information about the motion of each particle would have a considerable effect on the outcome, but on such a large scale it is generally found that these assumptions and approximations yield accurate and useful results.

A similar philosophy has been adopted for the study of magnets. In a strip of magnetic material each atom carries its own spin (its magnetic orientation) which influences the spins of other atoms close to it. The Ising model is a very basic

[1] This achievement should probably be attributed to Adolphe Quetelet [6].

[2] This is the order of magnitude of Avagadro's number, the number of atoms in one mole.

model that emulates this process and is one of the classic examples of how simple rules governing the microscopic behaviour of the atoms can lead to complex behaviour on the macroscopic scale. Here, atoms are arranged in a grid-like structure and only interact with their nearest neighbours. The grid naturally wants to be in a low energy state, and this can only be achieved when a large number of the neighbouring pairs have the same alignment[3]. This criterion produces complex patterns of alignment that sometimes resemble the end state in the board games Othello and Go [7]. As with these games it is difficult to predict the final outcome, simulating the problem on a computer is one way to find a result, but without that resource the problem of finding an exact configuration of alignments must be put aside in favour of finding the *probability* that the system will end up in a particular configuration.

Much of this is analogous to the problems faced in large scale social systems. To identify the individual behaviour of each person is next to impossible, even more so to then use this information to predict some future state, despite the huge databases and computational power we have today. If we are to understand the dynamics of social systems we need to take a leap of faith: to postulate that human beings are no more intricate, no more complex, than the particles in a gas or the atoms inside a magnet, at least not for the intents and purposes of mathematics.

The social systems I am referring to here are varied; so far I have mentioned the World-Wide Web, the network of movie actors, and the power grid as examples of complex networks. In each case the observed complex structure is the consequence of human decisions rather than physical forces. Despite this we still attempt a 'physical' description of the network. Instead of atoms or particles, we deal with web-sites, actors and power substations. One of the first things we observe about these systems is that they grow. Once we have a hypothesis suggesting exactly how that growth is happening then we can begin to employ the methods of statistical mechanics to see if the observed topological properties emerge.

[3]Specifically this is true only in a *ferromagnet*. In an *antiferromagnet* the opposite is true and the lowest energy is achieved when neighbouring pairs have opposing alignment.

1.1b Growing networks

Every introductory book or course on complex systems will mention both 'preferential attachment' and 'scale-free networks'. These are terms that were both coined in Réka Albert and Albert-László Barabási's 1999 paper which presents a mechanism (preferential attachment) by which networks grow, and a description of their topology (scale-free) after a long period of growth [8]. The technical details are discussed in the following section. These concepts have been a foundation for extensive subsequent research (including a good proportion of my own research). Along with Watts and Strogatz's small world paper, it has been cited by seemingly every article that mentions complex networks. It is hard to imagine what the state of research in this discipline would look like if this work had never been published; however, the concept had been proposed 23 years earlier without a fraction of the impact.

The 'preferential attachment' concept first originated in 1976 by Derek De Solla Price[4] who was interested in the network of citations in scientific literature [10]. The interest in citation analysis began in the 1950s with Eugene Garfield [11]; Price used several databases of scientific papers compiled by Garfield and others which recorded each reference from one paper to another. His statistical analysis showed that the distribution of the number of references to a paper follows a power-law, also known as a scale-free distribution for reasons we explain in detail in the next section, for now we refer to the examples Price himself gave:

> "It seems that, in any given year, about 35 percent of all the existing papers are not, cited at all, and another 49 percent are cited only once. This leaves about 16 percent of the papers to be cited an average of about 3.2 times each. About 9 percent are cited twice; 3 percent, three times; 2 percent, four times; 1 percent, five times; and a remaining 1 percent, six times or more."

Consider also the fact that there are papers with thousands of citations. Price explained the extreme asymmetry of the distribution by relating it to the principle of

[4] In his paper he cites the 1955 paper of Herbert Simon, here a similar model gives an explanation for power-law distributions in general [9].

cumulative advantage, also known as the Matthew effect[5], a virtuous cycle or the rich-get-richer phenomenon. He formulated a mathematical model where, in the spirit of statistical mechanics, he disregarded the complex details, i.e. the subject matter of every individual paper, and considered a single mechanism designed to mimic the way in which references are typically made. The process goes as follows: papers are added one at a time, each newly added paper references a number of old papers that are randomly selected. Additionally, to invoke the Matthew effect, the probability that any given paper is cited is proportional its current number of citations.

It is the 'what if' type of questions that are of most interest to theoretical physicists. The question asked here is what if the only factor that influences an authors decision whether to cite or not is the current level of success that the paper has already received? The answer Price found was that we get a citation distribution that looks remarkably similar to the real one.

The network of citations re-emerged in 1998 with another empirical study, this time by Sidney Redner [13]. Following this came the inevitable studies of the structure of the World-Wide Web [14]. They found that the distribution of pages linking to any given page follows a similar pattern to the citation network. This work was followed shortly by Barabási and Albert's reintroduction of the cumulative advantage/preferential attachment model [8]. While Price's model was seen at the time as an important contribution to Scientometrics (the meta-analysis of science), this time around it was apparent that the applications were much wider reaching, it ignited a big bang of complex network research that has continued to expand until the present day. In Chapter 2 we look extensively at citation networks and citation distributions.

[5]This term was first used in academic literature by the sociologist Robert K. Merton [12]. The concept was inspired by a passage in the Gospel According to Matthew "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath.".

1.1c Epidemics

Before the age of the Internet the word 'network' would most likely bring to mind images of rail maps, roads, or electrical networks. Common to each of these examples is that they are built by humans and therefore supposedly have an element of intelligent design, engineered to serve a purpose: to optimise transport of one form or another. What about networks that do not have a designer? The examples of complex networks we have so far discussed are of disordered self organising systems; and likewise, they aid the the flow of disorganised and unpredictable forces: information, computer viruses and disease.

Around the turn of the millennium online security was a major issue. The Internet - the physical infrastructure that supports the transmission of data - had no centralised controller to prevent it from growing like an out of control weed. This was also a time when viruses were prevalent and often smarter than the software employed to mitigate their threat. Following the discovery made by the Faloutsos brothers in 1999 that the network was in fact scale-free [6], academics in the complex network field began to apply their knowledge to this area [16]. Network science and epidemiology have become interlinked ever since.

Barabási and Albert followed up their breakthrough work with a paper that asked the 'what if' question of what would happen in the event of a failure in one or a number of nodes, how will this effect the flow of information? As seen in citation distributions, scale-free networks consist of a few very well connected nodes called 'hubs', these channel a huge proportion of the information flow. The vast majority of nodes have very few connections making them insignificant. Consequently, the system was found to be robust against failures given that the failures do not happen in one of the hubs. Targeted attacks, however, can cause catastrophic damage [17].

[6] Some doubt whether this is in fact true since the method by which they obtained their data has been brought into question [15]

It is natural to assume that the viruses themselves could be analysed in a similar way to the viruses of biology. The way a computer virus spreads from an infected machine to those connected to it via a network router is in many ways comparable to the movement of a biological infection by physical contact. The paper of Romualdo Pastor-Satorras and Alessandro Vespignani in 2000 introduced the concept of scale-free networks to epidemic modelling [18]. Previously, the standard methods involved either imagining each person in a grid-like structure who only infect those closest to them[7], or imagining that anyone can infect anyone else in the system with equal likelihood. Using data from real computer virus infections it was shown that the old models were inaccurate, in general they suggested the existence of an epidemic threshold (a level of transmissibility above which the virus will take over the system), but with a more realistic scale-free topology, it was discovered, this level does not exist - any virus will take over the system, irrespective of how contagious it is, given enough time.

The policies governments adopt when faced with an impending epidemic are now very much guided by network structure. For example, the spread of an epidemic can potentially be limited by manipulating the network of flight paths. Unfortunately however, cancelling flights and placing quarantine zones in airports comes at a considerable economic cost, it is therefore important to improve our understanding of the associated risk factors [19]. Moreover, epidemic models are frequently used to predict viral marketing campaigns, opinion dynamics leading up to elections, and the general diffusion of articles, images and other media that are shared through the World-Wide Web.

1.1d Current trends

The most popular sites on the Internet are those that facilitate and encourage communication and sharing. In the last ten years, sites like YouTube, Facebook and Twitter have helped to decentralise the sources of media and news. Most notably, the Arab uprisings in 2010 demonstrated the shift from the old paradigm of

[7]Recall that the Ising model, as well as many other models studied in traditional statistical mechanics, was originally constructed on a similar simple lattice.

centrally organised groups to the disordered and self-organised protest movements enabled by social media. Cultural phenomena too, such as Gangnam Style (currently the most watched video on YouTube), owe their success to a high level of local transmissibility in the epidemic sense, more than to any marketing campaign. Consequently, a lot of analysis has been applied to the data collected by these online platforms, ultimately the goal is to accurately predict human behaviour and social trends [20, 21].

Many of the large Silicon Valley companies now have their own research labs, blurring the distinction between research and software development. Data collection is instant, and comes in a constant stream, recording users' activity at every moment in time. While this presents the ominous threat of an Orwellian surveillance society, we can also hope that the data may be used to create better search engines, better content discovery systems, can aid security and create new communication channels. To summarize, the same factors that contributed most to the recent interest in networked systems i.e. the Internet, the World-Wide Web, and the availability of data, still exist and are becoming more relevant to people and to society each day.

1.1e Other areas

Meanwhile, complex networks have been infiltrating many areas of interdisciplinary research. This trend is in line with a general movement away from reductionism, i.e. studying each individual component of a system in greater and greater detail, towards the approaches that embrace the interconnectedness of the system. Biology has many examples. The functions of the mind are not localised in different spatial areas of the brain but are caused by the interactions between many different circuits of neurones [22]. Models are being formulated which describe the growth of the neural circuitry of the C. Elegans worm, suggesting the possibility that the human brain could one day be understood as a complex network [23]. Conservationists are finding ways to balance ecosystems which are dependent on the complex topology of food webs [24, 25]. The relationship between genetic sequences and disease depends on many complex interactions of individual genes, recent work has mapped these interactions and analysis has begun on the network produced [26, 27].

1.2 Modelling complex networks

1.2a Network

The use of the word 'network' is relatively recent, whereas the concept has been mathematically studied for centuries under the name *graph*. A graph is a set of objects called *vertices* (although in physics literature the more commonly used term is *node*) and a set of pairs of nodes called *edges* (can also use *arc*, *link*), if the order of the pairs is relevant then we say that the edges are *directed*.

The terms vertex and edge are an artifact of the origins of graph theory since these are the terms traditionally used to describe the geometric features of shapes[8]. Leonard Euler, who is credited with being the founder of graph theory, also published results in geometry. He proved, for example, that the formula $F-E+V=2$ applied to all convex polyhedra, where F, E and V are the number of faces, edges, and vertices of the shape respectively. In 1735, Euler solved the Königsberg bridge problem, a popular puzzle for tourists to attempt when visiting the German city, the challenge was to cross each of the city's seven bridges once and only once [6]. He showed that it was impossible using an abstract representation of the town where each land mass is a 'vertex' and each bridge is an 'edge'. This is widely accepted as the beginning of graph theory and subsequently complex networks.

1.2b Degree

The number of edges adjacent to a node is its *degree*. We could also define the degree of a node as the number of its neighbours. The *degree sequence* is the sequence of degrees corresponding to the set of nodes and is crucial to answering many questions about a network, including the following solution (found by Euler) to the Königsberg bridge problem. In general, if we are to make a continuous path through a graph, one that visits every edge only once, then each time we visit a node we need to leave that node via a different edge than the one by which we

[8]Although I am not aware of when these terms were first used, they can be found in translations of Euclid's Elements

arrived. Barring the first and last nodes we visit, there must be one edge to exit the node for every edge that takes us in. This puts a constraint on the degrees of the nodes: for the path to be possible only two nodes can have odd degree, the rest must be even. The graph that corresponds to the bridges of Königsberg does not meet this condition.

When we observe networks in the real world, the notion of 'degree' almost always corresponds to a meaningful quantity that we would like to know, usually it is some measure of importance, popularity or success. This is certainly the case in academia, where the success of a scientific paper is measured by the number of citations it has received from other scientific papers. The networks formed by mapping scientific papers and the references from one to another has received considerable attention in the complex network literature[9].

The mathematical questions in complex networks are generally focused on an analysis of a network's topology. Topology (which can also trace its roots back to Euler's geometry work) is the branch of mathematics concerned with the geometric properties of objects that do not depend on distances. It is therefore a word that encompasses the variety of properties and quantities associated with networks. Here we define some topological properties and give some examples to demonstrate their importance.

1.2c Degree distribution

Suppose we have a network $G(N, E)$ of N nodes and E edges. Let k be a non-negative natural number, then the *degree distribution* of G is the number of nodes whose degree is k.

The degree distribution is a useful metric to classify and discriminate different observed networks. The discovery of the universality of the power-law form of

[9]There are a few possible reasons why this might be: the relative simplicity of the process of citing, the availability of data, and the desire of academics to understand the dynamics of their own success certainly contribute.

degree distributions has provided an unexpected connection between many diverse complex systems. If we denote the number of nodes of degree k by n_k then the result for large values of k

$$n_k \sim k^{-\gamma}, \tag{1.1}$$

where γ is positive and usually between 2 and 3, has been found in citation networks [28, 13], food webs [29], word webs [30], to name three of many empirical studies. The diversity of these applications suggests a universality common to self organising networked systems, emphasising the importance of understanding networks in their own right, independent of their application. Moreover, the discovery of a power-law degree distribution is not a trivial result. Suppose that the connections in a graph are entirely random, then the degree distribution is completely different. *Random graphs* of this form were first studied by the prolific mathematician Paul Erdős and his collaborator Alfréd Rényi - the graph $G(N, p)$ of N nodes with edges existing between any node pair with probability p is (now) known as the Erdős-Rényi graph, its degree distribution is Poisson in the large N limit

$$n_k = N\frac{(Np)^k e^{-Np}}{k!}. \tag{1.2}$$

It is important to highlight the differences between Eq.(1.1) and Eq.(1.2) since they expose the differences between a random system, where there is only disorder, and a complex system where the disorder is in some way guided by underlying principles. Firstly, the distribution Eq.(1.1) has a long tail. In the network this translates to extremely well connected 'hub' nodes that are unlikely to be found in a random configuration. Second, the Poisson distribution has a bell curve form, i.e. the node degrees are distributed close to and around the mean. This is not the case with the power-law distribution, here the mean contains little information about the network and the smallest possible degree is also the most frequently found. The power-law distribution also has the intriguing property of scale-invariance. A function is scale-invariant if it keeps its form after re-scaling, Eq.(1.1) meets this condition since

$$n_{\lambda k} \sim \lambda^{-\gamma} n_k. \tag{1.3}$$

This ties the distribution to the field of fractal geometry and the concept of *self-*

similarity. A geometric object (that is infinite in size and also has infinite resolution) is *self-similar* if and only if it can be mapped onto itself with a change of scale [31]. The name 'scale-free network' is given to any network with a power-law degree distribution, however, while some of these networks have the property of self-similarity, most, including those created by preferential attachment, do not [32].

Also important are the moments of the degree distribution

$$M_n = \sum_{n=0}^{\infty} k^n n_k. \tag{1.4}$$

The first moment M_1 is a useful quantity as it is equal to the total degree of the network, or simply twice the total number of edges. Consequently, the mean degree is given by $\langle k \rangle = M_1/N$.

1.2d Clustering coefficient

Let i be a node in a network with degree k_i. Let E_i be the number of edges between the set of neighbours of i. Clearly $0 \geq E_i \geq \binom{N}{2}$. The *clustering* C_i of a node i is

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{1.5}$$

The *clustering coefficient* of a network is the average clustering over all of its nodes.

This definition comes from Watts and Strogatz although a very similar metric was introduced much earlier by social scientists looking at friendship networks [33]. The small-world model seeks to have a disordered network where the clustering coefficient can be tuned to a range of values by adjusting a single parameter. The model interpolates somewhere between a regular lattice and the Erdős-Rényi random graph. Starting from a ring lattice where nodes are distributed around a circle and edges exist between each node and the $\langle k \rangle$ nodes closest to it, each link is then re-wired with probability p to random nodes in the network. In this model, an increase in the clustering coefficient decreases the characteristic path length

(the average path length between any two nodes) and increases the 'small-world' effect. Clustering measures the concentration of triangles around the nodes and, as one might assume, social networks are found to be highly clustered owing to *triadic closure*, the process of friendships being formed by two people who have one or many friends in common, and *triad formation* where the introduced nodes link to pairs of nodes that are already linked to each other thus creating a triangle.

1.2e Static network models

Regular lattices, the Erdős-Rényi random graph, and the Watts-Strogatz model are all *static networks*. This class of network model accounts for a relatively small yet significant proportion of the complex network literature. A particularly useful static network construction comes from graph theory: Michael Molloy and Bruce Reed introduced random graphs with arbitrary degree sequences, known as the *configuration model* it is frequently cited in the complex network literature [34]. For instance, Mark Newman has argued that the size of the connected components in this model is equivalent to the problem of predicting the size of an epidemic outbreak on that network [35]. In this way static models are useful for studying processes like transport and epidemic spread *on* networks but they rarely answer the problem of why a network has such a topology to begin with.

1.2f Dynamic network models

As mentioned previously, the most prevalent method for growing networks is *preferential attachment*. Formally, the model proceeds as follows: in each time-step introduce a node to the network with m adjacent edges, attach the other end of each edge to nodes already in the network. Attach to a node i of degree k_i with probability

$$\Pi(i; k_i) = \frac{k_i}{M_1}. \tag{1.6}$$

The denominator here normalises the probability so that the sum over all nodes is equal to 1. This is how the model was formulated by Price and then later by Barabási and Albert, both motivate the mechanism by the rich-get-richer phenomenon, however, neither mention the fairly simple reason as to *why* well con-

nected nodes might be preferred to those with few connections: navigating a network naturally leads us to nodes with probability proportional to their degree. To better illustrate this important consideration it helps to refer to the observation of sociologist Scott Feld in 1991 in a paper titled "Why Your Friends Have More Friends than You Do" [36].

The title is clearly guilty of sensationalism, the actual content of the paper is mathematical by nature, proving the less insulting but equally counter-intuitive result: "most people have fewer friends than their friends have". This assertion is made without any assumptions being made about the topology of the social network. We compare the average degree of a node (in the social network that node is you) with the average degree of your friends (which are the neighbouring nodes). We can assume that saying 'their friend' is statistically equivalent to selecting *any* node in the social network that is reached by traversing an edge. The probability of randomly selecting a node of degree k is n_k/N whereas the probability of selecting a node of degree k by first randomly selecting an edge then selecting the node at one end of that edge is $kn_k/2E$. This is because the number of edges attached to nodes of degree k is kn_k, giving the same probability of selection given by Eq.(1.6) meaning that using this *local* neighbourhood method of selecting a node is equivalent to preferential selection. The mean degree of a randomly selected node obeys $\langle k \rangle = M_1/N$, the mean degree of preferentially selected nodes obeys

$$\langle k_{\text{pref}} \rangle = \frac{M_2}{M_1} = \langle k \rangle + \frac{\sigma^2}{\langle k \rangle} \tag{1.7}$$

which is always larger. Here we have introduced the variance of the distribution $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$.

The result suggests an effective method of choosing candidates to vaccinate in anticipation of an epidemic outbreak. Naturally it is more effective to target those people who are most social and have the potential to infect a large number of people. Instead of vaccinating a randomly selected sample of people, those people can be asked to nominate one friend. Feld's result dictates that on average that person will be better connected and therefore a more economic choice [37].

These sort of locally driven processes, which also include triadic closure and triad formation, have been the subject of many interesting works which we will review in more detail in Chapters 2 and 3. They also partly motivated the models I introduce in those chapters.

1.2g Solving for the degree distribution

The general framework we use for modelling the dynamics of a networked system goes as follows: We begin with a hypothesis stating how the individual nodes and edges of a network will change over time, this is what is referred to as the *model*. From this set of rules we derive topological characteristics of the network in the large t limit (t denotes time). We then compare the derived prediction with either real data or a programmed simulation of the model. Models of this type are a form of *stochastic process*. From an initial state, the state of the system one *time-step* later is determined by the set of rules defined in the model. Time, therefore, is always a variable in the model although in most cases it relates only very loosely to real time. In growing networks where a node is introduced in each iteration, time is equivalent to the number of nodes in the network, and we have $N \sim t$ for large t.

We can think of the *state* of the network to be described either by an *edge list*: a list of pairs of indexed nodes, or by an *adjacency matrix*: a square matrix where the value 1 in row i and column j represents an edge between the node indexed by i and the node indexed by j, the value 0 represents the absence of an edge. In each time-step the network transitions from one state to another according to probabilities that derive from the set of rules defined in the model.

