Unterschrift des Betreuers

TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

DIPLOMARBEIT

From Random Graphs to Complex Networks:

A Modelling Approach

Ausgeführt am Institut für

Diskrete Mathematik und Geometrie
der Technischen Universität Wien

unter der Anleitung von Ao.Univ.Prof. Bernhard Gittenberger

durch

Christina Knöbel

Herbeckstr. 46/6
1180 Wien

Datum

Unterschrift

Acknowledgments

First and foremost, I would like to thank Professor Dr. Bernhard Gittenberger for being the adviser any student could hope for: He was patient, reliable, and encouraging and always took the time to answer all my questions. I also appreciated the chance to obtain invaluable teaching experience, as well as the vast number of letters of recommendation he wrote for me.

In addition, I would like to thank Dr. Matthias Dehmer for pointing me towards some papers on complex networks I might have missed, as well as Dipl.-Ing. Martins Bruveris, Dipl.-Ing. Johannes Morgenbesser and Dipl.-Ing. Andrea Trautsamwieser for their helpful and motivating comments.

Special thanks goes to Wolfgang Müller for meticulously proof-reading my diploma thesis, fine-tuning my computer, and being a friend.

Last but not least, I want to thank my parents for always believing in me.

Contents

1 Introduction

Recent years have seen an upsurge in the study of so-called "complex networks." Various vast data bases are now available for investigation that were not given only a few years ago; they are mostly so large that research would not be possible without the help of very powerful, modern computers. Some examples are e-mail records, GPS navigation systems that capture travel patterns and the World Wide Web (www). Social Networks, which have been studied for a very long time already [23, 34], are now being looked into in a different way, i.e. through citation networks [2, 4, 17, 38]. All these examples would be called complex networks, and have influenced a major part of the work presented in this diploma thesis greatly.

To generalize from these examples, each consists of a very large set of objects that have some kind of relation to each other — that is, where the relation exists, it is the same type of relation between any pair of objects. (Compare for example with definition 2.1.) Mathematically, a network is nothing else than a graph [17].

The subject of this diploma thesis is the modelling of complex networks. The road map is as follows:
In this chapter, I will first state some basic graph theoretical preliminaries, and then go on to describe what typical properties of complex networks are, i.e. what we will be looking for in the models. This will be illustrated by motivating examples in chapter 2. Then, in chapter 3, I will briefly refresh some mathematical areas and state a few theorems used, and give explanations to some methods used by physicists. Chapter 4 will deal with the first model of random graphs, the Erdős-Réyni model. It is fairly long, and is the chapter where theorems are given with rigorous proofs. Chapters 5 and 6 are more modern, dealing with models that try to put the complex networks as we understand them at the time into models. If available, an outline of the proofs will be given. Finally, chapter 7 mentions some other models that were very important in the forming of the theory of complex networks, as well as some newer models.

1.1 Some Notes on the Models

There are several reasons to model these huge networks, and it is not only out of theoretical curiosity (Are they really random? Do they form according to a system?) that researchers have been striving to understand these interwoven systems [2].

Be it a model that could predict when a small power failure could lead to a major electricity shortage — see [46] — or an accurate model of the World Wide Web, the *web graph*, helping to solve problems that are computationally difficult directly on the web (for example, testing new algorithms [32]), the applications are manifold.

1.1.1 State of the Art

Bollobás nicely summarized what type of investigations of complex networks exist in [10]. Briefly, there are

- direct investigations of real networks, where nodes, degrees and so forth are counted and various properties are examined.
- These studies are followed by new models that try to explain why the properties measured are what they are.
- Often these models are examined via computer simulations,
- and/or a heuristic analysis of their properties.
- Very rarely in comparison, a mathematically rigorous study is successfully undertaken.

1.2 Graph Theoretical Preliminaries

Following definitions are necessary for the most basic understanding of this diploma thesis, and may be skipped if the basic concepts of graph theory are known to the reader. All definitions in this section are made with help of [8].

As stated above, a network is actually a graph:

Definition 1.1 (Graph)**.** *An undirected graph $G(V, E)$ consists of an ordered pair of sets, the* vertices *(or nodes) V of a graph and the* edges *E of a graph, where $E \subset V^{(2)}$, the set of unordered pairs of V.*

Definition 1.2. *A directed graph $G(V, E)$ (also called* digraph*) also consists of an ordered pair of sets; the difference is that here $E \subset V \times V$, the set of ordered pairs of V. These elements of E are called* arcs*.*

Most results will be brought for undirected graphs, thus when we write graph we mean an undirected graph. It will only be emphasized when a graph is directed.

When referring to the vertice or edge set of a certain graph $G(V, E)$, we will speak of $V(G)$ or $E(G)$, respectively. $|G|$ is defined as $|V(G)|$.

Definition 1.3. *An edge $\{u, v\}$* joins *(or connects, or links) the vertices u and v. It will also be denoted by uv. u is said to be a* neighbor *of v. These vertices are also called* adjacent*. u and v are both* incident *with the edge uv. For an arc (u, v) in a directed graph, we say that the arc begins in u and ends at v.*

Definition 1.4 (Multigraph)**.** *A graph that contains multiple edges and edges of the form $\{v, v\}$ (so-called* loops*) is called a* multigraph*. A graph without loops containing no multiple edges is called* simple*.*

Note that there are $\binom{n}{2}$ possible edges in a simple, undirected graph. A graph where all possible edges are present is called a *complete graph*; for $|V(G)| = n$, it is denoted by K_n.

$\tilde{G}$ is said to be a *subgraph* of G if $V(\tilde{G}) \subseteq V(G)$ and $E(\tilde{G} \subseteq G)$. This is written as $\tilde{G} \subseteq G$. For $W \subseteq V(G)$, $G - W$ means deleting all vertices in W and all the edges adjacent to them. We will denote this by $[G/W]$.

Two graphs G and G' are called *isomorphic* if there exists a bijective function $\Phi : V(G) \to V(G')$ such that for every $uv \in E(G)$, $\Phi(u)\Phi(v) \in E(G')$, we write this as $G \cong \tilde{G}$

For a vertex $v \in V(G)$ we denote by $\mathrm{d}(v)$ the *degree* of this vertex: It is defined as the number of edges adjacent to v. Similarly, the *out-degree* $\mathrm{d}^{\mathrm{out}}(v)$ in a directed graph is the number of arcs that begin in v and the *in-degree* $\mathrm{d}^{\mathrm{in}}(v)$ of a vertex v is the number of arcs that end in v.

A graph P of the form $E = \{v_0v_1, v_1v_2, \dots v_{n-1}v_n\}$ is called a *path*. v_0 is called the initial vertex of P and v_n the end vertex. The number of edges in a path is called the *path length*. If $v_0 = v_n$ then P is called a *circuit*. If a circuit C does not have a vertex u such that u is used twice (i.e. $v_iu \in E(C)$ and $v_ju \in E(C)$, $i \neq j$) C is called a cycle. A graph without a cycle is called a *forest*.

The vertices of a graph that are reachable from each other via paths are said to be part of the same *connected component*. A graph where the largest connected component is the graph itself is said to be *connected*. A graph G on n vertices, n large, whose greatest connected component consists of l vertices so that $l = O(n)$ is said to have a *giant component*. A connected forest is called a *tree*. A *recursive tree* is defined to be a labelled tree that is formed via a graph process. Starting with node 1, the *root*, each new vertex j connects to an older vertex of the system at time j so that the path from any vertex to the root is always ascending.

For Digraphs, we distinguish the *strongly connected* and *weakly connected*. The latter is the case when the digraph, seen as an undirected graph, is connected, for the former there must be directed paths from any vertex to any other vertex.

The *diameter* of a graph is defined as

$$\operatorname{diam} G = \max_{\{v_i,v_j\} \in V \times V} \min_{\substack{P \subset G \\ P \text{ is a path from } v_i \text{ to } v_j}} |E(P)|.$$

Note that the diameter of an unconnected graph is usually said to equal infinity.

A graph is said to be *bipartite* if V can be partitioned into two disjoint sets V_1 and V_2 such that there exists no edge in $E(G)$ that connects two vertices in the same set.