Ideally we want to calculate the probability, according to the model's rules, that the system will be in a particular given state at time t. This can in principle be done by calculating the transition probabilities between every possible state and conducting a Markov chain analysis. However, only in very simple cases is this possible. It has proven more productive to define the state as the degree distribution. This method of *coarse graining* retains the information about one topological characteristic (the degree distribution) but disregards finer details of the network

such as correlations of degree between neighbouring nodes. Additionally, the problem of finding the probability that the network will have degree distribution $n_k(t)$ at time t is generally not considered, particularly in the case of growing networks where the fluctuations away from the mean tend to be small[10]. For any evolving network model, one of the most immediate and tractable problems is solving for $\langle n_k(t)\rangle$, the average number of nodes of degree k as a function of time t. The associated transitions can be written in the form of a family of equations

$$\langle n_k(t+1)\rangle = \langle n_k(t)\rangle + \text{expectation of the change in } n_k \tag{1.8}$$

for $k = 0,1,2...$. When t is large these equations are well approximated by

$$\begin{aligned}\frac{\partial n_k(t)}{\partial t} &= \text{expectation of the number of } \textit{births} \text{ of nodes of degree } k \\ &\quad - \text{expectation of the number of } \textit{deaths} \text{ of nodes of degree } k.\end{aligned} \tag{1.9}$$

The last simplification comes from assuming that as t grows very large the system will reach a *steady state* where the proportion P_k of nodes of degree k, $P_k = n_k(t)/N(t)$ has converged to a constant value. The births and deaths expected in Eqs.(1.9) often depend on n_1, n_2 etc. meaning Eqs.(1.9) can not be solved independently. The solution therefore comes from solving some recursive formula for P_1, P_2 etc. that derives from Eqs. Eqs.(1.9).

The techniques and the motivations presented in Section 1.2 form the basis for the original research described in this thesis. Chapters 2-5 detail 4 projects, in each one we introduce models which are novel to the scientific and mathematical literature. In Chapters 2-4 we use analytical methods to solve for the degree distributions as well as various other topological features that characterise the networks. In Chapter 5 we use related analytical methods to examine electrical resistance in a random network model, the problem leads us to explore novel techniques reducing a stochastic model to a solvable recursion relation.

[10]Pavel Kraprivsky and Sidney Redner analyse these effects on the preferential attachment model [38]

1.3 Advances presented in this thesis

Here is a brief synopsis of each chapter, highlighting the developments each model contributes to the body of complex network research.

Chapter 2: Clustered networks and citation distributions

Just as citation networks were pivotal in the early development of scale-free network research, they also provide a connection between the degree distribution of a network and its clustering. When the author of a paper cites another paper, s/he is also likely to cite articles that are included in the bibliography of that paper. In the language of network analysis this process is known as triad formation. Although this concept has appeared several times in previous work, for the first time we create a model whose parameters can be directly inferred from observations of the citing behaviour of scientific authors, namely how many triads their papers create. The primary concern of this chapter is analytically deriving the power-law degree distribution from a simple model of triad formation, we conclude the chapter by demonstrating how the analysis can be applied to a real data-set. The results suggests that the topology of the network emerges only from the tendency of authors to create clusters through references they choose.

Chapter 3: Rewiring processes for evolving complex networks

Networks of the type studied in Chapter 2 are partly static in the sense that once an edge has been placed between two nodes it remains in that position for the rest of the network's lifetime. Much less is currently known about the class of network whose edges are dynamic (i.e. at any point could potentially be removed or rewired) yet they have a far wider scope for applications, particularly in social networks and the structure of the World-Wide Web. The model we introduce in this chapter also uses local processes (including triadic closure) and growth, and abstractly resembles a number of online social networks and content discovery systems. Since rewiring process allow for the degree of a node to decrease as well as increase the resulting topology is significantly different to those who only grow and the degree distribution does not follow a power-law. We find that nodes in

this network are capable of becoming outliers of the degree distribution, gathering many more links than their compatriots. They achieve this by forming tightly woven 'rich-clubs' with other nodes that make them less susceptible to having their adjacent edges rewired away from them.

Chapter 4: Splitting nodes to create extreme power-law degree distributions

The power-law result of preferential attachment can intuitively be attributed to two driving factors: Increasing nodes of an already high degree stretches the tail towards large values of k. Creation of nodes of degree 1 (the most recently added nodes) increases the frequency of nodes of low degree. The model we introduce in this chapter accelerates the creation of nodes of degree 1 in the preferential attachment model through a mechanism where nodes of degree k are *split* into k nodes of degree 1. The effect is an increase in the power-law exponent γ from 3 to any value greater than 3 depending on the frequency of splitting events. The model demonstrates that it is possible to create scale-free networks with significantly more inequality in the degree of nodes than achievable through the original model. Variations of this model are also considered. In one particular case we create scale-free networks without growth that have exponents between 2 and 3, previous models of non-growing networks have not found exponents in this range.

Chapter 5: Electrical resistance of a network and random Fibonnacci sequences

The previous chapters have focused mainly on the emergence of network topologies. Here we turn our attention to the relationship between the structure of a network and a process occurring on it, specifically electrical resistance. The non-linear nature of electrical resistance makes it an interesting process to highlight how the topology can affect the dynamics. In solving this model we also develop the technique of approximating a stochastic process with a random Fibonacci sequence. The technique can potentially be adapted for future research in many different areas.

Chapter 2

Triad formation, clustering and citation networks

We introduce a network evolution process motivated by the network of citations in the scientific literature. In each iteration of the process a node is born and directed links are created from the new node to a set of target nodes already in the network. This set includes m "ambassador" nodes and l of each ambassador's descendants where m and l are random variables selected from any choice of distributions p_l and q_m. The process mimics the tendency of authors to cite varying numbers of papers included in the bibliographies of the other papers they cite. We show that the degree distributions of the networks generated after a large number of iterations are scale-free and derive an expression for the power-law exponent. In a particular case of the model where the number of ambassadors is always the constant m and the number of selected descendants from each ambassador is the constant l, the power-law exponent is $(2l + 1)/l$. For this example we derive expressions for the degree distribution and clustering coefficient in terms of l and m. We conclude that the proposed model can be tuned to have the same power law exponent and clustering coefficient of a broad range of the scale-free distributions that have been studied empirically.

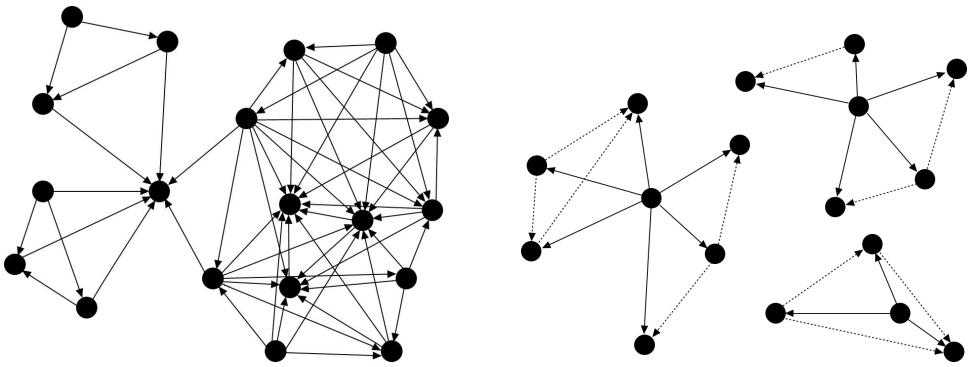
A crude but arguably effective measure of the worth of a scientific paper is the number citations made to it from other existing scientific articles. Empirical studies have shown that the number of articles with k citations (i.e. cited by k other articles) is proportional to k^{-3} [28, 13]. This distribution has certain properties we might expect, namely that the vast majority of papers written have few citations, creating little or no impact on future research, whereas a very small number of papers are extremely significant and have a very large number of citations. By modelling each paper as a node (vertex) and drawing directed edges from each paper to the papers it cites, it has been shown that the correct degree distribution is reproduced using *preferential attachment*; the process of creating nodes sequentially and linking them to nodes selected randomly with probability proportional to their degree (originally discussed in [10] although the term was coined later in [39]). The implication of this result is that authors of scientific articles are more likely to choose to cite articles that are already well cited rather than ones that have few or no citations. The attractiveness of highly connected nodes can be explained by a number of processes for example *redirection* [40], where nodes are selected randomly and a link is formed between one of its neighbours and a new node, and random walk models [41, 42, 43] where the new node is linked to the nodes occupied by random walkers on the network.

There is a growing literature offering more accurate representations of the way in which the citation network develops, much of this work can be found in the fields of Scientometrics, Bibliometrics, Informetrics and Webometrics [44, 45, 46]. A significant amount has been written concerning models that not only agree with the empirical data regarding degree distributions but also agree with other properties, for example in [46] the evolution of the citation network model is motivated by a coupling with the network of co-authors, other models account for the effect of time on the probability of receiving a citation [47, 48]. The model in [49] introduces tunable clustering (quantified by the clustering coefficient [50]) by extending

the preferential attachment model with an additional Triad Formation (TF) step. For each node that is introduced to the network, a node is selected by preferential attachment and linked to, then each neighbouring node is selected with probability p and also linked to from the new node, resulting in a triangle (triad) of edges. The forest fire model described in [51] extends the Triad Formation model by selecting multiple neighbours of the initially selected node, the process continues by then linking to a number of the neighbours of those neighbours and so on, at each stage a random variable from the binomial distribution determines the number of neighbours selected. In [52] the forest fire model, along with other models that attempt to mimic the network of citations in even greater detail, is tested against empirical data. The authors also examine the way the articles cited by any one paper, call it i, relate to one another forming a sub network called a reference graph of i (see Fig.(2.1a)). They observed that a clique structure is prevalent, i.e small groups of nodes that all link to each other, and incorporated this finding into their own model.

Much of the literature suggests that the high levels of clustering found in citation networks is a consequence of each author's choice to cite papers that are found in the bibliographies of the other papers they cite. This has been observed empirically [53], and modelled using a TF process where the initial nodes are selected randomly (rather than preferentially) [54]. A power-law degree distribution was found with an exponent that varies depending on the Triad Formation probability p, however, this model does not exhibit the exponential out-degree distribution observed in the data [55].

The models mentioned above and those considered in this paper belong to the class of evolving directed clustered scale-free networks that have applications beyond citation networks, the World-Wide Web being another well studied example. In these models the distributions of in-degree and out-degree are treated separately, often driven by a preferential linking mechanism where the probability of adding an edge from a node i is proportional to the out-degree of i, similarly the probability that the link will end at node j is proportional to the in-degree of



(a) The reference graph of a real paper.

(b) A typical set of nodes selected in one iteration of the model.

Figure 2.1: 2.1a shows the sub-network known as the reference graph of a real paper (taken from [52]) the nodes represent the papers cited by that paper and the edges represent the citations between them, it is highly clustered and many of the nodes are descendants of others. 2.1b shows the a typical sub-network structure of nodes that the proposed model links to, in this case three nodes are selected initially and 3, 5, and 6 of their descendants are also selected, the dotted lines represent the other links between descendants.

j [56]. Correlations between the in-degree and out-degree of nodes in such networks have been shown to emerge [57]. A detail of citation networks that makes analysis substantially easier is that the out-degree of a node i is fixed from the moment it is created. Consequently the evolution of the out-degree distribution can be disregarded, moreover we can control the out-degree distribution through an appropriate parametrization and ultimately answer the question of how the distribution of bibliography sizes affects the topology of the network.

In this chapter we introduce a variant of the TF model that uses a different parameter set to those previously studied. Using the distributions for the number of initial citations and the number of copied citations (which together give the out-degree distribution) as parameters, we show that the networks created by this process may have power-law in-degree distribution with any exponent greater than 2, and a clustering coefficient that ranges between 0 and 1. In Section 2.1 we describe the stochastic process that iteratively grows the network. We describe a simple case of the model in Section 2.2 and solve for the power-law exponent of the in-degree distribution in terms of two input parameters. In Section 2.3 we

formulate an expression for the in-degree distribution in terms of two input probability distributions then in Section 2.4 we find an expression for the clustering coefficient for the model in Section 2.2. In Section 2.5 we present numerical results that confirm the results of Sections 2.2 and 2.3. In Section 2.6 we discuss the strengths of our models and suggest how this work might be continued.

2.1 A growing network model with tunable clustering

Starting from a finite random network, at each iteration a node j is introduced and directed links are formed between j and a set of nodes that already exist in the network. Letting p_l and q_m be the probability distributions of the discrete random variables l and m, links are formed by the following process:

1. The value m is selected with probability q_m and steps 2 and 3 are repeated m times.
2. The value l is selected with probability p_l, a node i in the network is randomly selected from those which have out-degree l or greater, the edge $j \to i$ is added. Borrowing the terminology used in [51] we will refer to i as an "ambassador".
3. l of i's descendants are randomly selected and directed edges are added from j to each of these.

We are primarily interested in expressing the degree distributions for both incoming and outgoing edges and the clustering coefficient of the network as the number of iterations grows very large in terms of p_l and q_m ($l, m \in \mathbb{N}$). In the next section we solve for a simplified model where l and m are fixed (i.e. $p_r = \delta_{rl}$ and $q_r = \delta_{rm}$), we present the general solution in the section that follows.

2.2 A simple parametrisation of the model

We examine the network generated by the process described in Section 2.1 when $p_r = \delta_{rl}$ and $q_r = \delta_{rm}$, in this section we derive the degree distribution of this network. In this simplified model the growth of the network depends only on the fixed values l and m, thus the process can be described concisely as follows; in

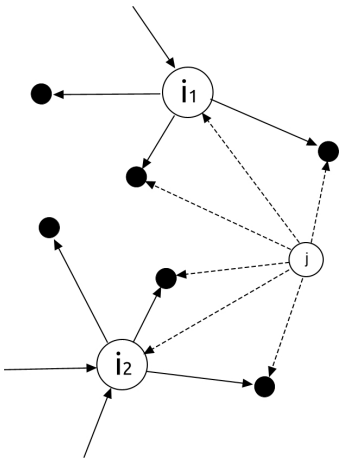


Figure 2.2: The new node j attaches to 2 randomly selected nodes i_1 and i_2 as well as 2 randomly selected descendants shown here in black, the dashed lines represent the new edges that are added in this iteration whereas the solid ones were added previously. This example illustrates one iteration in the growth process when $m = 2$ and $l = 2$.

each iteration, m ambassador nodes are randomly selected, l descendants of each ambassador are also selected, then a new node j is attached to each of the selected nodes (see Fig.(2.2)). This special case of the model is equivalent to the model proposed in [58] modified to have directed edges. We are interested in calculating the probability $P(k)$ of finding a node with in-degree k, in the citation model this represents the proportion of articles that are cited by k other papers. Let N be the total number of nodes, N increases by 1 with each iteration and every node has an out-degree of $m(l+1)$, the number of edges as N grows large is $E = m(l+1)N$. Consider a typical node i with in-degree k. There are two possible events which may cause the degree of i to increase to $k+1$: i can either be selected as one of the m ambassador nodes, or it can be selected as a descendant of another node j. In any given iteration, i will be selected as an ambassador with probability m/N. Alternatively j will be selected as an ambassador with probability m/N and then i will be one of the l selected descendants of j with probability $l/m(l+1)$. Let $P(i;k)$ denote the probability that in one iteration a node i with degree k is selected. Since there are k potential ancestors to i,

$$P(i;k) = \frac{m}{N}\left(1 + \frac{lk}{m(l+1))}\right). \tag{2.1}$$

It is possible for the same node to be selected two or more times in one iteration, for example if two of the selected ambassador nodes are a distance of one or two edges from each other. Since this possibility becomes less likely as N increases we do not account for it in our calculations. Let I_k be the number of nodes with in-degree k. For $k \geq 1$, I_k changes over time according to the rate equation

$$\frac{\partial I_k}{\partial N} = \frac{m}{N}\left[\left(1 + \frac{l(k-1)}{m(l+1)}\right) I_{k-1} - \left(1 + \frac{lk}{m(l+1)}\right) I_k\right]. \tag{2.2}$$

The first term on the right hand side accounts for the creation of a node of in-degree k that occurs when one of the new edges attaches to a node of in-degree $k-1$, the second term accounts for the destruction of a node of in-degree k when it is attached to by one of the new edges. For $k = 0$ the rate equation is

$$\frac{\partial I_0}{\partial N} = 1 - \frac{m}{N} I_0. \tag{2.3}$$

We are interested in finding $P_{\text{in}}(k)$ the probability of a node having in-degree k when N is very large. By assuming $P(k)$ grows linearly with N when N is large, we substitute $I_k = NP_{\text{in}}(k)$ into Eq.(2.2) to find

$$\left(\frac{l+1+m(l+1)}{l} + k\right) P_{\text{in}}(k) \quad = \quad \left(\frac{m(l+1)-l}{l} + k\right) P_{\text{in}}(k-1) \tag{2.4}$$

for $l \neq 0$. From Eq.(2.3) we also find

$$P_{\text{in}}(0) = \frac{1}{1+m} \tag{2.5}$$

and thus the in-degree distribution is expressed

$$P_{\text{in}}(k) = \frac{1}{m+1} \frac{\Gamma[(2l+1+m(l+1))/l]\Gamma[k+m(l+1)/l]}{\Gamma[m(l+1)/l]\Gamma[k+(2l+1+m(l+1))/l]}. \tag{2.6}$$

For large enough values of k, $P_{\text{in}}(k)$ has power-law form

$$P_{\text{in}}(k) \sim k^{-\gamma} \text{ where } \gamma = \frac{2l+1}{l}. \tag{2.7}$$

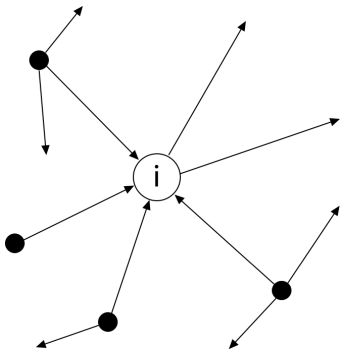


Figure 2.3: At each time-step i can be selected as the initial node, alternatively one of the black nodes may be selected and then i is selected as one of its descendants.

2.3 Degree distribution in the general case

Let $P_{\text{out}}(s)$ denote the proportion of nodes in the network that have out-degree s. Note that at the time of its creation, the out-degree of a node is fixed and, unlike its in-degree, does not change over time. Therefore, for sufficiently large networks, $P_{\text{out}}(s)$ is equal to the probability of creating a node with out-degree s within a single iteration. This can be written

$$P_{\text{out}}(s) = \sum_{n=1}^{\infty} q_n P\left(\sum_{i=1}^{n} x_i = s\right) \tag{2.8}$$

where the x_i are integer random variables that equal $l+1$ with probability p_l.
To calculate the in-degree distribution we again construct a rate equation from the probability that the degree of a typical node i will increase in one iteration. Let T_l be the number of nodes which have out-degree greater or equal to $l-1$, for the node i to be randomly selected as the ambassador in step 2 it must be one of these nodes. The probability that this is the case, multiplied by the probability that i is the one node randomly selected from the T_l nodes available, forms the probability that i is the ambassador given that l is the number of descendants chosen in step 2. Summing over all values of l returns $P_a(i)$ the probability that any node i is

selected as an ambassador, thus

$$\begin{aligned} P_a(i) &= \sum_{l=1}^{\infty} p_l \frac{T_l}{N} \times \frac{1}{T_l} \\ &= \frac{1}{N} \end{aligned} \tag{2.9}$$

Suppose i has in-degree k, and that as per step 2 only nodes with out-degree l or greater can be selected as an ambassador. Suppose also that the ambassador is an ancestor of i and has out-degree s where $s \geq l$ (see Fig.(2.3)). The expectation of the number of nodes that satisfy these conditions is $kP_{\text{out}}(s)$. The probability that each one is selected is $1/N$ given by Eq.(2.9). Once selected, the probability that of the s descendants i is one of those selected in Step 3, is l/s. Taking the product and summing over all l and all possible values of s returns $P_d(i;k)$ the probability that any node i with degree k is selected as a descendant, therefore

$$P_d(i;k) = \frac{k}{N}\Phi \tag{2.10}$$

where

$$\Phi(p,q) = \sum_{l=1}^{\infty}\sum_{s=l}^{\infty} \frac{l p_l P_{\text{out}}(s)}{s}. \tag{2.11}$$

The probability of a node i with degree k being linked to during step 2 or 3 of the process is $P_a(i) + P_d(i;k)$. Summing again over all possible values of m, the probability that the degree of i will increase by 1 during any iteration is

$$P(i;k) = \frac{\langle m \rangle}{N}(1 + k\Phi) \tag{2.12}$$

where

$$\langle m \rangle = \sum_{m=1}^{\infty} m q_m. \tag{2.13}$$

The associated rate equation is constructed in exactly the same way as Eq.(2.2), thus

$$\frac{\partial I_k}{\partial N} = \frac{\langle m \rangle}{N}\left[(1 + (k-1)\Phi)\, I_{k-1} - (1 + k\Phi)\, I_k\right]. \tag{2.14}$$

Letting $P_{\text{in}}(k) = I_k/N$ be the proportion of nodes that have in-degree at large N, Eq.(2.14) becomes

$$\left(\frac{1+\langle m\rangle}{\langle m\rangle\Phi} + k\right) P_{\text{in}}(k) = \left(\frac{1-\Phi}{\Phi} + k\right) P_{\text{in}}(k-1). \tag{2.15}$$

The rate equation for I_0 solves to find $P(0) = 1/(1+\langle m\rangle)$ and thus

$$P_{\text{in}}(k) = \frac{1}{\langle m\rangle + 1} \frac{\Gamma[(1+\langle m\rangle)/\langle m\rangle\Phi]\Gamma[k+(1-\Phi)/\Phi]}{\Gamma[(1-\Phi)/\Phi]\Gamma[k+(1+\langle m\rangle)/\langle m\rangle\Phi]}. \tag{2.16}$$

For large values of k, $P_{\text{in}}(k)$ has a power-law form

$$P_{\text{in}}(k) \sim k^{-\gamma} \text{ where } \gamma = 1 + \frac{1}{\langle m\rangle\Phi}. \tag{2.17}$$

2.4 Clustering

The clustering coefficient of a node i is defined as the number of edges between the neighbours of i divided by the number of pairs of nodes from the neighbours of i. If node i has d neighbours (ancestors and descendants) then this is

$$C_i = \frac{2E_i}{d(d-1)} \tag{2.18}$$

where E_i is the number of edges between the neighbours of i. Let $E(k)$ be the expectation of E_i when i has in-degree k, also let $\Theta(k)$ be the expectation of the number of times i has been selected as the ambassador node during step 2 of any previous iteration. From equations (2.9) and (2.10) we see that the kth edge is $(k-1)\Phi$ times more likely to be added as a result of one of i's neighbours being an ambassador rather than i being selected as an ambassador itself, so

$$\Theta(k) = \sum_{i=0}^{k-1} \frac{1}{1+i\Phi}. \tag{2.19}$$

We find $\langle C\rangle$, the mean of C_i over all nodes i in the network we studied in Section 2.2 where $p_r = \delta_{rl}$ and $q_r = \delta_{rm}$. This is the sum over all k of the product of $P_{\text{in}}(k)$ given by Eq.(2.6), and $C(k)$ the expectation of the clustering of a node of degree k.