1.3 Properties of Complex Networks

There does not seem to exist an exact definition of what makes a network complex. Vaguely speaking, a network is called complex when it has, in some sense, generated itself (see chapter 2 for examples) without a "construction plan" to guide the evolution. In most complex networks, there is still a certain dynamic in the network itself — it is still growing (or perhaps shrinking) over time, it is changing, optimizing itself again and again, vertices are born while other vertices or edges die.

Funnily, even though there are completely different types of complex networks that evolve at very different speeds — the World Wide Web is changing far quicker than protein networks that have taken centuries and more to form — there are similarities between them that give them the label of "complex network". In this section, most of these properties will be presented.

1.3.1 Big and Sparse

The quantity that is at the same time the easiest and the hardest to grasp is probably the size of these networks, size meaning the number of vertices. Most of them are *huge*. For example, in comparison, the network of protein-protein interactions of yeast proteome is comparatively small with only 1870 vertices and 2240 edges. However, collaboration networks of over a million nodes have been analyzed (see section 2.2), and those are still not the biggest networks available. The largest network being analyzed again and again, and because of its size often only in part, is the World Wide Web — it is growing so quickly that it is hard to say how many web pages there are. A recent statement by the official Google blog[1] states that the web had over a trillion (10^{12}) pages in 2008, while scientific papers only spoke of 11.5 billion in 2005. (See subsection 2.3.2. Numbers from [17].)

No matter how big the web may be today, these number are enormous. So big that they are very hard to imagine, and also so big that it is no longer possible to visualize the corresponding graph in any way that would measure all its properties [38].

Conversely, from the web to many other real world networks, these graphs are actually fairly *sparse* [11]. Sparse here means that not too many of the possible edges are present; this can be interpreted that globally — looking at the graph structure from far away — these graphs have a somewhat tree-like structure.

1.3.2 The Small-World Effect

What was coined as "six degrees of separation" is what we shall call the *small world effect*.

Roughly speaking, this means that the average distance between pairs of vertices in a graph G with n vertices will be small, i.e. if $d(v_i, v_j)$ denotes the distance between vertices v_i and v_j, then

$$\bar{l} = \frac{2}{n(n-1)} \sum_{i \geq j} d(v_i, v_j)$$

should be small, small meaning that $\bar{l}$ should scale about the same as the corresponding value of a classical random graph, $\bar{l} = O(\log n)$ [38]. (See section 4.6.)

Some authors also measure the diameter; a small diameter is of course a slightly stronger statement than a small mean distance. Both measurements need to be redefined when the graph is not connected.

1.3.3 The Clustering Coefficient

Most complex networks have a rather high number of *cliques*, that is complete subgraphs. Intuitively, this is fairly logical: Individuals in a system that have things in common will form groups.

The way to measure this is with the so-called *clustering coefficient* [38]. It is a measure for the number of triangles in a graph i.e. *cycles* of length three — especially social networks seem to have a high clustering coefficient, which seems clear: If A and B are friends, and B and C are friends, then it is likely that A and C will be friends as well. The clustering coefficient measures the transitivity of a graph.

[1] `http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html`

The clustering coefficient C can be defined in two different ways, as a local or as a global property. Globally, it measures the fraction of adjacent pairs of edges such that all three possible edges connecting them are present [10]:

$$C_g = \frac{3 \times \text{number of triangles in the network}}{\text{number of pairs of adjacent edges}}.$$

Note that there exist definitions that are not necessarily unambigious. For example, for following definition [38, 36]:

$$C_{\text{prob}} = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}}$$

needs a verbal explanation to be equivalent. Basically, C should measure the mean probability of two neighbors of a certain vertex being neighbors themselves. However, by considering the graph K_3, where the clustering coefficient should obviously be 1, we see that C_{prob} can not be right — if we count ordered triples of connected vertices, $C_{\text{prob}} = 1/2$, if we count unordered triples, $C_{\text{prob}} = 3$. A "connected triple" is defined in [36] to be a vertex which is connected to an unordered pair of other vertices. For different pairs the same vertex can then be counted several times.

Locally, the clustering coefficient is defined as

$$C_l(v) = \frac{\text{number of triangles connected to vertex } v}{\text{number of triples centered on vertex } v}$$

or, equivalently (see [10])

$$C_l(v) = \frac{\text{number of edges between neighbors of } v}{\binom{\mathrm{d}(v)}{2}}.$$

Note that for a graph of size n C_g can be obtained by C_l by [10]

$$C_g(G) = \left(\sum_{v=v_1}^{v_n} \binom{\mathrm{d}(v)}{2} C_l(v) \right) \Bigg/ \sum_{v_1}^{v_n} \binom{\mathrm{d}(v)}{2}.$$

If not otherwise stated, when talking of the clustering coefficient C we will mean C_g.

1.3.4 Scale-free Networks

The *total degree distribution* of a network of size n is defined as

$$P(k, n) = \frac{1}{n} \sum_{i=1}^{n} p(k, v_i, n)$$

where $p(k, v_i, n)$ is the probability of vertex v_i having degree k. It is thus the probability that a random vertex will have degree k. It is natural to expand this definition for directed graphs and in- and out-degree distributions [17].

For a given graph G on n vertices let $N_k(G)$ denote the number of vertices with degree k. Studies have shown that for various networks very different in type — social networks,

citation networks, the web graph — the empirical degree distribution, i.e. $\frac{1}{n}N_k(G)$ follows a *power law*. (See for example [11, 17, 38].)

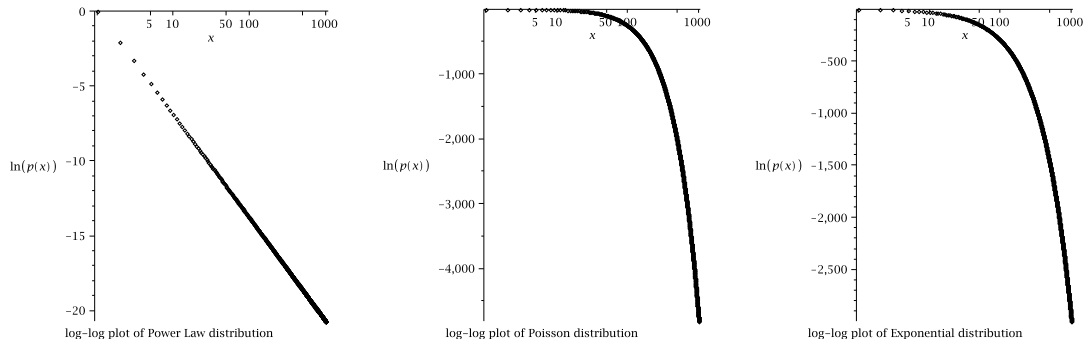
For $k \neq 0$, a power-law distribution $P(k)$ is defined as

$$P(k) \sim k^{-\gamma},$$

where $\gamma > 0$ is the exponent of the distribution. There is no scale in the network (i.e. it does not depend on the size n of the network) which is why this distribution is called *scale free*.

Notice that this distribution is also *fat tailed*. This means that even for very large k, there is still a fair possibility of there being a vertex of degree k. This distribution does not tend to zero as fast as a Poisson distribution, for example, or an exponential distribution. (See figure 1.1.) Also note that if the average degree of a network is finite, it must hold $\gamma > 2$.

It is important to point out that, seeing that real networks are always finite in number, this distribution will have a natural cut-off for an analyzed "real life" network.



(a) Power Law distribution (b) Poisson distribution (c) Exponential distribution

Figure 1.1: Log-Log plots for several distributions for values $1, \ldots, 1000$.

2 Touching upon Real Networks

2.1 Going from Social Networks ...

Applying graph theory to the social sciences is not a new concept. Networks of friendships have been investigated since at least the 1920's [23]. Understanding the structure of human interactions is interesting not only for its own sake, but also the implications of such theories are important: How information spreads, as well as diseases, can be modelled with network theory [37]. While there are several different approaches to social network analysis (for example, see [46, p. 47ff.]), for our purposes, this definition will be sufficient:

Definition 2.1. *A Social Network is a set of people or groups of people with some pattern of contacts or interactions between them [38, p. 174].*

Among the most fascinating studies concerning social networks are probably the so-called "small-world" experiments conducted by Milgram in 1967. Milgram asked a few hundred random individuals in Omaha, Nebraska, to convey a letter to some target person in Boston, Massachusetts. They were only to pass these letters on to other people they knew on a first name basis. About a quarter of these letters arrived; on average, each letter only passed through six people before arriving at its goal.