The contribution to E_i made by each neighbour j of i depends on the way in which the link was originally created, there are four cases to be considered. The first is where the link $i \to j$ was added when i was introduced to the network, in this case the expected contribution to E_i is the number of edges in the reference graph of i (see Fig.(2.1)). The second case is where i was selected as an ambassador and l edges are added to E_i. In the third case D_a edges are counted for those that were added when i was selected as a descendant of the ambassador j where the link from j to i was originally formed when i was selected as an ambassador in a previous iteration. Lastly, D_d edges are counted for those that were added when i was selected as a descendant of the ambassador j where the link from j to i was formed when i was originally selected as a descendant. Then

$$E(k) = E(0) + \Theta l + (k - \Theta)\left[\frac{\Theta}{k} D_a + \frac{1-\Theta}{k} D_d\right]. \tag{2.20}$$

When a node i is added to the network, E_i includes the edges between each ambassador node and its descendants as well as the edges between those descendants. The probability that an edge exists between two descendants of the same node is $C(0)$ so the expectation of E_i is

$$E(0) = m\left(l + \binom{l}{2} C(0)\right). \tag{2.21}$$

Combining this with Eq.(2.18) and solving gives

$$C(0) = \frac{2l}{(m-1)(l+1)^2 + 2l}. \tag{2.22}$$

In the instance where an ambassador node j is selected and i is linked to as one of j's descendants, the new node will link to a further $l-1$ neighbours of $m(l+1)-1$ possible descendants of j, those that are also neighbours of i will be counted in E_i. If j originally formed a link with i by selecting i as an ambassador, then l of i's descendants are also descendants of j, hence the expectation of the number of neighbours of i that are linked to is

$$D_a = \frac{l(l-1)}{m(l+1) - 1}. \tag{2.23}$$

If j originally formed a link with i by selecting i as the descendant of some other node, the expected number of of links between j and any of i's neighbours is $1 + (l-1)C(0)$ so the expectation of the number of i's neighbours linked to is

$$D_d = \frac{[1 + (l-1)C(0)](l-1)}{m(l+1) - 1}. \tag{2.24}$$

Combining Eqs. (2.18), (2.19), (2.20), (2.22), (2.23) and (2.24) gives an expression for the clustering of a node of in-degree k in terms of m and l ($l, m \geq 1$), multiplying by $P_{\text{in}}(k)$ given by Eq.(2.6) and summing over all k gives the mean clustering for the entire network. The clustering coefficient tends to 0 as m grows large. As l grows large the clustering also tends to 0 except when m is equal to one, in which case it tends to 1 (see Fig.(2.4)).

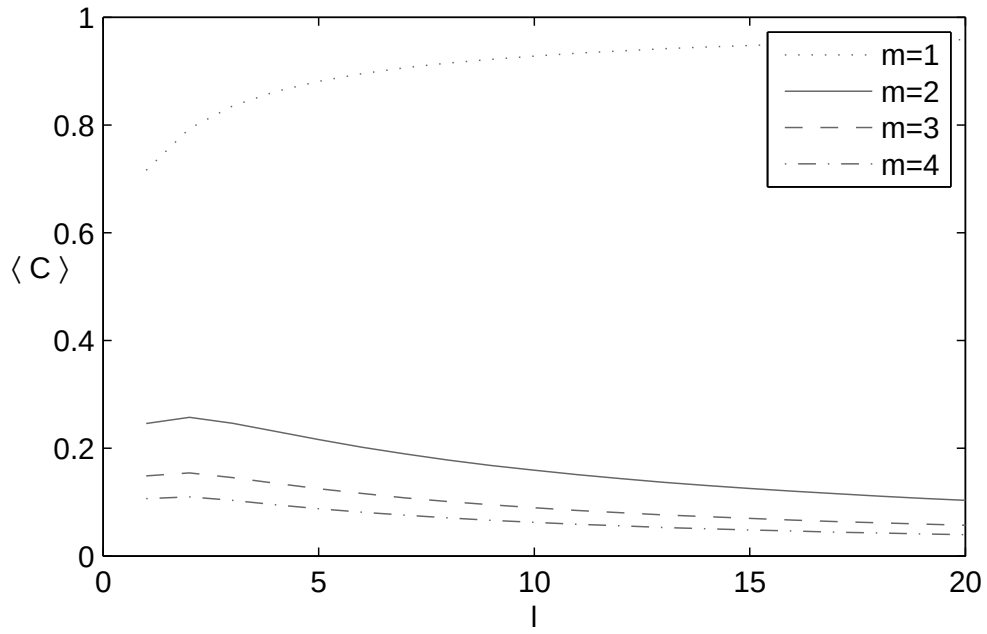


Figure 2.4: Using the formulae in Section 2.4 the average clustering coefficient is plotted on the vertical axis for the first 4 values of m, against l on the horizontal axis. The clustering tends to zero as l grows large for all values of m with the exception of $m = 1$.

2.5 Numerical results

It should be emphasized that the results found in previous sections are mean field approximations as N tends to infinity, it therefore cannot immediately be assumed that the derived results will be a fair description of any individual network grown following the proposed process. We consider the following:

1. Correlations between out-degree of a node and the in-degree of its descendants. Specifically in Eq.(2.10) where it is assumed that the out-degree of the neighbours of node i (i.e the black nodes in Fig.(2.3)) follow the distribution $P_{\text{out}}(s)$ regardless of the in-degree of i. In reality this might not be the case; imagine, for example, a node j with relatively large out-degree and i as one of its descendants, selecting j as an ambassador in future iterations is relatively unlikely to result in selecting i again unless the new node also has large out-degree (more specifically a large value of l in step 2 of the iteration), so the expectation is for i to have few ancestors each with large out-degree (the opposite is true if the out-degree of j is small). The effect of this has not been considered analytically, instead we show numerically that in practice there is no significant deviation from the mean field result.

2. Finite size corrections. For finite networks of size N, the existence of a largest degree $k_{\max}$ means the power-law degree distribution must fail around the largest values of k. These effects have been investigated for particular classes of preferential attachment based network [43, 38]. Once the asymptotic mean field solution $P(k)$ is known, the solution to average degree distribution on a network of size N is

$$N_k(N) \simeq NP(k)F(\xi) \text{ where } \xi = k/k_{\max}. \tag{2.25}$$

   From the generated data (discussed below) we observe in Fig.(2.5) a similar form of scaling function $F(\xi)$ as observed in [43, 38]. The function $F(\xi)$ can be derived by considering the average of all possible values of $N_k(N)$ for every N starting from an initial value for $N_1(1)$, under the specific circumstances however, the initial conditions must be chosen carefully for each possible

choice of our parameters. It is impractical to derive $F(\xi)$ for every possibility here, instead we show that the model passes a suitable goodness-of-fit test even when finite size effects are neglected.

In the numerical tests we grew a network in three phases, initially a small number of nodes with large out-degree are created (the degree must be large enough to allow T_l to be non-zero for all l), then a phase of creating new nodes with a random number of out links to randomly selected nodes already in the network, finally the process described in Section 2.1 is applied for a large number of iterations. To assess the goodness-of-fit of the results in Eqs. (2.6) and (2.16) we compare the degree distribution of a simulated network of size N to the distribution given by drawing N values from a uniform pseudo-random number generator adapted to output the value k with probability given by Eq.(2.16)[1]. The degree distribution of the simulated network is then compared against the mean field prediction Eq.(2.16) using a suitable measure of similarity, in this case we choose the Kolmogorov-Smirnov statistic. Lastly, over a large number of trials (we chose 10^3) the pseudo-random distribution is measured against the model, the p-value for this test is the proportion of trials in which the simulated data is closer to the model (i.e. a lower KS statistic) than the random data. Here we have followed the methodology of [60], developed for use in empirical studies where the data are not likely to be as clean as those generated in a computer simulation. The authors suggest that a p-value greater than 0.1 is evidence enough for the model to be accepted. We ran this test for 10^3 networks generated first using the pair of distributions $p_r = (1/3)(\delta_{r1} + \delta_{r2} + \delta_{r3})$ and $q_r = (1/3)(\delta_{r2} + \delta_{r3} + \delta_{r4})$ then another 10^3 networks using $p_r = \delta_{r3}$ and $q_r = \delta_{r4}$. Fig.(2.6) shows the proportion of these trials that achieved particular p-values, while the p-value varies greatly, only a very small proportion are less than 0.1.

We ran the simulation for a large number of different distributions p_l and q_m and found that the numerical results agreed with the analytically derived formulae, Figures (2.7) and (2.8) show two typical examples. The log-binned values are the

[1]The method used is described in [59]. The uniform pseudo-random number generator we use is MatLab's `rand()` function.

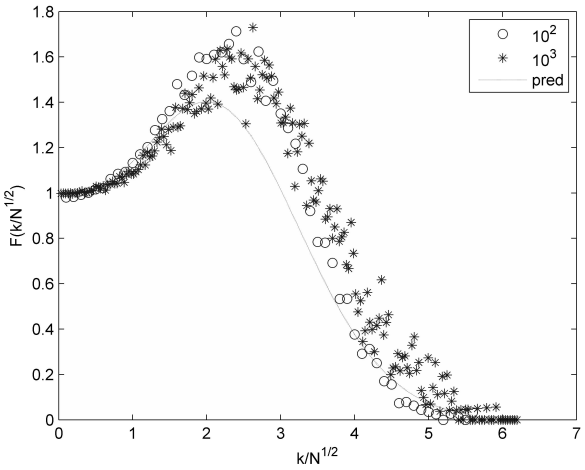


Figure 2.5: The correction function $F(\xi)$ in Eq.(2.25) for the simplest case of the model. Here $m = 1$, $l = 1$ and an initial condition of three nodes, one with out-degree 2 connected to the other two. The dashed line represents the analytical prediction from [38].

means of I_k over a ranges of k that increases logarithmically with k. In these examples the first bin is just the first value of I_1, the second is the mean of I_2 and I_3, the third is the next 4 values and so on. We were able to compute the clustering coefficient only for networks with no more than approximately 10^3 nodes, we found that for networks where the out-degree of the nodes is large the simulated result tended to be higher than the analytical result, this exposes the assumption in the analytical calculations that ambassador nodes will not be close to each other in the network. This discrepancy gets smaller as the network grows larger as one would expect.

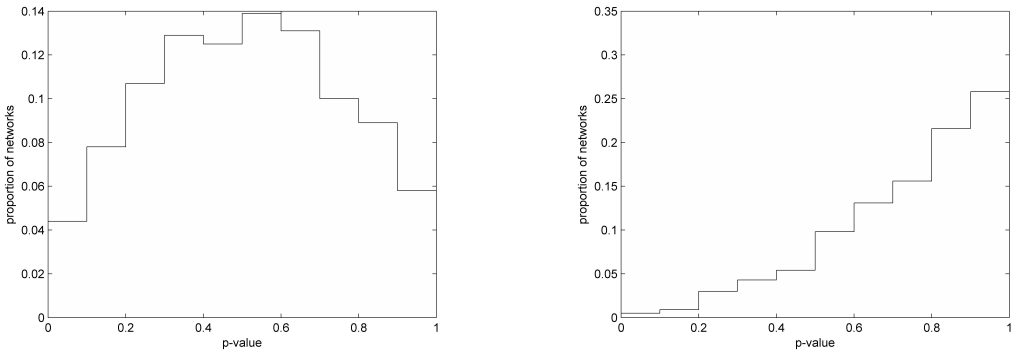


Figure 2.6: Histograms showing the numerically derived frequencies that given p-values were achieved. The networks in the figure to the left use the parameters $p_r = (1/3)(\delta_{r1} + \delta_{r2} + \delta_{r3})$ and $q_r = (1/3)(\delta_{r2} + \delta_{r3} + \delta_{r4})$ and the networks in the figure to the right use the parameters $p_r = \delta_{r1}$ and $q_r = \delta_{r1}$, each network has $N = 10^3$.

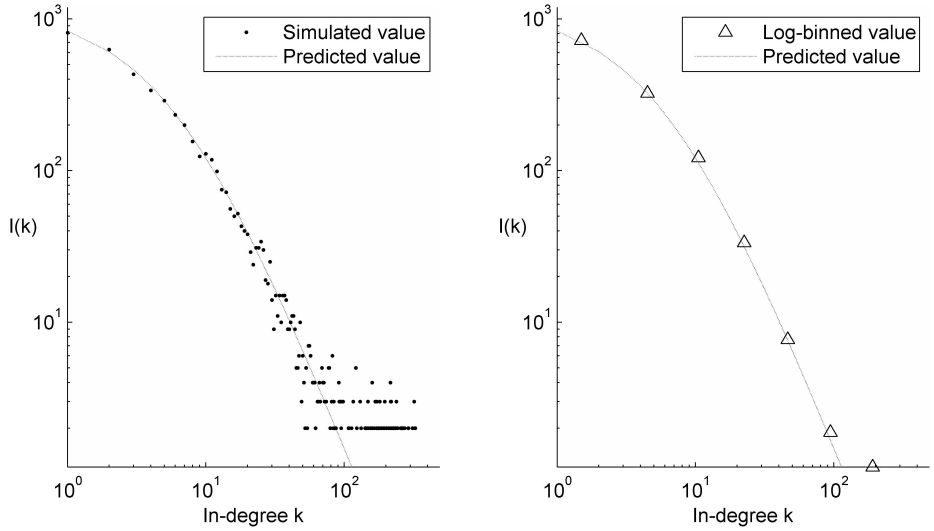


Figure 2.7: The number of nodes $I(k)$ of in-degree k for each value of k, results here are taken from the simulation of the model when $p_r = \delta_{r3}$ and $q_r = \delta_{r4}$, the dotted line shows the predicted result derived from Eq.(2.6), the right hand figure shows that the log-binned values agree very well with the prediction with the exception of the largest values of k. This network contained 6×10^3 nodes, the first 300 were added randomly.

2.6 Conclusion

There are two particular strengths of this model that are worth highlighting. The first is tunability; the feature that a wide range of results for the clustering and power-law exponent can be achieved by inputting the appropriate parameter values. In the simplified model l can be tuned to achieve exponents between 2 and 3, by adjusting m the clustering is tunable to a restricted range of values (see Fig.(2.4)). It is not difficult to find distributions in the full model that allow the clustering to be tuned to any value between 0 and 1, however as we showed in Section 2.3 the exponent in the distribution depends on both m and l. Tunable networks are particularly useful to study processes on networks such as epidemic spread; since the results they obtain depend largely on the topologies of the underlying networks, flexibility allows the extent of the effects of clustering and degree distribution to be analysed in greater detail [61]. The second strength of this model is its generality; the property that there are a wide range of parameter values that can be used as input to the model. As there are no restrictions on the probability distributions involved it is possible to choose those that most closely match the empirical data. We conclude this chapter with suggestions about how the model can be compared to a real citation network. Results are shown in Fig.(2.9).

2.6a Using the model for data analysis

Suppose we have the edge-list or adjacency matrix of a citation network, for each node we are able to extract the sub-network of its descendants that we call its reference graph. In the model proposed in Section 2.3 we presented a method of selecting a set of target nodes such that the set of possible reference graphs can be tuned to somewhat mimic reality. Our aim then is to infer from the data suitable distributions p_l and q_m so that the artificial reference graphs produced resemble the actual set of reference graphs observed. Since the reference graphs themselves bear a certain amount of disorder, we could potentially put an arbitrary amount of effort into finding the optimal way to make this inference. Instead, we simply choose the easiest method. This involves counting, for each node i, the number of its descendents that are not descended from any other node in the reference

graph, this value will be a data point in the distribution of ambassadors. For each of those ambassadors we count the number of its descendants that also belong to the reference graph of i, these values make up the distribution of descendants. We can then normalise these distributions to use as our values of q_m and p_l respectively. For example, applying this to the real reference graph in Fig.(2.1a) we would add to the frequency of $m = 4$ in the distribution of ambassadors, and add to the values $l = 2, 3, 5$ and 6 in the distribution of descendants. The imperfections of this method are visible in this example as one can easily see that some of the nodes have been neglected while others have been counted twice.

We applied this method to the freely available High Energy Physics data-set downloaded from the Stanford University website [62]. Once the distributions had been found we used them as parameters in our model and compared the results (Fig.(2.9)). Using the method described in [60] for determining the power-law exponent we found that the degree distribution of the real network begins to follow a power-law at $k \geq 100$, at which point the exponent is calculated to be $\gamma = 2.87$. From simulating the network using the inferred parameters and an equal number of nodes we found exponents in the range $1.8 < \gamma < 2.8$ depending on the choice of initial conditions. Additionally the mean value for the number of descendants was found to be $\langle l \rangle = 4.8$, from the result found in Section 2.2 this corresponds to $\gamma = 2.2083$. The difference between the model's prediction and the data suggest that the rules of the model do not entirely describe the citing behaviour of scientific authors in high energy physics. This is not surprising, some degree of preferential selection and intrinsic quality would be expected, our model does not take these factors into account.

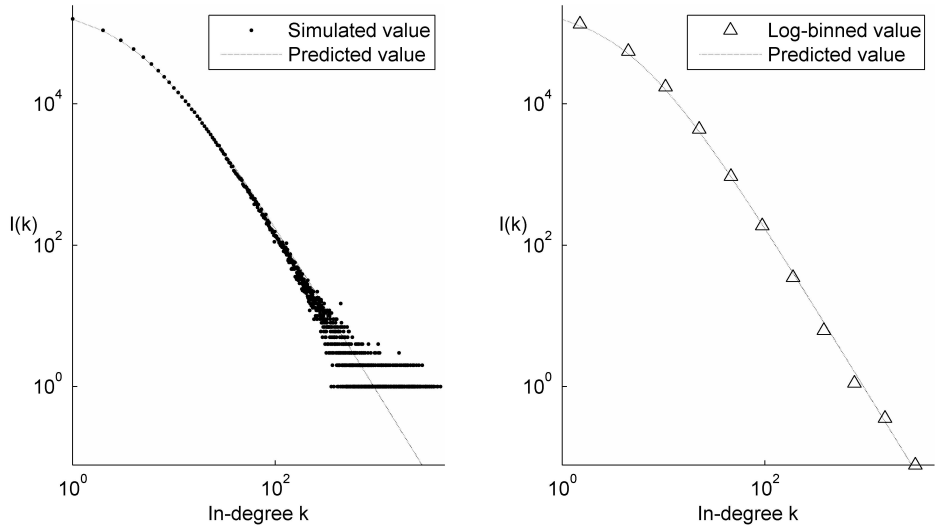


Figure 2.8: The in-degree from the simulation of the model when $p_r = (1/3)(\delta_{r1} + \delta_{r2} + \delta_{r3})$ and $q_r = (1/3)(\delta_{r2} + \delta_{r3} + \delta_{r4})$, the dotted line shows the predicted result given by Eq.(2.16). This network contained 10^6 nodes, the first 10^3 were added randomly.