Albeit not aiming to prove anything about networks, this experiment later coined the phrase "six degrees of separation," which is not only the title of a play [25] but also the basic idea behind the small world effect as discussed in subsection 1.3.2.

The classical approach to social networks has some flaws that render it difficult to work with from the point of view of the sciences. First of all, due to the nature of human interactions, it is often necessary to obtain information by directly interviewing people via questionnaires and the like. This is time and cost intensive, and leads to comparatively small sample sizes. The information obtained is often inaccurate, because respondants interpret questions differently. For example, there is no universal definition of the word *friend* [38, p. 175].

Other methods of analyzing social networks needed to be found; the vast databases now available facilitated research. Investigations of so-called *collaboration networks* started. A link between two individuals exists here exactly when a trace of their collaboration is in the database.

For example, the network of movie actors (nearly half a million people) has been studied with the help of the *Internet Movie Database*[1] [39]. This network can be studied in two ways:

1. As a bipartite graph $G(A \cup M, E)$, where there are edges connecting the nodes A (actors) and the nodes M (movies). There is an edge between $a_i \in A$ and $m_j \in M$ iff a_i has starred in m_j

[1] `http://www.imdb.com/`

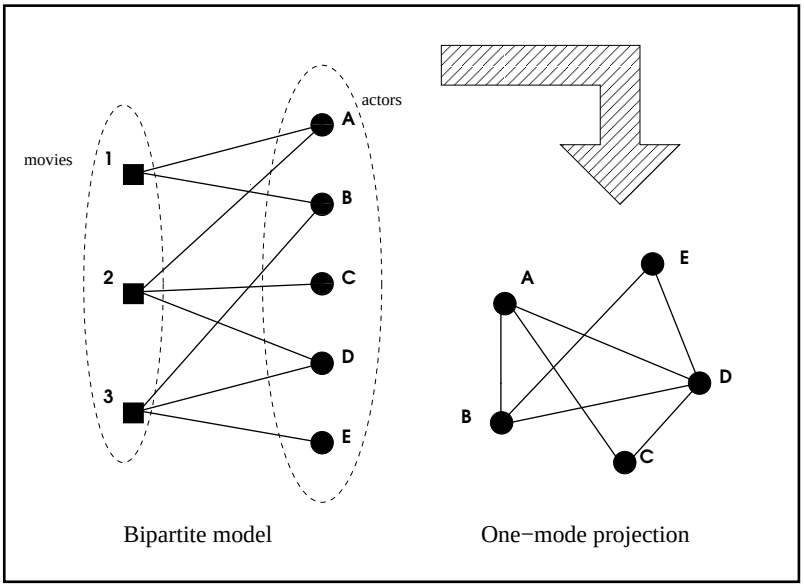


Figure 2.1: One mode projection

2. As the so-called *one-mode projection.* where all vertices represent actors, and there is a link between two actors if they have played in a movie together. (See Fig. 2.1.)

Note that in the one mode projection, some information is lost. Also, as most films are made with more than two actors, the projection of the vertices obviously causes many triangles, which will influence the clustering coefficient. It is also necessary to point out that even though two actors appeared in the same film, this does not imply that they had any other social contact apart from that; as such, it is questionable if the net of movie actors is representative of human interactions.

2.2 . . . over Scientific Collaboration Networks . . .

The idea behind scientific collaborations is similar. One of the first papers treating this topic from a graph theoretical view was written by Newman who investigated several concepts of complex networks in four vast databases [37].