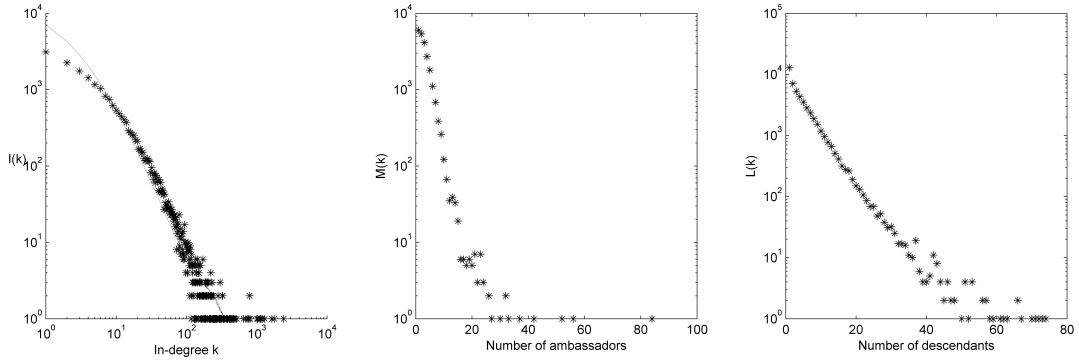


Figure 2.9: The panel on the left shows the citation distribution of the *high energy physics* dataset. The dashed line shows the results of the model when we use the distributions shown in the other two panels as parameters. These distributions are inferred from the data, $M(k)$ is equivalent to q_m and is the number of papers that link to k ambassador nodes. $L(k)$ is equivalent to p_l and is the number of occurrences of k papers being linked to from an ambassador.

Chapter 3

Edge rewiring and local rules

The effects of link rewiring are considered for the class of directed networks where each node has the same fixed out-degree. We model a network generated by three mechanisms that are present in various networked systems; growth, global rewiring and local rewiring. During a rewiring phase a node is randomly selected, one of its out-going edges is detached from its destination then re-attached to the network in one of two possible ways; either globally to a randomly selected node, or locally to a descendant of a descendant of the originally selected node. Although the probability of attachment to a node increases with its connectivity, the probability of detachment also increases, the result is an exponential degree distribution with a small number of outlying nodes that have extremely large degree. We explain these outliers by identifying the circumstances for which a set of nodes can grow to a very high degree.

The question of how complex patterns can be produced by the collective behaviour of many interacting agents such as particles, cells or people, is one of the most important considerations in complex systems science. The techniques of statistical physics that originated from the study of gasses and magnets have been adapted to address this question to explain a much wider range of emergent phenomena seen in biological and social systems. Fundamentally, mathematical models are used to derive statistical information about the system as a whole from the assumptions made about its constituent agents, or more specifically, the "rules" that govern their interactions. While in most physical systems agents interact with their closest neighbours in a spatial sense, many other systems are not constrained in this way, these are typically modelled as networks where the concept of distance between two points is redefined as the path-length between two nodes. An example of a local rule is triadic closure, the creation of a link between two nodes separated by a path-length of 2.

When the growth and evolution of a network is driven by local rules, nodes tend to be selected with a frequency proportional to how well connected they are. This is simply because a node with x connections is present in the neighbourhood of x other nodes, in other words there are x possible ways to discover the node via a local search. It is not suprising then, that the scale-free networks generated by global preferential attachment can also be created by numerous processes that use only local rules i.e. with no global knowledge of the network structure [41].

Typically in these models, a network will begin as a small set of nodes connected by edges, then with each iteration, more nodes are introduced and connections made, thus increasing the degree of those that are already there. Networks of this type are partly static in the sense that once an edge has been placed between two nodes it remains in that position for the rest of the network's lifetime. The class of

network whose edges are dynamic, i.e. at any point could potentially be removed or rewired, has far wider scope of application.

This chapter examines networks that combine dynamic edges with locally driven processes. Our model is a iterative process that evolves a network, the parameters are the rate of growth, and the rates of local and global (random) rewiring. We examine only networks with directed edges and nodes of a fixed out-going degree. For particular regions of the parameter space, we examine in detail a phenomenon whereby a small set of nodes, owing to their position in the network, gather significantly larger number of connections than those outside the set. These considerations lead to a good approximation of the extreme tail of the degree distribution, giving probabilities for the existence of outlying nodes of the distribution, sometimes referred to as dragon kings [63].

In Section 3.2 we introduce a model of growth and rewiring in directed networks and show the main results. The following sections describe the mathematical models and their solutions. In Section 3.3 we find the distribution of cycles of size n in the initial randomly wired graph. In Section 3.4 we find a formula for the degree distribution in the large t limit. In Section 3.5 we model the total degree of the dominant nodes and for selected parameter values derive the degree distribution tail.

3.1 Related work

Local rules for growing networks have been in the literature for some time [41, 40]. In the model most similar to the one presented here [64], the preferential attachment mechanism is generalised to include rewiring events. They find both exponential and power-law degree distributions depending on the choice of parameters. Preferential attachment in rewiring has been studied on a network of fixed size with the interesting conclusion that a power-law degree distribution can be achieved without a growing network [65]. This result relies on the use of a non-linear attachment kernel (heavily biased towards nodes with large degree) to

ensure that nodes with large degree continue to grow in spite of the preferential detachment that also occurs through rewiring. This work has been extended to bipartite networks [66] which have an advantage of being free of degree correlations between neighbouring nodes, thus the results in [43] for the mean field solution to the degree distribution are exact. The same model also exhibits a condensation phenomenon, also know as gelation [40], where one node becomes connected to almost every other, this is relevant to the study of the dominant nodes presented here.

A large body of literature, much of which is commercially motivated, comes from the analysis of the network properties of web 2.0 systems [67, 68]. We believe our results here are relevant in this field since rewiring, local dynamics and directed links are present in many of these self-organising systems. Twitter, for example, gives its users the option to "unfollow" other users meaning the edges are not static as they are in the majority of complex network models. Local rules, specifically triadic closure contribute to the growth of the network [69], however the distribution does not follow a power-law [70].

Recommendation algorithms designed to facilitate sharing online news articles, music, films etc. connect users together based on the similarity of the content they have responded to positively. The content a user is exposed to in this way is limited to a small number of items shared by her neighbours. When the algorithm updates the links based on the most recent data, we can expect the strength of the similarity between her and her second neighbours to increase, making triadic closure likely. The network topologies of these networks has been studied in [71]. In this work the network is treated as a static object at one instant in time, clustering is found to be significantly higher than the random network which suggests that it is likely that triadic closure plays a part in the dynamics of the network. The evolution of a theoretical model network [72] considers directed edges between "leaders" and "followers" that are rewired periodically according to a similarity score. A scale-free structure is found but the authors do not go into detail about the rewiring dynamics. The network evolution of recommendation networks perhaps

deserves more attention since it exhibits cumulative advantage effects that have consequences for many commercial areas.

3.2 A growing network model with local and global rewiring

Let $G(N, mN)$ be a random graph in which each of the N nodes has m out-going directed edges, the destination of each of the directed edges is selected randomly. Throughout this chapter we use 'degree' to refer to the in-coming degree of a node. In each time-step the network develops in one of the following ways

- **Local rewiring:** With probability p, randomly select a node and rewire one of its out-going edges to a randomly selected descendant of one of its descendants (see Fig.(3.1)).
- **Global rewiring:** With probability q, randomly select a node and rewire one of its out-going edges to a randomly selected node.
- **Growth:** With probability r, introduce a node to the network with m out-going edges, attach the edges to randomly selected nodes in the network.

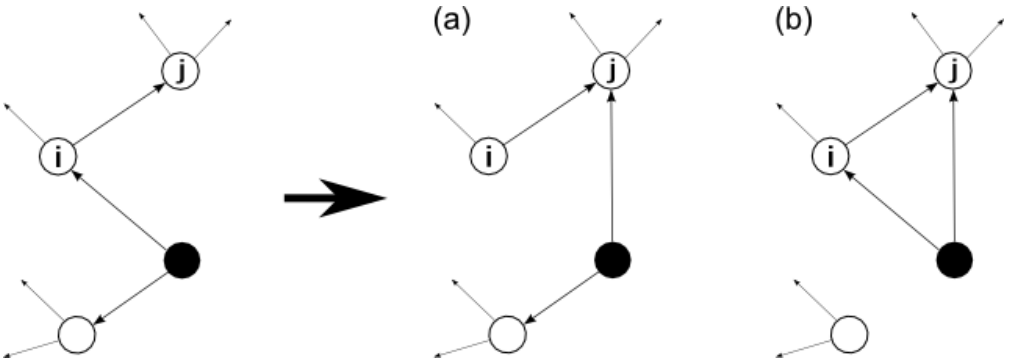


Figure 3.1: The two possible ways to locally rewire. The left image shows part of the network before rewiring. We consider two possible interpretations of our model. In both, the black node is initially selected, the target node can be any second descendant of the black node, in this example it is j. One of the out-going edges from the black node is then rewired to the target node, it can either be the node that connects the black node to j, shown in (a), or it can one which completes the triad, shown in (b).

For convenience we set $r = 1-p-q$. As we iterate this process, the binomial degree distribution of the initial network converges towards an exponential distribution for every choice of p, q and m (Fig.(3.2)). When q is small and p is relatively

large we observe additional dynamics where we see a small number of outlying nodes with degrees much higher than predicted by the exponential distribution (Figures (3.2b) and (3.3)). These are the conditions for "rich-clubs" to develop, small sets of nodes whose growth in degree is magnified by the fact that the set has very few out-going links. The outlying nodes, which we call 'dominant nodes', exist because their out-going edges belong to small cycles. This is illustrated most easily in the case where $m = 1$; over time the outliers increase in degree until the cycle they belong to is broken, at this point the degree rapidly falls while a new dominant node begins its rise (Fig.(3.4a)). For sufficiently small q, the node remains dominant long enough to reach a state where its degree, on average, is neither increasing or decreasing, this causes a small spike in the tail of the degree distribution (Fig.(3.4b)).

3.3 Random graphs with directed edges and fixed out-degree

For the mN edges in the network, each is attached to the node i with probability $1/N$. The probability that i has degree k is the probability of k successes in mN trials. Letting P_k denote the probability that any node has degree k we have

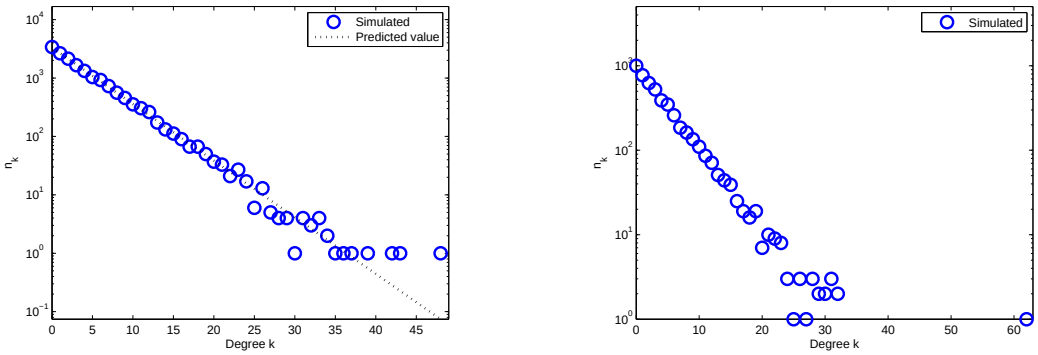
$$P_k = \binom{mN}{k} \left(\frac{1}{mN}\right)^k \left(1 - \frac{1}{mN}\right)^{N-k}. \tag{3.1}$$

Let $l_{i,j}$ be the length of a path from node i to node j where no nodes are visited more than once, and let L_n be the average number of such paths that have $l_{i,j} = n$. We can find solutions for the average of L_n over the network ensemble from the recursion

$$L_n = L_{n-1}\frac{m(N-n)}{N}. \tag{3.2}$$

The fraction on the left hand side is the probability that the next edge in the path does not link to any of its ancestor nodes in the path or to itself. We have $L_0 = N$ so

$$L_n = \frac{m^n}{N^{n-1}}\frac{(N-1)!}{[N-(n+1)]!}. \tag{3.3}$$



(a) $p = 2/3$, $q = 1/6$, $m = 4$. **(b)** $p = 9/10$, $q = 1/20$, $m = 4$.

Figure 3.2: The degree distribution of the network after 10^5 iterations, starting from an initial random network of 10 nodes. The line in (3.2a) shows the predicted result in Eq.(3.23). In (3.2b) an outlier exists owing to the high rate of local rewiring compared with the other mechanisms.

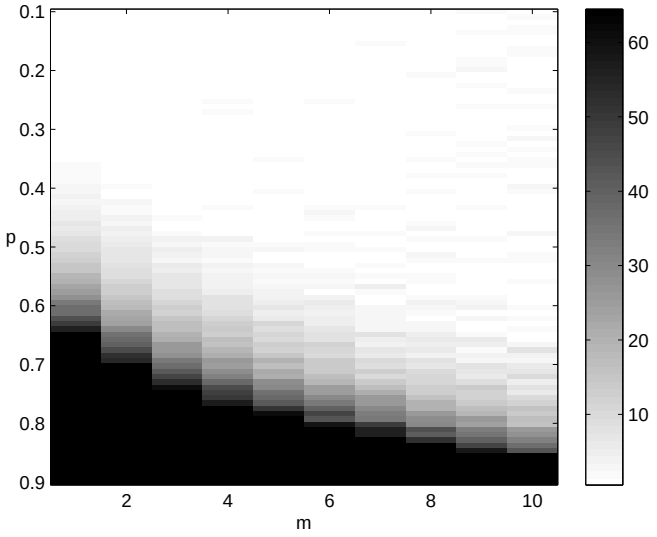
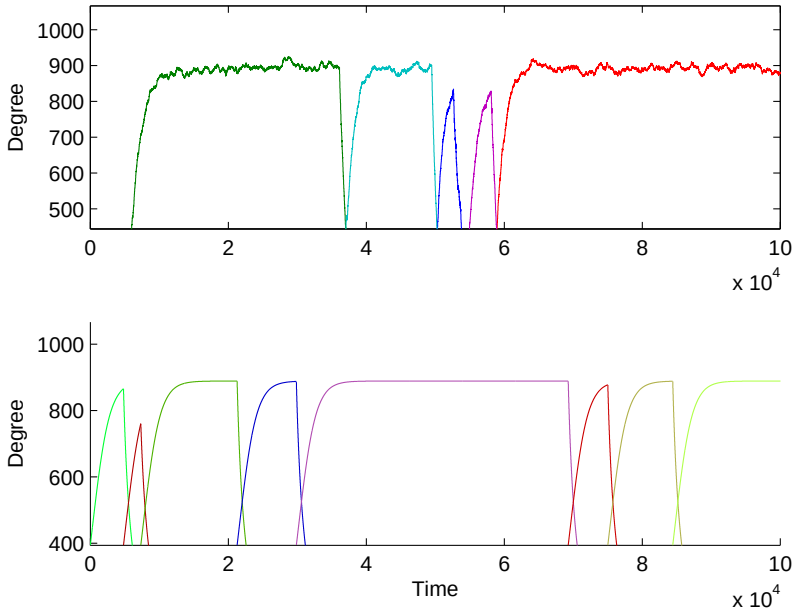
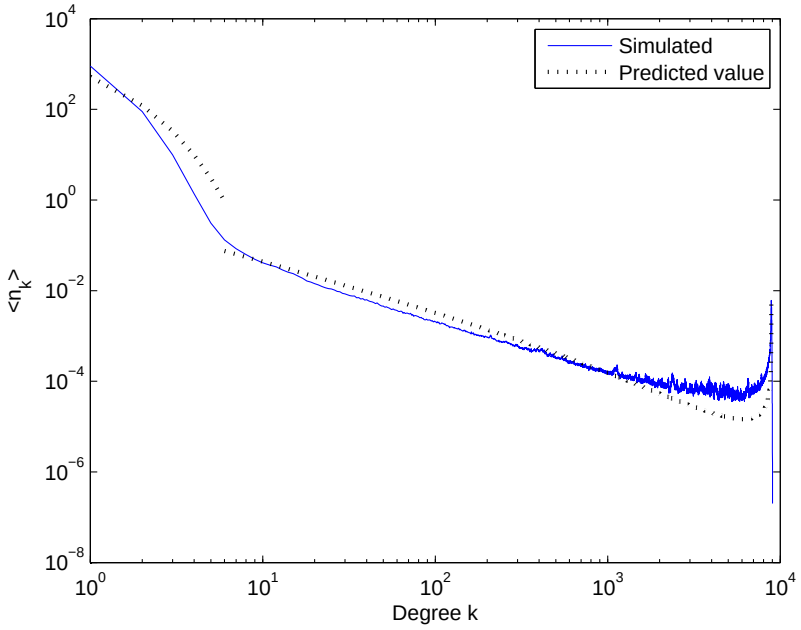


Figure 3.3: The level of agreement quantified by the Kolmogorov-Smirnov statistic between the prediction for the degree distribution Eq.(3.25) and the corresponding numerical simulation. The results presented are for the special case where growth is excluded i.e. when $r = 0$ and $q = 1 - p$, $N = 10^3$. We consider the model to be accurate up to a KS value of 20 since this is the value found when we test the prediction of Eq.(3.25) against data generated by pseudo-random numbers drawn from the same probability distribution.



(a) Time dependent dynamics of dominant nodes ($p = 0.9$, $N = 10^3$).



(b) Mean degree distribution of the whole network ($p = 0.9$, $N = 10^4$).

Figure 3.4: Shown here are results when $r = 0$ (no growth) and $m = 1$. Shown in (3.4a) is an example of how the degree of the most connected nodes changes over time (top), and the equivalent approximation using the method outlined in 3.6. The distribution is divided into two regimes; the exponential part when $\langle n_k \rangle \geq 1$ and the tail. The prediction comes from Eq.(3.37) for the first part and Eq.(3.32) for the second.

This also gives a formula for the average number of cycles C_n of length n

$$nC_n = L_{n-1}\frac{m}{N} \tag{3.4}$$

giving

$$C_n = \frac{m}{nN^{n-1}}\frac{(N-1)!}{(N-n)!}. \tag{3.5}$$

It is important to note that every network in this class will have at least one cycle and that every node either belongs to a cycle or is connected to a cycle by a directed path.

3.4 Degree distribution

In a single time-step the probability of attaching to a node i with degree k_i is

$$\Pi_a(k_i) = p\left[\frac{k_i}{mN}\right] + q\left[\frac{1}{N}\right] + r\left[\frac{m}{N}\right]. \tag{3.6}$$

This assumes that node degree correlations do not effect the attachment probability, i.e. the degree of a parent node of i is approximated well by the mean degree m. Therefore the number of edges that can potentially be rewired to i is mk_i, multiplying by the probability $1/m$ that once selected, i will be the node redirected to gives the first term on the left hand side of Eq.(3.6). The probability of removing an adjacent edge from i is

$$\Pi_r(k_i) = (p+q)\left[\frac{k_i}{mN}\right]. \tag{3.7}$$

We are interested in finding $n_k(t)$, the number of nodes with in-coming degree k. At $k = 0$

$$\frac{\partial n_0}{\partial t} = r + \frac{(p+q)}{mN}n_1 - \frac{q+rm}{N}n_0. \tag{3.8}$$

The terms on the right hand side respectively represent the addition of a node to the network, creation of a node of degree 0 by removing an edge from a node of degree 1, and destruction by attaching an edge and making it a node of degree 1.

Similarly for $k \geq 1$,

$$\frac{\partial n_k}{\partial t} = \frac{(p+q)}{mN}[(k+1)n_{k+1} - kn_k] + \frac{p}{mN}[(k-1)n_{k-1} - kn_k] + \frac{q+rm}{N}[n_{k-1} - n_k]. \tag{3.9}$$

The first pair of terms on the right hand side represent the mean change in n_k by either creating or destroying a node of degree k by removing one of its edges, the second pair are similar except for attachment by local rewiring, the third is for global rewiring.