The concept of investigating scientific collaborations is not completely new, either. In the mathematics community, a very central figure was *Paul Erdős* (one of the two founders of the theory we will use in chapter 4). He wrote over 1,400 papers during his lifetime, which outnumbers all other mathematicians. In fact, Paul Erdős is so central that mathematicians started calculating their *Erdős numbers*, which is the distance (via co-authorship of papers) they have to this exceptional scientist. Having co-authored a paper with Paul Erdős gives Erdős number one, a co-author of someone who has Erdős number one has Erdős number two, and so on. Erdős numbers are so popular that there is an application on the site of the *American Mathematical Society*[2] to calculate the distance between any two mathematicians, with Erdős being the default value.

[2] `http://www.ams.org/mathscinet/collaborationDistance.html`

	MEDLINE	Los Alamos	SPIRES	NCSTRL
Total papers	2,163,923	98,502	66,652	13,169
Total authors	1,520,251	52,909	56,627	11,994
Mean papers/author	6.4	5.1	11.6	2.55
Mean authors/paper	3.754	2.530	8.96	2.22
Collaborators/author	18.1	9.7	173	3.59
Cutoff z_c	5,800	52.9	1,200	10.7
Exponent τ	2.5	1.3	1.03	1.3
Size of giant component	1,395,693	44,337	49,002	6,396
As a percentage	92.6%	85.4%	88.7%	57.2%
Second largest component	49	18	69	42
Mean distance	4.6	5.9	4.0	9.7
Maximum distance	24	20	19	31
Clustering coefficient C	0.066	0.43	0.726	0.496

Table 2.1: Findings from four different networks [37].

As seems natural, in [37] two scientists are linked if they published a paper together. Newman's data came from MEDLINE (papers on biomedical research), the Los Alamos e-Print Archive (preprints in theoretical physics), SPIRES (papers and preprints in high energy physics) and NCSTRL (preprints in computer science), where he investigated the papers issued during the five year period 1995–1999.

Newman checked all the essential questions of complex networks. Some of his findings (more sub-cases were investigated) are summarized in table 2.1. On average, authors wrote four papers between 1995 and 1999; each paper has an average of three authors, which slightly biases the clustering coefficient C. Note that the average number of collaborators of an author varies strongly between the different papers. For NCSTRL (computer science), it is below four, while for MEDLINE the number already rises to 18, reaching 173 for SPIRES. This can be explained by the nature of the sciences involved: papers published in SPIRES are by high-energy experimentalists, where many people are involved per paper just to run the experiments.

Newman also tried to fit the number of collaborators z with a power-law form, which did not work. However, he succeeded quite well in fitting the data by a power-law form with an exponential cutoff:

$$P(z) \sim z^{-\tau} e^{-z/z_c}$$

where τ and z_c are constants. One explanation for the cutoff is the finiteness of the data — only five years of numbers. The values of τ and z_c are given in table 2.1, they vary considerably.

Note that for all these sets of researchers, the giant component of their collaboration graphs comprises more than half of all vertices, in three of them around 90%. This, and the fact that the second largest components are all truly tiny in comparison, show how strongly connected the networks of these scientists are.

They also exhibit the "small-world" property: the average distance between a random pair of vertices is around six. In fact, splitting the data into the different (sub-)groups of scientists, we can see in Fig. 2.2 that this data can be fitted to match the definition of the "small-world" property, as the average distance of two random researchers is plotted

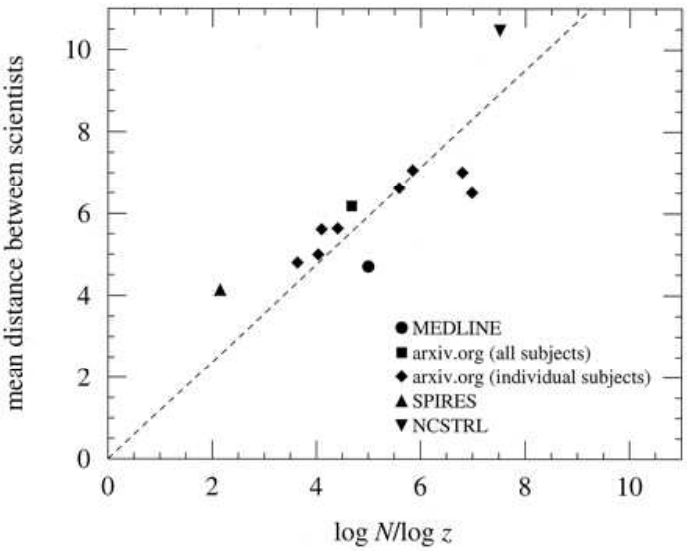


Figure 2.2: Average distance of researchers to average distance in random graph [37]

against the average distance in a random graph with the same number of vertices and edges.