As t grows large, the proportion of node of degree k will converge to constant values. Therefore in the asymptotic limit as $t \to \infty$ Eq.(3.9) reduces to the following second order recursion relation, found by substituting $N(t) = rt$ and $P_k = n_k(t)/N$.

$$\left[r + (q+rm) + \frac{2p+q}{m}k\right] P_k = [(q+rm) + \frac{p}{m}(k-1)]P_{k-1} + \frac{p+q}{m}(k+1)P_{k+1} \tag{3.10}$$

and Eq.(3.8) becomes

$$[q + (m+1)r]\, P_0 = r + \frac{p+q}{m}P_1. \tag{3.11}$$

We introduce the generating function

$$g(x) = \sum_{k=0} P_k x^k, \tag{3.12}$$

following the method outlined in the supplementary material at the end of this chapter we get

$$\left[\frac{-p}{m}x^2 + \frac{2p+q}{m}x - \frac{p+q}{m}\right] g'(x) + [-(q+rm)x + r + q + rm]g(x) \tag{3.13}$$
$$= -\frac{p+q}{m}P_1 + (r+q+rm)P_0.$$

The right hand side equates with Eq.(3.11) to give

$$\frac{p}{m}(1-x)\left(x - \frac{p+q}{p}\right) g'(x) + [r + (q+rm)(1-x)]g(x) = r \tag{3.14}$$

for $q, r \neq 0$. We solve by writing it as

$$g'(x) + \frac{m[r + (q + rm)(1 - x)]}{p(1 - x)\left(x - \frac{p+q}{p}\right)} g(x) = \frac{rm}{p(1 - x)\left(x - \frac{p+q}{p}\right)} \tag{3.15}$$

then multiplying both sides by the integrating factor

$$\begin{aligned} I &= \exp\left[\int \frac{m[r + (q + rm)(1 - x)]}{p(1 - x)\left(x - \frac{p+q}{p}\right)} dx\right] && (3.16) \\ &= (x - (p + q)/p)^{m[(q+rm)/p - r/q]} (x - 1)^{mr/q}, && (3.17) \end{aligned}$$

Eq.(3.14) can then be rewritten as

$$\begin{aligned} &\frac{d}{dx}\left[(x - 1)^{\lambda}(x - (p + q)/p)^{\mu} g(x)\right] && (3.18) \\ &\qquad = -\frac{rm}{p}(x - (p + q)/p)^{\mu - 1}(x - 1)^{\lambda - 1} && (3.19) \end{aligned}$$

where

$$\lambda(p, q, m) = \frac{rm}{q} \tag{3.20}$$

and

$$\mu(p, q, m) = m\left[\frac{q + rm}{p} - \frac{r}{q}\right]. \tag{3.21}$$

Notice that the terms in λ and μ are simply the ratios of the different rates of attachment by the three different mechanisms in the process.The solution is

$$g(x) = -\frac{rm}{p}(x - (p + q)/p)^{-\mu} \sum_{n=0}^{\infty} \binom{\mu - 1}{n} \left(-\frac{q}{p}\right)^{\mu - n - 1} \frac{(x - 1)^n}{n + \lambda} \tag{3.22}$$

To return the degree distribution P_k we equate the coefficients of x^k in the expansion of $g(x)$with Eq.(3.12). This is easily done when μ is a positive integer, for example when $\mu = 1$,

$$P_k = \frac{q}{p + q}\left(\frac{p}{p + q}\right)^k \tag{3.23}$$

and $\mu = 2$

$$P_k = \frac{m(1-p-q)}{(p+q)^2}\left(\frac{p}{p+q}\right)^k\left[\left(\frac{q}{1+\lambda}-\frac{q}{\lambda}\right)k-\left(\frac{p}{1+\lambda}+\frac{q}{\lambda}\right)\right]. \tag{3.24}$$

In fact when μ is any positive integer the form of P_k is the product of an exponential part and a polynomial in k of order μ. In the case of a network with fixed size N, $r = 0$, we solve Eq.(3.14) to find

$$P_k = \frac{p^k(1-p)^{m(1-p)/p}}{k!}(k-1-\alpha)(k-2-\alpha)...(-\alpha) \tag{3.25}$$

where

$$\alpha = -\frac{(1-p)m}{p}. \tag{3.26}$$

3.5 Dominant nodes

Consider the extreme example where $m = 1$ and $p = 1$, the steady state solution for the degree distribution is a network comprising of one node of degree N which is linked to by every node in the network including itself. Hence, as p approaches 1 we anticipate the existence of nodes with degree much higher than predicted in Section 3.4, and a possible alteration to the topology of the entire network. The mathematical formulation of the model in Section 3.4 (Equations (3.8) and (3.9)), did not account for this and so we model specifically the degree of the nodes which are likely to dominate the network. Previous work has examined the similar concept of gelation, where a gel node takes a finite proportion of the network's N nodes as N goes to infinity [40, 43]. To become dominant a node must belong to a subset of nodes called a 'rich-club'; a small set of nodes characterised by the large number of links between its members relative to the small number of links that leave the set [73]. In this section we present the equation that describes the dynamics of the total degree of the rich-club before taking a detailed look at the simplest case, when $m = 1$ and $r = 0$.

Let R be a subset of n_R nodes, let $k_R^{\text{in}}(t)$ denote the total number of in-coming edges adjacent to R and $k_R^{out}(t)$ the number of out-going edges. Using a continuum

approximation

$$\frac{\partial k_R^{\text{in}}}{\partial t} = [q + rm]\frac{n_R}{N} + p\frac{k_R^{\text{in}}}{m^2 N}\left(\frac{mN - k_R^{\text{in}}}{N}\right) - \left(\frac{k_R^{\text{out}}}{mn_R}p + q\right)\frac{k_R^{\text{in}}}{mN}. \tag{3.27}$$

The first term on the left hand side comes from attachment during growth or global rewiring, the second term comes from local rewiring and is the product of the probability that a second neighbour of R is selected, and the probability that once selected it will rewire to R (it assumes only one edge exits from the neighbour to R), the last term shows the decrease when one of the edges coming into R is rewired away, k_R^{out}/mn_R is the probability that the edge which guides the local rewiring is one that leaves the set R. When $N >> n_R$ and $O(1/N)$ terms are disregarded Eq.(3.27) becomes

$$\frac{\partial k_R^{\text{in}}}{\partial t} = \frac{k_R^{\text{in}}}{m^2 N}\left(p - p\frac{k_R^{\text{in}}}{N} - qm - p\frac{k_R^{\text{out}}}{n_R}\right). \tag{3.28}$$

If a set R exists such that this derivative is positive, i.e. if

$$k_R^{\text{in}} > N\left(\frac{qm}{p} - \left[1 - \frac{k_R^{\text{out}}}{n_R}\right]\right) \tag{3.29}$$

then the nodes in R will begin to dominate the network. However, the edges in this model are transient, and R will only maintain its structure until one of its internal edges is selected for rewiring. We investigate only the simple case where $r = 0$, $m = 1$ and $q = 1 - p$.

3.5a $r = 0$, $m = 1$ and $q = 1 - p$

Suppose R is a single node. Let $k(t) = k_R^{\text{in}}(t)$. The solution to Eq.(3.28) is

$$T_{\text{down}}(k_\tau, k) = \frac{N}{1-p}\ln\left[\frac{k_\tau}{N(1-p)/p + k_\tau}\frac{N(1-p)/p + k}{k}\right]. \tag{3.30}$$

Here $T_{\text{down}}(k_\tau, k)$ represents the average time taken for R to decrease from degree k_τ to k. Suppose R is self-cyclic (meaning that its one outgoing edge links back on itself). Now, if an edge adjacent to R is selected for local rewiring it will be rewired

to exactly the position it was in initially. The solution to Eq.(3.28) becomes

$$T_{\rm up}(k_\tau, k) = \frac{N}{2p-1} \ln \left[\frac{N(2p-1)/p - k_\tau}{k_\tau} \frac{k}{N(2p-1)/p - k} \right]. \tag{3.31}$$

Here $T_{\rm up}(k_\tau, k)$ represents the average time taken for R to increase from degree k_τ to k.

To predict the tail of the degree distribution $\langle n_k \rangle$ we assume that it is proportional to the expectation of the length of time for which a dominant node has degree k. Suppose i is a node of degree k_i which becomes self-cyclic. The probability that the degree of i will grow to size $k+1$ or greater is the probability that i will not be selected for global rewiring in $T_{\rm up}(k_i, k+1)$ consecutive iterations. Given that this occurs, the total time for which i has degree k is given by Eqs.(3.30) and (3.31). Putting this together we get

$$\langle n_k \rangle \approx C \left(1 - \frac{1-p}{N} \right)^{T_{\rm up}(k_i, k+1)} [T_{\rm up}(k, k+1) + T_{\rm down}(k+1, k)] \tag{3.32}$$

where C is the constant of proportionality and depends on k_i. In Section 3.6 we show how the mean of k_i can be approximated and the results are plotted in Fig.(3.4b). Eq.(3.32) only approximates the shape of the tail of the degree distribution, it should be noted that we have neglected the period of time for which a node i has degree k but its outgoing (self-cyclic) edge gets rewired before i reaches degree $k+1$, for this reason $\langle n_k \rangle$ quickly approaches infinity as k approaches its upper bound.

3.5b Effect on the rest of the network

Previously we have used the mean degree to approximate the number of second neighbours of any given node, and hence the attachment probability for local rewiring. In cases where a significant proportion of the edges are attached to a small number of dominant nodes the expectation of the number of second neighbours of a node is less and Eq.(3.25) fails to give an accurate prediction (see Fig.(3.3)). If we let $\langle k_{-e} \rangle$ be the mean degree of the network excluding any num-

ber of edges then the equivalent of Eq.(3.6) is

$$\Pi_a(k_i) = (1-p)\frac{1}{N} + p\frac{\langle k_{-e}\rangle k_i}{m^2 N} \tag{3.33}$$

which gives

$$mN\frac{\partial n_k}{\partial t} = (k+1)n_{k+1} - kn_k + (1-p)m[n_{k-1} - n_k] + \frac{p\langle k_{-e}\rangle}{m}[(k-1)n_{k-1} - kn_k]. \tag{3.34}$$

This can be solved in a similar way to before, but since we are not considering growth we can adopt a simpler method, used in [65], and assume that for large t a steady state has been reached and the left hand side is 0. Eq.(3.34) can be rewritten

$$kn_k - (k+1)n_{k+1} = \left[(1-p)m + \frac{p\langle k_{-e}\rangle}{m}(k-1)\right] n_{k-1} - \left[(1-p)m + \frac{p\langle k_{-e}\rangle}{m}k\right] n_k. \tag{3.35}$$

We immediately see that

$$kn_k = \left[(1-p)m + \frac{p\langle k_{-e}\rangle}{m}(k-1)\right] n_{k-1} \tag{3.36}$$

and so we find

$$n_k = \left(\frac{p\langle k_{-e}\rangle}{m}\right)^k \frac{1}{k!}\left(k-1-\alpha\right)\left(k-2-\alpha\right)...\left(-\alpha\right) n_0 \tag{3.37}$$

where

$$\alpha = -\frac{(1-p)m^2}{p\langle k_{-e}\rangle}. \tag{3.38}$$

Knowing that the sum over all k is N we also find

$$n_0 = N(1-p)^{-\alpha}. \tag{3.39}$$

3.6 Estimating the mean degree of dominant nodes

We consider a model that describes the time dependent behaviour of the dominant nodes with the following certain simplifying assumptions:

1. At any time there will be exactly one self-cyclic node whose degree increases according to Eq.(3.31).
2. The times for which nodes remain self-cyclic are geometrically distributed with mean $N/(1-p)$.
3. After the out-edge of a self-cyclic node is rewired globally its degree decreases according to Eq.(3.30).

Additionally we assume that the degree of a node when it initially becomes self-cyclic is k_0, which we find by simultaneously solving

$$k_{\text{top}} = \frac{N\beta}{\left(\frac{N\beta}{k_0} - 1\right)\exp\left(\frac{1-2p}{1-p}\right) + 1} \tag{3.40}$$

where k_{top} is the degree of a self-cyclic node after the average amount of time it remains cyclic (from Eq.(3.31)), $\beta = (2p-1)/p$, and

$$k_0 \approx \frac{k_{\text{top}}^2}{2N}. \tag{3.41}$$

To understand this approximation consider that when global rewiring of the self-cyclic node occurs, it may rewire to form a 2-cycle with probability k_t/N, then when local rewiring happens on one of the edges in the 2-cycle a self-cyclic node is created and the expectation of its degree is $k_t/2$. If this does not occur then we assume that the new self-cyclic node has small degree (close enough to 0 to be ignored). Eq.(3.41) is the expected outcome of those two possibilities. Solving Eqs. (3.40) and (3.41) gives

$$k_0 = \frac{N}{2}\left(\frac{\beta}{2(1-\theta)}\left[1 + \sqrt{1 - \frac{8\theta(1-\theta)}{\beta}}\right]\right)^2 \tag{3.42}$$

where

$$\theta = \exp\left(\frac{1-2p}{1-p}\right). \tag{3.43}$$

Through numerical investigation we determine that k_0 is real valued for $p > 0.77$.

The expectation of the number of nodes that have degree $k > k_0$ at any time t is given by the length of time a self-cyclic node has degree k divided by the mean

length of time a node remains self-cyclic. For $k > k_0$,

$$n_k \approx \frac{1-p}{N}\left(1-\frac{1-p}{N}\right)^{T_{\rm up}(k_0,k+1)}[T_{\rm up}(k,k+1)+T_{\rm down}(k+1,k)]. \qquad (3.44)$$

The average number of edges linking to dominant nodes is

$$\langle k_{-e}\rangle = \sum_{k'=k_0}^{N} k' n_{k'}. \qquad (3.45)$$

Fig.(3.4a) compares the model described here, and the mean found from simulating the actual model.

3.7 Conclusion

The model presented is one of the simplest possible treatments of rewiring in directed networks and although we have not related it to any particular application, these results add to the understanding of this class of network as a whole. We have looked at local rules that naturally lead to the preferential selection of nodes for attachment, and global rules that select nodes randomly. Edges are selected with equal probability for rewiring which leads to nodes being selected proportionally to their degree. The combined effect of these two mechanisms is a network with predominantly a exponential degree distribution. The vast majority of nodes do not accumulate edges to create a long (power-law) tail. Instead we find a small number of dominant nodes which conspire to develop an immunity to local detachment causing a large number of links to condense around them.

Appendix: Solving the first order recursion relation

This section details the first part of the solution to Eq.(3.10). For the recursion relation

$$[m_0 k + c_0]P_k + [m_{-1}(k-1) + c_{-1}]P_{k-1} + [m_1(k+1) + c_1]P_{k+1} = 0 \tag{3.46}$$

first multiply by x^k

$$m_0 k P_k x^k + c_0 P_k x^k + m_{-1}(k-1)P_{k-1}x^k + c_{-1}P_{k-1}x^k + m_1(k+1)P_{k+1}x^k + c_1 P_{k+1}x^k = 0. \tag{3.47}$$

Rewrite this as

$$\begin{aligned} x m_0 k P_k x^{k-1} + c_0 P_k x^k + x^2 m_{-1}(k-1)P_{k-1}x^{k-2} + x c_{-1}P_{k-1}x^{k-1} \\ + m_1(k+1)P_{k+1}x^k + x^{-1}c_1 P_{k+1}x^{k+1} = 0. \end{aligned} \tag{3.48}$$

Summing over $k \geq 1$

$$\begin{aligned} x m_0 \sum_{k=1} k P_k x^{k-1} + c_0 \sum_{k=1} P_k x^k + x^2 m_{-1} \sum_{k=0} k P_k x^{k-1} + x c_{-1} \sum_{k=0} P_k x^k \\ + m_1 \sum_{k=2} k P_k x^{k-1} + x^{-1} c_1 \sum_{k=2} P_k x^k = 0. \end{aligned} \tag{3.49}$$

Introduce the generating function

$$g(x) = \sum_{k=0} P_k x^k \tag{3.50}$$

and we have

$$x m_0 g'(x) + c_0[g(x) - P_0] + x^2 m_{-1} g'(x) + x c_{-1} g(x) + m_1[g'(x) - P_1] + x^{-1} c_1[g(x) - P_0 - P_1 x] = 0 \tag{3.51}$$

or

$$[x^2 m_{-1} + x m_0 + m_1]g'(x) + [x c_{-1} + c_0 + x^{-1} c_1]g(x) = [m_1 + c_1]P_1 + [c_0 + c_1 x^{-1}]P_0 \tag{3.52}$$

Since $g(1) = 1$ and $g'(1) = \langle k \rangle$

$$[m_{-1} + m_0 + m_1]\langle k \rangle + [c_{-1} + c_0 + c_1] = [m_1 + c_1]P_1 + [c_0 + c_1]P_0 \tag{3.53}$$

Chapter 4

Node splitting: a fragmentation process for nodes

We introduce a collection of complex networks generated by a combination of preferential attachment and a previously unexamined process of 'splitting' nodes. The splitting mechanism involves selecting a node of degree k, either randomly or preferentially, then fragmenting it into k nodes of degree 1. Using the rate equation method we derive the degree distributions for several models, the first of which is a power-law with an exponent tunable to very large values. We also consider the case where only one node of degree 1 is created through splitting and the remaining edges are rewired. Lastly we introduce a splitting mechanism where a node of degree k will fragment to 2 nodes of degree k' and $k - k'$.

The foundation for the work presented in this chapter is Barabási and Albert's *preferential attachment* model. Here a new node is created at each time-step and linked to m existing nodes in the network, by design the likelihood of linking to a node of degree k is proportional to k. After many iterations the proportion of nodes which have degree k has been shown to have power-law behaviour: $P(k) \sim 2m^2k^{-3}$ where $P(k)$ is the proportion of nodes having degree k [8]. An extension of this model incorporates the addition of links between existing nodes [74]; originally introduced to describe the social network of scientific collaborations, it has also been used to model the interactions of words in human language [30]. The degree distribution for this model still follows a power-law although it is now composed of two regimes divided by a critical point where the exponent changes.

In the broader field of statistical mechanics, a substantial body of research concerns the coalescence and fragmentation of clusters of particles, applications in this field span a variety of subjects including astrophysics [75], polymerization [76] and aerosols [77]. Despite the diversity of applications the basic model remains the same; two clusters containing either a number, in the discrete case, or mass, in the continuous, of identical particles of sizes x and y *coalesce* at a rate $K(x, y)$ into a cluster of size $x + y$. In the discrete setting, the number of clusters of size x at time t denoted by $n(x, t)$ obeys the *Smolochowski coagulation equation*,

$$\frac{\partial}{\partial t}n(x,t) = \frac{1}{2}\sum_{y=1}^{x-1} K(y, x-y)n(y,t)n(x-y,t) - n(x,t)\sum_{y=1}^{\infty} K(x,y)n(y,t) \quad (4.1)$$

where the first term on the right hand side accounts for the creation of a cluster of size x from the coalescence of two smaller clusters and the second term accounts for the loss of a cluster of size x when it coalesces with another. Exact general solutions have not been found, however in the special cases where $K(x, y) = 1$ for example, representing two clusters coalescing at each time step, and $K(x, y) = xy$,

where clusters coalesce at a rate proportional to their size, exact solutions do exist [78]. Conversely, equivalent equations for fragmentation are constructed in a similar way. If coalescence and fragmentation are simultaneously present in a model then complete fragmentation or complete coalescence into one supercluster can be avoided, in this case the distribution of cluster sizes at large t is independent of t. A model of this type has been used to describe the herding behaviour of traders in financial markets [79]. Here, traders are the particles of the system and clusters represent groups of traders sharing information and therefore trading in the same way. The clusters of this model coalesce over time and at random times will rapidly fragment into unclustered individuals. It was shown that the size of the clusters at large t follows a power-law distribution with exponential cut-off, it has also been proposed as a possible reason why variations in share price do not follow a Gaussian distribution [80]. The networks presented in this chapter extend the lexicon of complex networks by translating the previous model into a network environment. By considering link formation between nodes to be equivalent to coalescence, and by introducing a new process that we shall refer to as 'splitting' to parallel the fragmentation process described above, we reproduce the cluster size distribution as a network degree distribution.

We define splitting as the replacement of a single node of degree k with k nodes of degree one (see Fig.(4.1)) and examine the topologies of networks created through this splitting process alongside other growth processes. The evolution of the networks studied here are driven also by the preferential attachment mechanisms outlined in [81], first in Section 4.1, where new nodes are linked to existing nodes in the network chosen with probability proportional to their degree, and secondly in Section 4.2 edges are attached between pairs of existing nodes, again with probability proportional to their degree.

We use $n_k(t)$ to denote the number of nodes of degree k at time t and introduce the following two quantities

$$N(t) = \sum_{k=1}^{\infty} n_k(t) \tag{4.2}$$

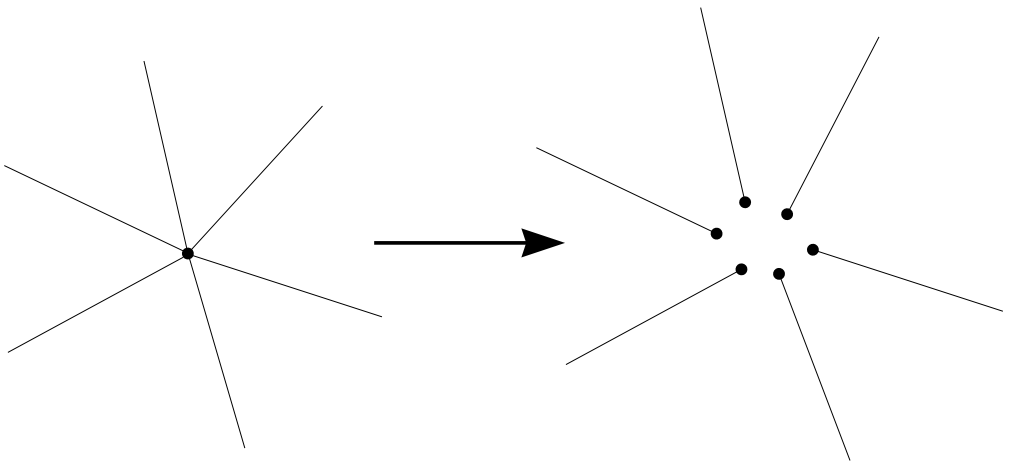


Figure 4.1: The effect of "splitting" a node of degree 6 in a network.

and

$$M(t) = \sum_{k=1}^{\infty} k n_k(t) \tag{4.3}$$

where $N(t)$ is the total number of nodes and $M(t)$ is the total degree of the network at time t.

4.1 Splitting in a node attachment model

4.1a Random splitting

At each time step the network may develop in one of the two following ways:

(a) With probability p, a node is introduced and attached by an edge to an existing node, the probability that the end of the edge attaches to a node of degree k is proportional to k.

(b) With probability $1-p$, a node of degree k is randomly selected and split into k nodes of degree 1.

At time t, for nodes with degree $k \geq 2$, $n_k(t)$ evolves according to

$$\frac{\partial n_k}{\partial t} = \frac{p}{M}[-k n_k + (k-1) n_{k-1}] - \frac{1-p}{N} n_k. \tag{4.4}$$

The first term on the right comes from the loss of a node of degree k that happens when the new edge is attached to it, the second term comes from the creation of

a node of degree k when the new edge attaches to a node of degree $k-1$. The last term comes from the loss of a node of degree k when it is split into k nodes of degree 1. For nodes of degree $k = 1$,

$$\frac{\partial n_1}{\partial t} = -\frac{p}{M} n_1 + p + \frac{(1-p)}{N} \sum_{k=2}^{\infty} k n_k. \tag{4.5}$$

We can can substitute Eq.(4.4) and Eq.(4.5) into the rate equation for $M(t)$,

$$\frac{\partial M}{\partial t} = \frac{\partial n_1}{\partial t} + \sum_{k=2}^{\infty} k \frac{\partial n_k}{\partial t}, \tag{4.6}$$

to verify that $M(t) = 2pt$. Similarly, the rate equation for $N(t)$ is found to be

$$\frac{\partial N}{\partial t} = (1-p)\left(\frac{M}{N} - 1\right) + p. \tag{4.7}$$

Assuming $N(t)$ grows linearly with time i.e $N(t) = \alpha t$, where α is a time independent constant, Eq.(4.7) becomes

$$\alpha = (1-p)\left(\frac{2p}{\alpha} - 1\right) + p \tag{4.8}$$

hence

$$\alpha = \frac{2p - 1 + \sqrt{1 + 4p - 4p^2}}{2} \tag{4.9}$$

It should be noted that α ranges between 0 at $p = 0$ and increases to 1 at $p = 1$, the slow rate of growth in the size of the network owes to the large likelihood of a node of degree 1 being selected for splitting resulting in no change to the network. For this reason we find the fastest rate of growth at $p = 1$ (the usual preferential attachment model without splitting).

The probability P_k ($\equiv n_k/N$) that a randomly selected node will have degree k is solved by substituting $n_k(t) = \alpha t P(k)$ into Eq.(4.4) giving

$$P_k = \frac{1}{2}[-kP_k + (k-1)P_{k-1}] - \frac{1-p}{\alpha} P_k \tag{4.10}$$

and so

$$P_k = \frac{k-1}{k+2+4(1-p)/(2p-1+\sqrt{1+4p-4p^2})} P_{k-1} \tag{4.11}$$

thus

$$P_k \sim k^{-\gamma} \text{ where } \gamma = 3 + \frac{4(1-p)}{2p-1+\sqrt{1+4p-4p^2}}. \tag{4.12}$$

Comparing this to the case when $p = 1$, we see that while the power-law structure is not changed by the process of splitting, the exponent can take any value greater than 3 depending on the value of p.

4.1b Preferential splitting

We consider a modification of this model; the same preferential attachment of nodes occurs with probability p, however with probability $1 - p$ nodes are also preferentially selected for splitting with probability proportional to their degree. The rate equations are

$$\frac{\partial n_k}{\partial t} = \frac{p}{M}[-kn_k + (k-1)n_{k-1}] - \frac{1-p}{M} kn_k \tag{4.13}$$

for $k \geq 2$ and

$$\frac{\partial n_1}{\partial t} = -\frac{p}{M} n_1 + p + \frac{(1-p)}{M} \sum_{k=2}^{\infty} k^2 n_k. \tag{4.14}$$

Following similar analysis to above the distribution of node degrees in this network is found to follow

$$P_k \sim p^{k-1} k^{-\gamma} \text{ where } \gamma = 1 + 2p \tag{4.15}$$

where the power-law behaviour is only recovered when $p = 1$ and we return to the Barabási-Albert [8, 81].

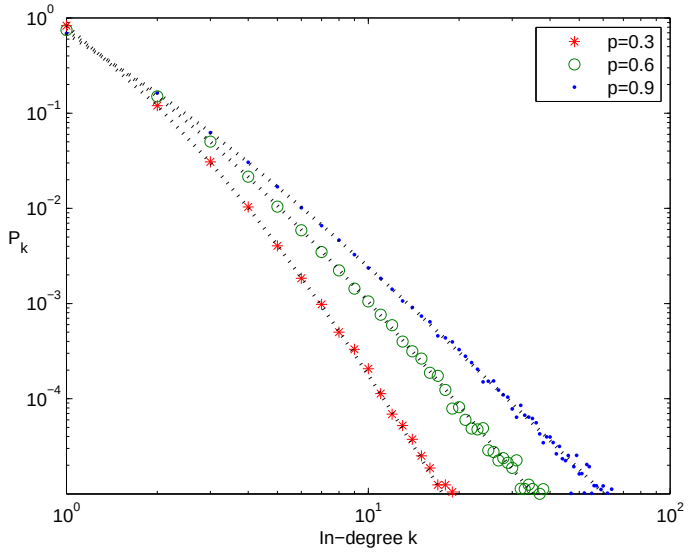


Figure 4.2: The degree distribution for the first splitting model Eq.(4.11). Each simulation was run for 10^5 iterations starting from an initial network consisting of 2 nodes and 1 edge. We show the means over 10^2 trials. The dotted lines are the predicted results.

4.1c Components

We compute the number of components $Q(t)$ of the network, that is how many disconnected pieces there are at time t, this changes according to the rate equation

$$\frac{\partial Q}{\partial t} = (1-p)\sum_{k=1}^{\infty}(k-1)P(k) \tag{4.16}$$

$$= (1-p)\left(\frac{M}{N}-1\right) \tag{4.17}$$

since the number of components will increase only when a node of degree k is selected for splitting, causing Q to increase by $k-1$ with probability P_k. Solving gives

$$Q = (1-p)\left(\frac{2p}{\alpha}-1\right)t \tag{4.18}$$

$$= \frac{-1+\sqrt{1+4p-4p^2}}{2}t \tag{4.19}$$

concluding that the number of components grows at its fastest rate when $p = 1/2$, with $Q(t) = (\sqrt{2}-1)t/2$. The model in the following section can be seen as a modification of the node attachment model: with probability p, edges are introduced to the network and each end is attached to a node of degree k with probability proportional to k, nodes of degree 1 are no longer introduced to the network and thus are generated only when splitting occurs (with probability $1-p$), we find that a combination of both processes is necessary for the network to grow indefinitely.

4.2 Splitting in an edge attachment model

There are two processes that may occur at each time step in the construction of this network

(a) With probability p, an edge is attached between two existing nodes, the probability that each end of the edge attaches to a node of degree k is proportional to k.

(b) With probability $1-p$, a node of degree k is randomly selected and split into k nodes of degree 1.

The rate equation for the behaviour of the number of nodes with degree k, $n_k(t)$, is

$$\frac{\partial n_k}{\partial t} = \frac{2p}{M}[(k-1)n_{k-1} - kn_k] - \frac{1-p}{N}n_k \tag{4.20}$$

for $k \geq 2$. The first two terms on the right hand side represent the preferential attachment process seen also in Section 4.1, the last term accounts for the loss of a node of degree k by being selected for splitting. When $k = 1$ the rate equation is

$$\frac{\partial n_1}{\partial t} = -\frac{2p}{M}n_1 + \frac{1-p}{N}\sum_{k=2}^{\infty} kn_k. \tag{4.21}$$

The first term on the right hand side accounts for the loss of a node of degree 1 that occurs when the new edge is linked to it, the second term accounts for the increase caused by splitting a node of degree k into k nodes of degree 1. As before,

we substitute Eq.(4.20) and Eq.(4.21) into Eq.(4.6) to find $M(t) = 2pt$, and also

$$\frac{\partial N}{\partial t} = \frac{\partial n_1}{\partial t} + \sum_{k=2}^{\infty} \frac{\partial n_k}{\partial t} = (1-p)\frac{M-N}{N}. \tag{4.22}$$

Assuming $N(t)$ grows linearly with time i.e $N(t) = \beta t$, where β is a time independent constant, we have

$$\beta = (1-p)\left(\frac{2p}{\beta} - 1\right) \tag{4.23}$$

with the solution

$$\beta = \frac{p - 1 + \sqrt{(1-p)(1+7p)}}{2}. \tag{4.24}$$

In contrast to the node attachment model, the limiting values $p = 0$ and $p = 1$ both produce networks that do not grow with time ($\beta = 0$), the rate of increase of $N(t)$ here has its maximum of $\beta = 2(2\sqrt{2} - 1)/7$ at $p = (3 + \sqrt{2})/7$. It is now possible to find the probability P_k that a randomly selected node will have degree k by solving Eq.(4.20). First, note that for large t, $n_k = \beta t P_k$, then Eq.(4.20) becomes

$$P_k = (k-1)P_{k-1} - kP_k - \frac{1-p}{\beta}P_k \tag{4.25}$$

giving

$$P_k = \frac{k-1}{k+1+(1-p)/\beta}P_{k-1} \tag{4.26}$$

thus

$$P_k \sim k^{-\gamma} \text{ where } \gamma = 2 + \frac{2(1-p)}{p - 1 + \sqrt{(1-p)(1+7p)}}. \tag{4.27}$$

Again the edge attachment model can be modified in such a way that the candidates for splitting are selected preferentially, the distribution is found in a similar way to those in the previous sections:

$$P_k \sim \left(\frac{2p}{1-p}\right)^{k-1} k^{-\gamma} \text{ where } \gamma = 1 + \frac{2p}{1-p}. \tag{4.28}$$

Exponential cut-off is present for all values of p except when $p = 1/3$, instead the distribution follows a power-law with exponent -2.

Table 4.1: The results of the previous two sections are reviewed here. "Random splitting" refers to those models where the candidates for splitting have been selected with equal probability whereas "preferential" refers to models that use preferential selection.

Model	Size (N)	Degree distribution P_k	Number of components
Node attachment, random splitting	$\left\lfloor \frac{2p-1+\sqrt{1+4p-4p^2}}{2} \right\rfloor t$	$k^{-\gamma}$, $\gamma = 3 + \frac{4(1-p)}{2p-1+\sqrt{1+4p-4p^2}}$	$\frac{-1+\sqrt{1+4p-4p^2}}{2} t$
Node attachment, preferential splitting	-	$p^{k-1}k^{-\gamma}$, $\gamma = 1 + 2p$	$\frac{-1+\sqrt{1+4p-4p^2}}{2} t$
Edge attachment, random splitting	$\left\lfloor \frac{p-1+\sqrt{(1-p)(1+7p)}}{2} \right\rfloor t$	$k^{-\gamma}$, $\gamma = 2 + \frac{2(1-p)}{p-1+\sqrt{(1-p)(1+7p)}}$	-
Edge attachment, preferential splitting	-	$\left(\frac{2p}{1-p}\right)^{k-1} k^{-\gamma}$, $\gamma = 1 + \frac{2p}{1-p}$	-

The degree distributions for each of the models studied here are shown in Table 4.1 along with the number of nodes and number of components for certain models. We compare the degree distribution of the edge attachment model with preferential selection for splitting with that of the cluster equivalent studied in [80]. This model evolves by assembling agents into clusters at each time step by the usual coalescence process with probability a, or with probability $1-a$ a node is selected and the cluster containing it is split, by which we mean a cluster of size k becomes k clusters of size 1. It was shown that the distribution of cluster sizes follows a power-law with an exponential cut-off, the power-law exponent is $-5/2$. We observe that the process of attaching edges in the network model is comparable to the coalescence of clusters, and the splitting processes of the two different formulations are similar, the degree distribution also follows a power-law with exponential cut-off and at the value $p = 3/7$ the exponent becomes $-5/2$. The models presented in this paper demonstrate splitting as a mechanism to necessitate the creation of nodes, thus the network grows indefinitely contrary to the cluster based model which has a fixed number of agents. Additionally, at each time step the degree of a selected node may only increase by at most one whereas during coalescence clusters can potentially grow by any amount, these differences prove to be enough not to allow an exact equivalence between the distribution of cluster sizes and the degree distribution.

4.3 Splitting and rewiring

We show that the *rewiring* mechanism, discussed in the previous chapter, can be incorporated into a similar model. The creation of nodes of degree 1 can easily be replaced with an equivalent global rewiring process. We introduce a model parametrised only by the number of nodes N and the number of edges E. The network does not grow in time but instead develops by the following rule:

> At each time-step a node of degree k is randomly selected, $k-1$ of its edges are rewired to existing nodes with probability proportional to their degree.

The mean degree is given by

$$\langle k \rangle = \frac{2E}{N}. \tag{4.29}$$

The rate equations are

$$\frac{\partial n_k}{\partial t} = \frac{\langle k-1 \rangle}{2E}[(k-1)n_{k-1} - kn_k] - \frac{n_k}{N} \tag{4.30}$$

for $k \geq 2$ and

$$\frac{\partial n_1}{\partial t} = 1 - \frac{n_1}{N} - \frac{\langle k-1 \rangle}{2E} n_1. \tag{4.31}$$

To solve for n_k we put the left hand sides to zero (since there is no growth) then rearrange to find

$$n_k = \frac{k-1}{k + \langle k \rangle / \langle k-1 \rangle} n_{k-1} \tag{4.32}$$

and

$$n_1 = \frac{N}{2 - 1/\langle k \rangle}, \tag{4.33}$$

which give the solution

$$n_k = \frac{N\Gamma(\gamma)}{2 - N/2E} \frac{\Gamma(k)}{\Gamma(k+\gamma)}, \tag{4.34}$$

where

$$\gamma = 1 + \frac{2E}{2E - N}. \tag{4.35}$$

For large k, the network follows a power-law degree distribution with scaling exponent γ. Models of this type, that produce scale-free networks of finite size, are currently rare, and previously studied models [65, 43] have not been able to achieve the range of exponents found here. Results are plotted in Fig.(4.3).

4.4 Splitting by mitosis

While the type of splitting we have introduced is a useful mechanism for generating networks of various topologies, a direct application is yet to be found. Other forms of splitting are more natural, in particular the process where a mode of degree k is split into two nodes of degrees k' and $k - k'$. This *mitosis* mechanism describes what we are likely to observe when a networked cell of some type splits into two. The converse of this process has been studied in [82] where it was shown that

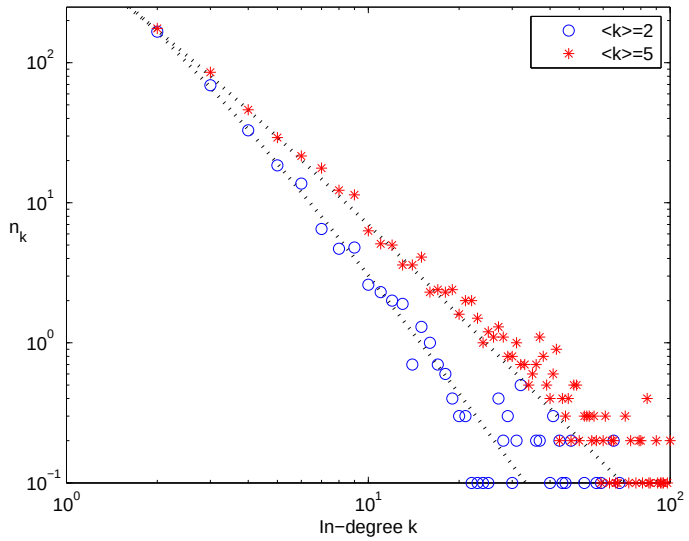


Figure 4.3: The degree distribution for the splitting-rewiring model. The network has 10^3 nodes, the simulation was run for 10^5 iterations and averaged over 10 trials.

repeated aggregation of nodes creates a scale-free network. In fact, if we were to combine the aggregation method described there with the splitting mechanism in section 4.1a we would get the exact network equivalent of the cluster dynamics of [80]. Performing the associated calculations the network equivalent was found to be too similar to the previously studied model to be of any particular interest.

We introduce the simplest possible mitosis model, where we perform the splitting step in a way that makes the associated equations easy to solve:

> At each time step introduce an edge and attach each end to nodes in the network with probability proportional to their degree. If a node is found to have degree $L+1$ then we divide the node into two, with degrees s and $L+1-s$, where s can be 1, 2 ... L with equal probability.

Time t here is equivalent to E the number of edges since we add exactly one in each time step. Growth comes from the splitting of nodes and so we need only one parameter L. Observing that the expectation of the increase in nodes of degree k by splitting a node of degree L is $2/L$, and this happens with probability $2Ln_L/2E$

with each added edge, we can write

$$\frac{\partial n_k}{\partial E} = \frac{2}{2E}[(k-1)n_{k-1} - kn_k] + \frac{4}{2E}n_L. \tag{4.36}$$

If we assume that N grows linearly at rate α, and $n_k = NP_k$, Eq.(4.36) reduces to

$$(k+1)P_k = (k-1)P_{k-1} + 2P_L \tag{4.37}$$

which has the solution $P_k = 1/L$ for $1 \geq k \geq L$. We can easily find that $N = 4E/L$. This seemingly trivial result is worth remarking on. There is no immediate intuitive reason why the solution should be so simple. In fact, when we replace the preferential attachment condition with one of random attachment, giving

$$(\alpha+2)P_k = 2P_{k-1} + \frac{2}{L}P_L \tag{4.38}$$

where α is the rate of growth, we find the less remarkable result

$$P_k = \frac{2}{L}\left[1 - \left(\frac{2}{\alpha+2}\right)^k\right]. \tag{4.39}$$

Simulations of both models are plotted in Fig.(4.4).

4.5 Conclusion

Mechanisms that drive the evolution of networks is a topic of interest that is relevant in many different fields, in this work we have introduced one such mechanism and deduced some of the macroscopic qualities of its resulting network. The effect of *splitting* on the growth of a network is not immediately obvious and using two different kinds of preferential attachment models we have deduced some important network properties at large t. We have found that by adjusting the frequency of splitting events on the node attachment network we can affect its growth and create scale free networks with degree distributions that are not achievable through preferential attachment alone. On the edge attachment model splitting drives the creation of new nodes, the maximum growth is reached when $p = (3+\sqrt{2})/7$, as before, the frequency of splitting events decreases the proportion of nodes of high

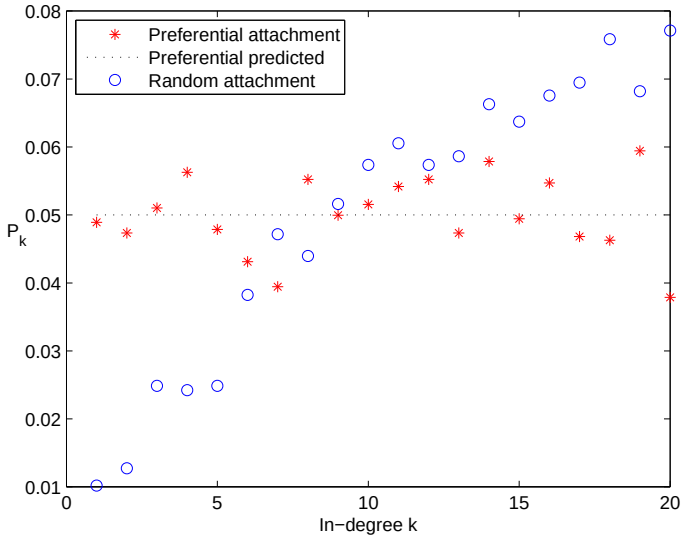


Figure 4.4: The mitosis model with both random attachent and preferential attachment. Nodes split into 2 nodes when their degree exceeds 20. The figure shows the degree distribution after 10^4 iterations.

degree. Through rewiring we have created a scale-free network without growth, but also has scaling exponents in the range of most of the scale-free networks observed in the real world. Finally we have shown that limiting the growth of a node by splitting it when it reaches a maximal degree considerably affects the network structure.

Chapter 5

Electrical resistance and random Fibonacci sequences

We consider a particular electrical network with the structure of a random tree: starting from a root vertex, in one iteration each leaf (a vertex with zero or one adjacent edges) of the tree is extended by either a single edge with probability p or two edges with probability $1-p$. With each edge having a resistance equal to 1, the total resistance R_n between the root vertex and a busbar connecting all the vertices at the n^{th} level is considered. Representing R_n as a dynamical system we show that $\langle R_n \rangle$ approaches $(1+p)/(1-p)$ as $n \to \infty$, the distribution of R_n at large n is also examined. Additionally, expressing R_n as a random sequence, we find that it is related to the Legendre polynomials and that it converges to the mean with $|\langle R_n \rangle - (1+p)/(1-p)| \sim n^{-1/2}$.

For several years random sequences have been a topic of interest for a number of researchers. While this body of work has been accepted as a branch of statistical physics, the current literature is primarily focused on problems of a purely mathematical conception, namely the idea of a random Fibonacci sequence introduced in [83] and expanded on in [84], [85] and [86]. The natural response to these analyses is to consider areas in applied science where a random sequence may be characteristic of the phenomena being studied, these are disordered systems whose behaviour is non-deterministic in that the state of the system after a short step in time could be any of a number of possibilities (according to certain probabilities), much in an analogous way to a random Fibonacci sequence. One success of statistical mechanics has been the widespread utilization of complex (random) networks to model naturally occurring phenomena, for this reason random sequences that mimic the properties of random network problems have the potential to become a fruitful topic of research.

The example considered here extends a number of well studied problems involving networks of electrical resistors, [87] and [88] are concerned with the resistance between two sites on a lattice where each edge is a resistor, and [89] goes further by examining the percolation that occurs when these resistors are overloaded, destroying the corresponding edge and breaking the lattice. On a similar theme, this chapter attempts to find the resistance across a particular class of random network where each edge represents a resistor. This chapter concerns a theoretical application of the equations associated with electrical resistance, the aim being to find results regarding the total resistance of a random network where each edge represents a resistor, the particular problem chosen has provided an opportunity to show that random sequences can be useful in studying problems in physics.

5.1 Finding the resistance of a randomly grown tree

Here we study a network grown from a single vertex by the repeated process of appending either one branch or two (with probabilities p and $1-p$ respectively) to those vertices created in the previous iteration (see Fig.(5.1)), the tree after n steps is denoted T_n [$\equiv T_n(p)$]. To simplify the problem all edges are chosen to have a resistance of 1Ω, throughout the chapter the units of resistance Ω will not be displayed. In this chapter we ask the following question: as a function of p, what is the resistance R_n [$\equiv R_n(p)$] between the root vertex and a busbar connecting all the vertices at the n^{th} level, and what happens when $n \to \infty$? The problem is

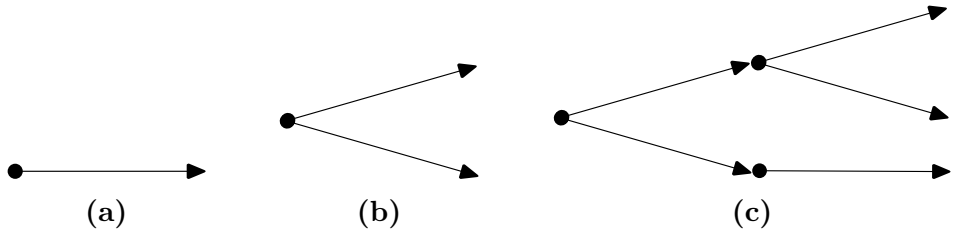


Figure 5.1: The network begins as a single vertex and at each iteration one or two new edges are attached to each 'leaf' of the tree, the tree may grow along a single branch as in Fig.(5.1a) with probability p, or split in to two branches as in Fig.(5.1b) with probability $1-p$. Fig.(5.1c) shows what a tree may look like after 2 iterations, this particular realization will occur with probability $(1-p)\times(1-p)\times p$. At the far right of the tree all the leaf vertices will, by design, join to a single busbar to complete the circuit, therefore the network can be seen as a complex combination of resistors in series and in parallel, Fig.(5.1c) for example will have total resistance 6/7 (assuming the resistance across each edge is 1).

interesting since the equations governing electrical resistance will take a different form depending on whether a given vertex branches in two or not - if it does then the formula for the resistance across the two parallel edges with resistances R_1 and R_2 given by

$$\frac{1}{R_{Total}} = \frac{1}{R_1} + \frac{1}{R_2}$$

is used, edges connected in series use the formula $R_{Total} = R_1 + R_2$. The random combination of these equations makes the question of the total resistance difficult to solve, moreover the problem increases rapidly in complexity as the network grows.

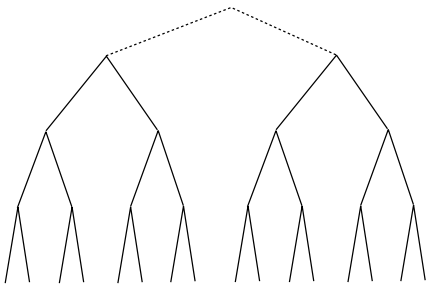


Figure 5.2: $T_n(0)$ is equivalent to joining together two copies of $T_{n-1}(0)$ by two parallel edges to a new root vertex.

The remainder of this chapter describes ways in which these problems are mitigated and approximate solutions are found for $\langle R_n \rangle$, its distribution $P_n(R)$ as well as the rate of convergence to the mean as n increases. In Section 5.2 the exact solutions for the two special cases, $p = 0$ and $p = 1$, are presented. In Section 5.3 a simplified model is used to approximate $T_n(p)$ and the mean and second moment are approximated for general p and large n. Section 5.4 introduces a method to generate T_n accurately and the corresponding numerical results are compared with those of 5.3. We present a random sequence model in Section 5.5 from which we obtain the convergence towards the mean as n increases.

5.2 Solution in the limiting cases

At the extreme values, $p = 1$ and $p = 0$, upper and lower bounds for $R_n(p)$ are easily found: in the first case there is no branching so $T_n(1)$ is composed of a line of n edges connected in series; supposing the network grows with n, the equation for resistance in series gives

$$R_n = R_{n-1} + 1$$

and so $R_n \sim n$ as $n \to \infty$. In the second case the network branches at every vertex (thus $T_n(0)$ is a *complete binary tree*) and so both equations for resistance in series and in parallel are needed. As illustrated in Fig.(5.2), T_n is equivalent to joining two networks, T_{n-1}, by two parallel edges from the root vertex to a newly created root vertex (to verify this, observe that T_n has 2^n end points (leaves) and $2 \times T_{n-1}$

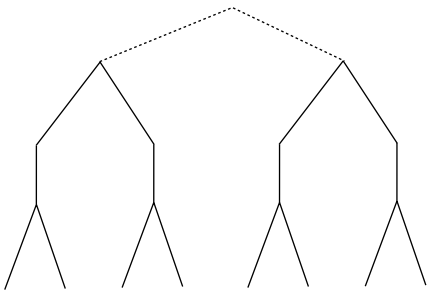


Figure 5.3: In the simplified model at each iteration the network is either extended by an edge from the root vertex or joined with a duplicate of itself as shown.

has $2 \times 2^{n-1}$). The consequent resistance equation is

$$\begin{aligned} \frac{1}{R_n} =& \frac{1}{R_{n-1}+1} + \frac{1}{R_{n-1}+1} \\ \Rightarrow R_n =& \frac{R_{n-1}+1}{2}, \end{aligned}$$

the solution to this being $R_n \to 1$ as $n \to \infty$.

5.3 A simplified model to approximate R_n

When $0 < p < 1$ we wish to find the mean $\langle R_n \rangle$ as a function of p as well as higher moments and also the distribution $P_n(R)$. Since this is not easily obtained, the simplified network T_n^* is studied. In this section, for the resistance R of T_n^* when $n \to \infty$ the first and second moments are found and the distribution $P(R)$ is expressed in two different forms.

With probability p, T_n^* is constructed by joining the root vertex of T_{n-1}^* to a new vertex or, with probability $1-p$, T_n^* is found by joining two duplicates of T_{n-1}^* to a new vertex (see Fig.(5.3)). Since this network retains the same proportion of split branches as T_n one expects it to be a close approximation. The corresponding resistances are given by

$$R_n = \begin{cases} R_{n-1}+1 & \text{with probability } p \\ \frac{R_{n-1}+1}{2} & \text{with probability } 1-p \end{cases} \tag{5.1}$$

We can immediately write down a recursive formula for the average,

$$\langle R_n \rangle = \frac{1+p}{2} \left[\langle R_{n-1} \rangle + 1 \right], \tag{5.2}$$

which indicates the steady state value for which R_n converges to: $\langle R \rangle = (1 + p)/(1 - p)$. Solving Eq.(5.2) gives

$$\langle R_n \rangle = \left(\frac{1+p}{2} \right)^n R_0 + \frac{2}{p-1} \left(\frac{1+p}{2} \right)^{n+1} + \frac{1+p}{1-p} \tag{5.3}$$

which clearly shows that $\langle R_n \rangle$ converges at an exponential rate. Importantly however, since the distribution of all possible values $P_n(R)$ here has been neglected, and only the mean $\langle R_n \rangle$ was used in this derivation, we discard this result in favour of the conflicting results shown in Section 5.5 which do take into account the distribution of R_n.

The distribution of R at iteration n obeys

$$P_n(R) = \int P_{n-1}(R')dR' \left[(1-p)\delta \left(\frac{R'+1}{2} - R \right) + p\delta \left(R' + 1 - R \right) \right]$$

where $\delta(x)$ is the Dirac delta function. This simplifies to

$$P_n(R) = 2(1-p)P_{n-1}(2R-1) + pP_{n-1}(R-1). \tag{5.4}$$

For the second moment, following directly from Eq.(5.4),

$$\langle R_n^2 \rangle = \int_1^\infty R^2 P_n(R)dR = 2(1-p) \int_1^\infty R^2 P_{n-1}(2R-1)dR + p \int_1^\infty R^2 P_{n-1}(R-1)dR$$

is solved with changes of variable, $u = R - 1$ and $v = 2R - 1$, and the knowledge that for any natural number n, $\int_0^1 P_n(R)dR = 0$ and $\int_1^\infty P_n(R)dR = 1$ which follows from Eq.(5.1). The resulting recursive formula is

$$\langle R_n^2 \rangle = \frac{1+3p}{4} (\langle R_{n-1}^2 \rangle + 2\langle R_{n-1} \rangle + 1), \tag{5.5}$$

as $n \to \infty$ the second moment is found to be $\langle R^2 \rangle = (3 + 10p + 3p^2)/3(1 - p)^2$.

Additionally, $P_n(R)$ converges to an invariant distribution as n increases to infinity, from Eq.(5.4) the invariant distribution satisfies

$$P(R) = 2(1-p)P(2R-1) + pP(R-1). \tag{5.6}$$

Using $\tilde{P}(k)$ to denote the Laplace transform of $P(R)$, the solution of Eq.(5.6) when transformed,

$$\tilde{P}(k) = \mathcal{L}(P(R)) = 2(1-p)\int_1^\infty P(2R-1)e^{-kR}dR + p\int_1^\infty P(R-1)e^{-kR}dR,$$

simplifies to the recursive equation

$$\begin{aligned}
\tilde{P}(k) =& \frac{(1-p)e^{k/2}}{1-pe^{-k}}\tilde{P}(k/2) \\
=& \frac{(1-p)e^{-k/2}}{1-pe^{-k}}\frac{(1-p)e^{-k/4}}{1-pe^{-k/2}}\tilde{P}(k/4) \\
& \vdots \\
=& \prod_{r=0}^{\infty}\frac{(1-p)e^{-k/(2^{r+1})}}{1-pe^{-k/(2^r)}}\tilde{P}(k) = e^{-k}\prod_{r=0}^{\infty}\frac{1-p}{1-pe^{-k/2^r}}
\end{aligned} \tag{5.7}$$

The inverse Laplace transform will recover an expression for $P(R)$, this is described as

$$P(R) = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT}\tilde{P}(k)e^{Rk}dk \tag{5.8}$$

$$= \text{the sum of the residues of } \tilde{P}(k)e^{Rk}. \tag{5.9}$$

These residues reside at the points on the complex plane where the denominator in Eq.(5.7) is equal to zero, i.e for the s^{th} root

$$k_s = 2^s\log(p) \text{ for } s = 0, 1, 2, ... \tag{5.10}$$

The residue is calculated by taking the limit

$$\begin{aligned}\text{Res}[\tilde{P}(k),k_s] =& \lim_{k\to 2^s\log(p)}\left[(k-2^s\log(p))\times e^{Rk}\prod_{r=0}^{\infty}\frac{(1-p)e^{-k/(2^{r+1})}}{1-pe^{-k/(2^r)}}\right] && (5.11)\\ =& e^{2^s\log(p)R}\prod_{r\neq s}^{\infty}\frac{(1-p)e^{-2^s\log(p)/(2^{r+1})}}{1-pe^{-2^s\log(p)/(2^r)}} && (5.12)\\ &\times(1-p)e^{-2^s\log(p)/(2^{s+1})} && (5.13)\\ &\times\lim_{k\to 2^s\log(p)}\frac{k-2^s\log(p)}{1-pe^{-k/(2^s)}}. && (5.14)\end{aligned}$$

In the limit in the above expression, both numerator and denominator tend towards zero, therefore L'Hopital's rule is applicable, giving

$$\begin{aligned}\lim_{k\to 2^s\log(p)}\frac{k-2^s\log(p)}{1-pe^{-k/(2^s)}} =& \lim_{k\to 2^s\log(p)}\frac{1}{(p/2^s)e^{-k/(2^s)}} && (5.15)\\ =& 2^s. && (5.16)\end{aligned}$$

With this, and some tidying up of Eq.(5.11), the expression for the s^{th} residue becomes

$$\begin{aligned}\text{Res}[\tilde{P}(k),k_s] =& e^{2^s\log(p)R}(1-p)e^{-2^{-1}\log(p)}\prod_{r\neq s}^{\infty}\frac{(1-p)e^{-2^{s-r-1}\log(p)}}{1-pe^{-2^{s-r}\log(p)}}\times 2^s && (5.17)\\ =& 2^s p^{2^s R}\frac{1-p}{\sqrt{p}}\prod_{r\neq s}^{\infty}\frac{(1-p)p^{-2^{s-r-1}}}{1-p^{1-2^{s-r}}} && (5.18)\end{aligned}$$

and so

$$P(R) = \sum_{s=0}^{\infty}2^s p^{2^s(R-1)}(1-p)\prod_{r\neq s}^{\infty}\frac{(1-p)}{1-p^{1-2^{s-r}}}. \qquad (5.19)$$

We now simplify Eq.(5.19) with an approximation. Using the expansion $1/(1-X) = 1+X+X^2+...$ with $X = pe^{-k/2^r}$, Eq.(5.7) can be written

$$\tilde{P}(k) = e^{-k}\prod_{r=0}^{\infty}(1-p+pe^{-k/2^r}-p^2e^{-k/2^r}+p^2e^{-k/2^{r-1}}-...)$$

Focusing only on terms up to and including multiples of p^2, multiplying out the

brackets and recalling the translation property of $\delta(x)$,

$$\begin{aligned}\tilde{P}(k) \approx & e^{-k} + p\sum_{r=0}^{\infty}(e^{-k(1+1/2^r)} - e^{-k}) \\ & + p^2\left(\sum_{r=0}^{\infty}(e^{-k(1+1/2^{r-1})} - e^{-k(1+1/2^r)}) \right. \\ & \left. + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(e^{-k(1+1/2^i+1/2^j)} - e^{-k(1+1/2^i)} - e^{-k(1+1/2^j)} + e^{-k})\right)\end{aligned}$$

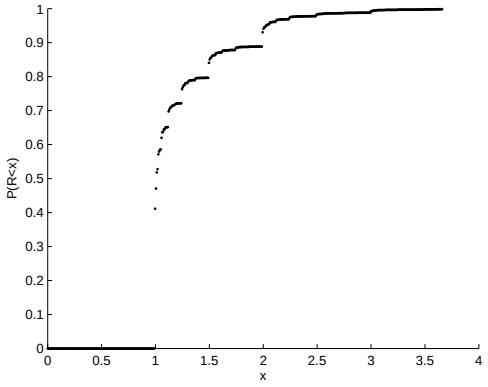
can be expressed as

$$\begin{aligned}\tilde{P}(k) &\approx \int e^{-kR}\delta(R-1)dR + p\sum_{r=0}^{\infty}\int e^{-kR}\delta(R-1-1/2^r)dR - ... \\ &= \int e^{-kR}P(R)dR\end{aligned}$$

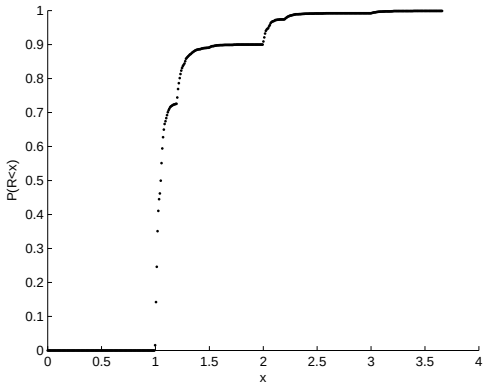
where

$$\begin{aligned}P(R) =& \delta(R-1) + p\sum_{r=0}^{\infty}[\delta(R-1-1/2^r) - \delta(R-1)] \\ & + p^2\sum_{r=0}^{\infty}[\delta(R-1-1/2^{r-1}) - \delta(R-1-1/2^r)] \\ & + p^2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left[\delta(R-1-1/2^i-1/2^j) - \delta(R-1-1/2^i)\right. \\ & \left. -\delta(R-1-1/2^j) + \delta(R-1)\right]. \qquad (5.20)\end{aligned}$$

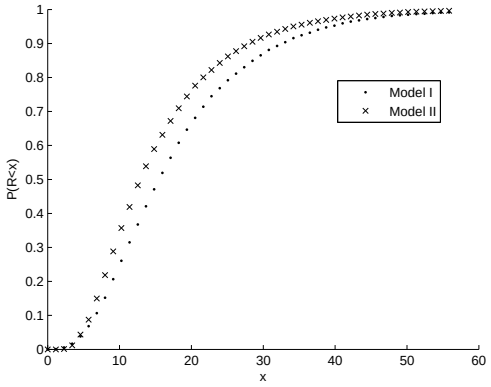
As values of p go towards zero, the values identified by the delta functions in the above expression constitute an increasingly significant proportion of $P(R)$. This can be seen in Fig.(5.4a) with the largest probability occurring at $R = 1$, corresponding to the first order term as well as lower order terms, other notable values of R correspond to the values identified by the delta functions that are multiplied by p^2 in Eq.(5.20), 2, 1.5, 1.25, etc..



(a)



(b)



(c)

Figure 5.4: Cumulative probability distribution plotted with the output of 10^4 realizations of the simplified model (Model I) in Fig.(5.4a) and the accurate model (Model II) in Fig.(5.4b), $n = 10^6$ with $p = 0.1$. Fig.(5.4c) shows both the simplified and accurate models together, this time with $p = 0.9$.

5.4 Comparison of the resistances of T_n^* and T_n: numerical results

In this section we compare the results of Section 5.3 with numerical results generated by simulation of the dynamical system Eq.(5.1) (Table 5.1). Additionally we introduce a second system which accurately reproduces the behaviour of R_n for the original network T_n. The similarity of the simplified model with the accurate representation is then shown by a comparison of each models mean (Fig.(5.5) and Table 5.1), variance (Table 5.1) and distribution (Figs.(5.4a),(5.4b) and (5.4c)) generated numerically.

5.4a Constructing T_n

With probability p, T_n is constructed by joining the root vertex of T_{n-1} to a new vertex or, with probability $1-p$, T_n is found by joining two trees T_a and T_b to a single root vertex, where both T_a and T_b are possible realizations of T_{n-1}. The resistance of T_n is then given by

$$R_n = \begin{cases} R_{n-1}+1 & \text{with probability } p \\ \dfrac{1}{1/(R_a+1)+1/(R_b+1)} & \text{with probability } 1-p \end{cases} \tag{5.21}$$

where R_a and R_b are distributed according to $P_{n-1}(R)$. Comparisons are shown in table 5.1 and Fig.(5.5), for the simple model it was shown that the mean converges to $(1+p)/(1-p)$ and the sum of the squares converges to $(3+10p+3p^2)/3(1-p)^2$ from which the variance is calculated.

5.5 Using random sequences to predict the convergence of R_n

To obtain the rate at which $\langle R_n \rangle$ converges to the mean we consider a third model. Expressing $\langle R_n \rangle$ as a recurrence relation we show it to be equivalent to a well known family of orthogonal polynomials, these polynomials are expressed as an integral

Table 5.1: A comparison is made between the simplified model and the accurate model, here the results for the mean and variance of the resistance in the simple model (Eq.(5.1)) predicted in Section 5.3 are shown in the first two columns, the next two columns show numerical results obtained from 10^6 realizations of the same model, the last two columns show numerical results for the accurate model (Eq.(5.21)). Numerical results of the accurate model are also shown in Fig.(5.5), illustrating the accuracy of the simplified model as an approximation to the accurate one.

p	Predicted		Model I (Simplified)		Model II (Accurate)	
	Average	Variance	Average	Variance	Average	Variance
0.1	1.22222	0.164618	1.22266	0.173222	1.19601	0.132801
0.2	1.5	0.416667	1.49595	0.422622	1.43467	0.337122
0.3	1.85714	0.816326	1.86260	0.850533	1.74874	0.666383
0.4	2.33333	1.48148	2.32556	1.45536	2.17774	1.19880
0.5	3	2.66667	3.00448	2.68774	2.72414	2.10841
0.6	4	5	3.96578	4.77015	3.60162	3.92301
0.7	5.66667	10.3704	5.67442	10.4559	5.04931	8.19541
0.8	9	26.6667	8.98651	26.7466	8.36457	23.7019
0.9	19	120	18.8040	116.624	17.3566	104.912

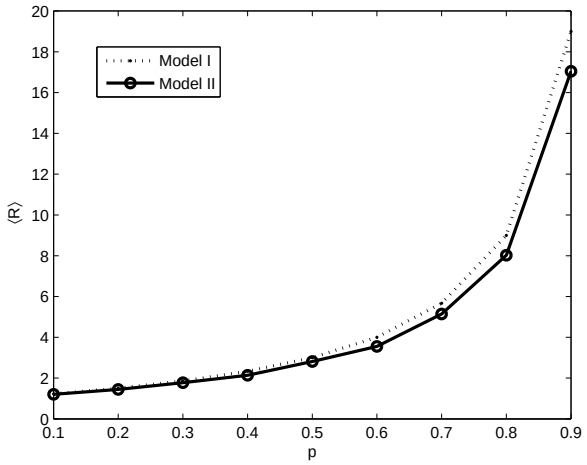


Figure 5.5: Comparison of the accurate model and the simplified model ($n = 10^6$).

which is solved to retrieve the average as a function of p and n when n is large. In this model the tree T_n is either extended by an edge from the root vertex as before (with probability p) or two duplicates of T_q are connected to a newly created root, where q is a randomly selected integer from $[0, n-1]$. The resistance is then given by following the random sequence

$$R_n = \begin{cases} R_{n-1} + 1 & \text{with probability } p \\ \dfrac{R_q + 1}{2} & \text{with probability } 1-p \text{ and } q \in [0, n-1] \end{cases} \tag{5.22}$$

If the T_q are chosen with equal probability then for large n one would expect the distribution of R_q to approach that of R_n, this specifically describes a system of either attaching a single edge to the root vertex of T_n (with probability p) or selecting a previous T_q and attaching it to its own duplicate (with probability $1-p$). Letting $Q_n(q)$ be the probability that the value $q \in [0, n-1]$ is chosen, the distribution of R_n obeys the integral equation

$$\begin{aligned} P_n(R) = &\int (1-p) \sum_{q=0}^{n-1} Q_n(q) P_q(R') \delta \left(R - \frac{R'+1}{2} \right) dR' \\ &+ \int p P_{n-1}(R) \delta(R - (R'+1)) dR' \end{aligned}$$

which reduces to

$$P_n(R) = 2(1-p) \sum_{q=0}^{n-1} Q_n(q) P_q(2R-1) + p P_{n-1}(R-1)$$

From this the average $A_n = \langle R_n \rangle = \int R P_n(R) dR$ is found to obey

$$A_n = (1-p) \sum_{q=0}^{n-1} Q_n(q) \left(\frac{A_q + 1}{2} \right) + p(A_{n-1} + 1). \tag{5.23}$$

A similar argument to the following can be found in [86], the distribution $Q_n(q)$ can be written as $Q_n(q) = Q(q)/b_n$ where

$$b_n = \sum_{q=0}^{n-1} Q(q). \tag{5.24}$$

Then Eq.(5.23) becomes

$$A_n = \frac{(1-p)}{b_n} \sum_{q=0}^{n-1} Q(q) \left(\frac{A_q + 1}{2} \right) + p(A_{n-1} + 1) \tag{5.25}$$

Subtracting Eq.(5.25) from the equivalent equation for A_{n+1}, and observing from Eq.(5.24) that $Q(n) = b_{n+1} - b_n$, it is found that

$$, 2b_{n+1}A_{n+1} + 2pb_nA_{n-1} = (p+1)[(b_{n+1} + b_n)A_n + (b_{n+1} - b_n)]. \tag{5.26}$$

In the case where the previous T_n are chosen with equal probability, $Q_n(q) = 1/n$ for all $q \in [0, n-1]$, $b_n = n$ ($b_{n+1} = n+1$, obviously) and Eq.(5.26) becomes

$$2(n+1)A_{n+1} + 2pnA_{n-1} = (p+1)[(2n+1)A_n + 1] \tag{5.27}$$

equivalently

$$nA_n = \frac{p+1}{2}[(2n-1)A_{n-1} + 1] - p(n-1)A_{n-2}. \tag{5.28}$$

A solution can be obtained with the help of some known results in orthogonal polynomials, this is possible by first observing that the transformation

$$A_n = \frac{1+p}{1-p} + p^{\frac{n}{2}} B_n \tag{5.29}$$

when substituted into Eq.(5.28) leaves

$$nB_n = \frac{p+1}{2\sqrt{p}}(2n-1)B_{n-1} - (n-1)B_{n-2}, \tag{5.30}$$

the recursion relation for the Legendre polynomials $B_n = P_{n-1}(x)$ at $x = (p+1)/2\sqrt{p}$ [90].

$$P_n(x) = \frac{1}{\pi} \int_0^{\pi} [x + \sqrt{x^2 - 1} \cos \phi]^n d\phi. \tag{5.31}$$

This can be easily solved using Laplace's method as it can be expressed in the form

$$P_n(x) = \frac{1}{\pi} \int_0^{\pi} \exp\{nf(\phi)\} d\phi$$

with

$$f(\phi) = \log[x + \sqrt{x^2-1}\,\cos\,\phi]$$

$$f'(\phi) = -\frac{\sqrt{x^2-1}\,\sin\,\phi}{x + \sqrt{x^2-1}\,\cos\,\phi} \tag{5.32}$$

$$f''(\phi) = -\frac{\sqrt{x^2-1}\,\cos\,\phi}{x + \sqrt{x^2-1}\,\cos\,\phi} + \frac{(x^2-1)\,\sin^2\phi}{[x + \sqrt{x^2-1}\,\cos\,\phi]^2}. \tag{5.33}$$

From Eq.(5.32) it can be seen that stationary points of f exist at $\phi = 0$ and $\phi = \pi$, putting these into Eq.(5.33) yields

$$f''(0) = -\frac{\sqrt{x^2-1}}{x + \sqrt{x^2-1}} \tag{5.34}$$

$$\text{and } f''(\pi) = \frac{\sqrt{x^2-1}}{x + \sqrt{x^2-1}} \tag{5.35}$$

Using Laplace's method on the integral in Eq.(5.31), as $n \to \infty$

$$\begin{aligned} P_n(x) &\sim \frac{1}{\pi}\frac{1}{2}\sqrt{\frac{2\pi}{n|f''(\phi_0)|}}\,\exp\{nf(\phi_0)\} \\ &= \frac{1}{2\pi}\sqrt{\frac{2\pi}{n}}\sqrt{\frac{x+\sqrt{x^2-1}}{\sqrt{x^2-1}}}(x+\sqrt{x^2-1})^n \end{aligned}$$

and so

$$P_n\left(\frac{1+p}{2\sqrt{p}}\right) \approx \frac{1}{\sqrt{n\pi}}\sqrt{\frac{1}{1-p}}p^{-n/2}.$$

Note that only half of the value is taken since the maximum is on the boundary of the integral. Relating this back to the formula for the average value of the random sequence [Eq.(5.29)],

$$A_{n+1} = \frac{1+p}{1-p} + p^{\frac{n+1}{2}}P_n\left(\frac{p+1}{2\sqrt{p}}\right),$$

the resulting equation showing the rate of convergence

$$A_{n+1} = \frac{1+p}{1-p} + \frac{1}{\sqrt{\pi n}}\sqrt{\frac{p}{1-p}}, \tag{5.36}$$

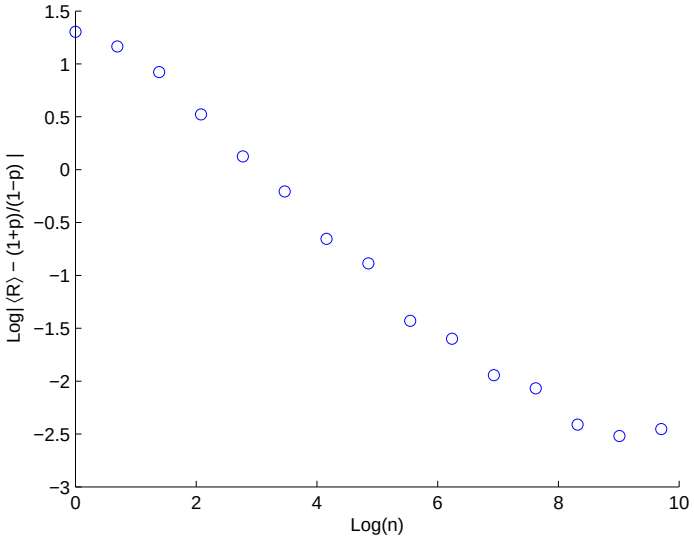


Figure 5.6: Convergence of random sequence representation of $\langle R_n \rangle$, here $p = 0.7$. The gradient in this plot agrees with the predicted result in Eq.(5.36) that the convergence to the mean behaves as $n^{-1/2}$.

also retrieves the value for $\langle R_n \rangle$ for large n seen in Section 5.3.

5.6 Conclusion

We have introduced a class of random resistor networks with a tree-like structure characterized by a single parameter p. By constructing similar yet simplified tree networks that retained the important property of the proportion of branching points to non-branching points, approximations were made to the resistance of the network with a slight loss of accuracy. For a growing network it was established for the approximation that when $p = 1$ the resistance diverges but for all other values of p the resistance converges to $(1+p)/(1-p)$ with $|\langle R_n \rangle - (1+p)/(1-p)| \sim n^{-1/2}$. Our analysis revealed that the structure of the probability distribution of R_n when n is large is intricate and as p decreases certain values begin to dominate the distribution.

Chapter 6

A final remark

Networked systems are affecting life more than ever. Through the growth of networks in transport and communication, recent decades have witnessed the emergence of a global society which, in principle, cuts through many of the barriers which have prevented success, popularity or wealth from reaching those who most deserve it. To believe, however, that a more networked society will facilitate a fairer distribution of success is naively optimistic. Self-organising networks are a facilitator of inequality. We have shown how the locally concentrated ways in which networks typically develop naturally lead to cumulative advantage effects. In Chapter 2 these mechanisms alone managed to reproduce the hugely asymmetric distribution of success in scientific literature. The analysis provides a mathematical foundation for the development of tools to create a fairer, more objective measure of the intrinsic quality of scientific work using the available data.

Citation networks are by no means unique: similar local triadic processes are observed in social networks and recommendation systems. The results of Chapter 3 show that there are many aspects to the dynamics of popularity in a network, an area where very little is currently understood. Through the mathematical analysis of models we can uncover the hidden factors that determine the dynamics of node connectivity and begin to interpret data and manage systems in an informed way. The contributions made throughout this work will hopefully prove valuable in achieving this goal.

References

[1] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1, pp. 107–117, 1998.

[2] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[3] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[4] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Apr 1998.

[5] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.

[6] I. Stewart, *Seventeen Equations that Changed the World*. Profile, 2012.

[7] D. Parlett, *The Oxford history of board games*, vol. 5. Oxford University Press New York, 1999.

[8] A.-L. Barabsi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[9] H. A. Simon, "On a class of skew distribution functions," *Biometrika*, vol. 42, no. 314, pp. 425–440, 1955.

[10] D. D. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science*, vol. 27, no. 5, pp. 292–306, 1976.

[11] G. Eugene, "Citation indexes for science," *Science*, vol. 122, no. 3159, pp. 108–111, 1955.

[12] R. K. Merton, "The Matthew effect in science," *Science*, vol. 159, no. 3810, pp. 56–63, 1968.

[13] S. Redner, "How popular is your paper? an empirical study of the citation distribution," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 4, pp. 131–134, 1998.

[14] R. Albert, H. Jeong, and A.-L. Barabási, "Internet: Diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.

[15] W. Willinger, D. Alderson, and J. C. Doyle, *Mathematics and the internet: A source of enormous confusion and great potential*. Defense Technical Information Center, 2009.

[16] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *SIGCOMM Comput. Commun. Rev.*, vol. 29, pp. 251–262, Aug. 1999.

[17] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[18] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, pp. 3200–3203, Apr 2001.

[19] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 7, pp. 2015–2020, 2006.

[20] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, "The dynamics of protest recruitment through an online network," *Scientific reports*, vol. 1, 2011.

[21] L. Weng, F. Menczer, and Y.-Y. Ahn, "Predicting successful memes using network and community structure," *arXiv preprint arXiv:1403.6199*, 2014.

[22] M. P. van den Heuvel and O. Sporns, "Rich-club organization of the human connectome," *The Journal of neuroscience*, vol. 31, no. 44, pp. 15775–15786, 2011.

[23] V. Nicosia, P. E. Vértes, W. R. Schafer, V. Latora, and E. T. Bullmore, "Phase transition in the economically modeled growth of a cellular nervous system," *Proceedings of the National Academy of Sciences*, vol. 110, no. 19, pp. 7880–7885, 2013.

[24] J. Terborgh and J. A. Estes, *Trophic cascades: predators, prey, and the changing dynamics of nature*. Island Press, 2010.

[25] D. Garlaschelli, G. Caldarelli, and L. Pietronero, "Universal scaling relations in food webs," *Nature*, vol. 423, no. 6936, pp. 165–168, 2003.

[26] C. Boone, H. Bussey, and B. J. Andrews, "Exploring genetic interactions and networks with yeast," *Nature Reviews Genetics*, vol. 8, no. 6, pp. 437–449, 2007.

[27] M. Vidal, M. E. Cusick, and A.-L. Barabasi, "Interactome networks and human disease," *Cell*, vol. 144, no. 6, pp. 986–998, 2011.

[28] D. J. d. S. Price, "Networks of scientific papers," *Science*, vol. 149, pp. 510–515, July 1965.

[29] R. V. Sole and M. Montoya, "Complexity and fragility in ecological networks," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1480, pp. 2039–2045, 2001.

[30] S. N. Dorogovtsev and J. F. F. Mendes, "Language as an evolving word web," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1485, pp. 2603–2606, 2001.

[31] K. Falconer, *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2013.

[32] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, no. 7024, pp. 392–395, 2005.

[33] P. W. Holland and S. Leinhardt, "A method for detecting structure in sociometric data," *American Journal of Sociology*, vol. 76, no. 3, p. 492, 1970.

[34] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random structures & algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[35] M. E. Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.

[36] S. L. Feld, "Why your friends have more friends than you do," *AJS*, vol. 96, no. 6, pp. 1464–77, 1991.

[37] R. Cohen, S. Havlin, and D. Ben-Avraham, "Efficient immunization strategies for computer networks and populations," *Physical Review Letters*, vol. 91, no. 24, p. 247901, 2003.

[38] P. Krapivsky and S. Redner, "Finiteness and fluctuations in growing networks," *Journal of Physics A: Mathematical and General*, vol. 35, no. 45, p. 9517, 2002.

[39] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, Jan 2002.

[40] P. L. Krapivsky and S. Redner, "Organization of growing random networks," *Phys. Rev. E*, vol. 63, p. 066123, May 2001.

[41] A. Vázquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations," *Physical Review E*, vol. 67, no. 5, p. 056104, 2003.

[42] J. Saramäki and K. Kaski, "Scale-free networks generated by random walkers," *Physica A: Statistical Mechanics and its Applications*, vol. 341, pp. 80–86, 2004.

[43] T. Evans and J. Saramäki, "Scale-free networks from self-organization," *Physical Review E*, vol. 72, no. 2, p. 026138, 2005.

[44] L. Bornmann and H.-D. Daniel, "What do citation counts measure? a review of studies on citing behavior," *Journal of Documentation*, vol. 64, no. 1, pp. 45–80, 2008.

[45] L. Leydesdorff and S. Milojević, *Scientometrics*, pp. 1–20. Elsevier, 2012.

[46] K. Börner, J. T. Maru, and R. L. Goldstone, "The simultaneous evolution of author and paper networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5266–5273, 2004.

[47] M. Wang, G. Yu, and D. Yu, "Measuring the preferential attachment mechanism in citation networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 18, pp. 4692 – 4698, 2008.

[48] Z.-X. Wu and P. Holme, "Modeling scientific-citation patterns and other triangle-rich acyclic networks," *Phys. Rev. E*, vol. 80, p. 037101, Sep 2009.

[49] P. Holme and B. J. Kim, "Growing scale-free networks with tunable clustering," *Phys. Rev. E*, vol. 65, p. 026107, Jan 2002.

[50] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[51] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, KDD '05, (New York, NY, USA), pp. 177–187, ACM, 2005.

[52] F.-X. Ren, H.-W. Shen, and X.-Q. Cheng, "Modeling the clustering in citation networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 12, pp. 3533 – 3539, 2012.

[53] M. V. Simkin and V. P. Roychowdhury, "Read before you cite!," *arXiv preprint cond-mat/0212043*, 2002.

[54] M. V. Simkin and V. P. Roychowdhury, "Copied citations create renowned papers?," *arXiv preprint cond-mat/0305150*, 2003.

[55] A. Vazquez, "Statistics of citation networks," *arXiv preprint cond-mat/0105031*, 2001.

[56] B. Tadic, "Dynamics of directed graphs: the world-wide web," *Physica A: Statistical Mechanics and its Applications*, vol. 293, no. 12, pp. 273 – 284, 2001.

[57] P. L. Krapivsky, G. J. Rodgers, and S. Redner, "Degree distributions of growing networks," *Phys. Rev. Lett.*, vol. 86, pp. 5401–5404, Jun 2001.

[58] R. Toivonen, J.-P. Onnela, J. Saramki, J. Hyvnen, and K. Kaski, "A model for social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 371, no. 2, pp. 851 – 860, 2006.

[59] D. E. Knuth, *Art of Computer Programming, Volume 2: Seminumerical Algorithms, The*, vol. 2. Addison-Wesley Professional, 3 ed., 1997.

[60] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[61] X. Wu and Z. Liu, "How community structure influences epidemic spread in social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 23, pp. 623 – 630, 2008.

[62] http://snap.stanford.edu/data/.

[63] D. Sornette, "Dragon-kings, black swans and the prediction of crises," *arXiv preprint arXiv:0907.4290*, 2009.

[64] R. Albert and A.-L. Barabási, "Topology of evolving networks: Local events and universality," *Phys. Rev. Lett.*, vol. 85, pp. 5234–5237, Dec 2000.

[65] Y.-B. Xie, T. Zhou, and B.-H. Wang, "Scale-free networks without growth," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 7, pp. 1683–1688, 2008.

[66] J. Ohkubo, K. Tanaka, and T. Horiguchi, "Generation of complex bipartite graphs by using a preferential rewiring process," *Phys. Rev. E*, vol. 72, p. 036120, Sep 2005.

[67] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proceedings of the 16th international conference on World Wide Web*, pp. 835–844, ACM, 2007.

[68] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pp. 29–42, ACM, 2007.

[69] D. M. Romero and J. M. Kleinberg, "The directed closure process in hybrid social-information networks, with an analysis of link formation on twitter.," in *ICWSM*, 2010.

[70] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web*, pp. 591–600, ACM, 2010.

[71] P. Cano, O. Celma, M. Koppenberger, and J. M. Buldu, "Topology of music recommendation networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 16, no. 1, p. 013107, 2006.

[72] T. Zhou, M. Medo, G. Cimini, Z.-K. Zhang, and Y.-C. Zhang, "Emergence of scale-free leadership structure in social recommender systems," *PLoS One*, vol. 6, no. 7, p. e20648, 2011.

[73] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *Nature Physics*, vol. 2, no. 2, pp. 110–115, 2006.

[74] A. Barabsi, H. Jeong, Z. Nda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mechanics and its Applications*, vol. 311, no. 34, pp. 590 – 614, 2002.

[75] S. Chandrasekhar, "Stochastic problems in physics and astronomy," *Rev. Mod. Phys.*, vol. 15, pp. 1–89, Jan 1943.

[76] R. M. Ziff, "Kinetics of polymerization," *Journal of Statistical Physics*, vol. 23, pp. 241–263, 1980.

[77] C. Seigneur, A. B. Hudischewskyj, J. H. Seinfeld, K. T. Whitby, E. R. Whitby, J. R. Brock, and H. M. Barnes, "Simulation of aerosol dynamics: A comparative review of mathematical models," *Aerosol Science and Technology*, vol. 5, no. 2, pp. 205–222, 1986.

[78] D. J. Aldous, "Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists," *Bernoulli*, vol. 5, no. 1, pp. pp. 3–48, 1999.

[79] V. M. Eguiluz and M. G. Zimmermann, "Transmission of information and herd behavior: An application to financial markets," *Phys. Rev. Lett.*, vol. 85, pp. 5659–5662, Dec 2000.

[80] R. D'Hulst and G. Rodgers, "Exact solution of a model for crowding and information transmission in financial markets," *IJTAF*, vol. 3, pp. 609–619, 2000.

[81] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," *Phys. Rev. Lett.*, vol. 85, pp. 4629–4632, Nov 2000.

[82] M. Alava and S. Dorogovtsev, "Complex networks created by aggregation," *Physical Review E*, vol. 71, no. 3, p. 036107, 2005.

[83] D. Viswanath, "Random Fibonacci sequences and the number 1.13198824...," *Mathematics of Computation*, vol. 69, no. 231, pp. pp. 1131–1155, 2000.

[84] E. Ben-Naim and P. L. Krapivsky, "Growth and structure of stochastic sequences," *Journal of Physics A: Mathematical and General*, vol. 35, no. 41, p. L557, 2002.

[85] C. Sire and P. L. Krapivsky, "Random Fibonacci sequences," *Journal of Physics A: Mathematical and General*, vol. 34, no. 42, p. 9065, 2001.

[86] I. Krasikov, G. J. Rodgers, and C. E. Tripp, "Growing random sequences," *Journal of Physics A: Mathematical and General*, vol. 37, no. 6, p. 2365, 2004.

[87] J. H. Asad, A. Sakaji, R. S. Hijjawi, and J. M. Khalifeh, "On the resistance of an infinite square network of identical resistors theoretical and experimental comparison," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 52, pp. 365–370, 2006.

[88] J. H. Asad, R. S. Hijjawi, A. Sakaj, and J. M. Khalifeh, "Remarks on perturbation of infinite networks of identical resistors," *International Journal of Theoretical Physics*, vol. 44, pp. 471–483, 2005.

[89] D. Stauffer and A. A., *Introduction to Percolation Theory.* London, Washington, DC: Taylor and Francis, 2 ed., 1991.

[90] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions.* New York: Dover, 1975.