These networks exhibit very high clustering coefficients. One explanation is (as with movie actors) that a paper written by three or more authors already induces at least one triangle. However, the values here are so high that this cannot be the only explanation. Interestingly, MEDLINE's clustering coefficient is much lower in comparison. This may be because of the hierarchical structure of biological laboratories, which could cause tree-like networks without many loops.

Other studies of collaboration networks followed, which have treated new models, for example, see [5]. To summarize Newman's findings, overall, they fulfilled our expectations which we described in section 1.3. We shall see that collaboration networks are not the only networks that "perform" in this way.

2.3 . . . to the Internet and www . . .

Two of the driving factors in network theory are the internet and the World Wide Web (www). These massive networks are intriguing in their size, fast dynamics and almost complete self-organization. Apart from purely scientific reasons, this research is important in applications: web crawls and search engines, network stability questions and understanding the sociology of content creation are only a small number of possible examples [12].

2.3.1 The Internet

First note the difference between the internet and the www [17, p. 34 ff]. The internet is made up of physical components on different levels, such as computers that have activated their connection to the net (hosts), servers that provide service to the web, and routers that arrange traffic across the internet. Edges in this network are the connections between the different components, they are undirected.

In the beginning of 2001, it contained about 100 million (10^8) hosts. Going up one level, we study the internet at the router or interdomain level. Mid 2000 there existed about 150 000 routers in total, a year later there were about 220 000.

Dorogovtsev and Mendes [17] compare several studies of the internet at interdomain level done between 1999 and 2002. At the time, this graph was quite small and sparse; in 1999, it had only $n = 5287$ vertices, their number and the number of edges connecting them fluctuates considerably. From 1997 to 1999, the average degree increased from 3.42 to 3.8, meaning that connections grow stronger than vertices. The average distance $\bar{l}$ between two nodes was always below four, and the ratio of $\bar{l}$ to the equivalent number of the corresponding random graph was 0.6, meaning that the internet shows signs of having the "small-world" effect. The maximum length of separation was around 11, and the clustering coefficient C of 0.2 was considerably higher than that of a classical random graph. The degree distribution manifested a power-law with exponent 2.2.

Note that, because we are talking about physical components here — hardware, cables, and so on — geographical as well as economical influences matter: For example, the fluctuations in n are due to providers opening or going out of business, while locations of routers etc. have been shown to be closely related to population density.

2.3.2 The World Wide Web

In comparison, the www consists of documents (pages) containing information [11, 17]. When these pages refer to each other, they are connected by hyperlinks. The webpages are the vertices of the web-graph, while the connections are the arcs.

Note that this gives a directed graph. For each page, we are thus looking at links coming in and going out, so we also distinguish in- and out-degrees.

The web is growing quickly [11]. In 1997, there were supposedly 320 million web pages. In 1999, around 800 million were found. A current and accurate estimate is hard to come by, but in 2005 it was stated that the web had about 11.5 billion pages, while shortly after this number was claimed to be 53.7 billion, with 34.7 billion of these pages indexed by *Google*[3].

The directedness of the web changes our definition of the giant component and the way we see the structure of the net. We say the www has a *bow-tie structure*, and we call the giant strongly connected component (GSCC) the *core* or *knot* of the bow. We call vertices connecting to the core part of the *giant in component* (GIN) and vertices that have edges connecting from the GSCC part of the *giant out component* (GOUT). Note that the GSCC is the intersection of GIN and GOUT. The remaining vertices are either not connected to the giant component, or they are so-called *tendrils* that connect to the GOUT, or lead away from the GIN. For a schematic view of the (weakly) connected component of the web, see figure 2.3. The connectivity of the web is becoming stronger, as the apparent growth of the core shows: While in 2000, it was estimated that a third of all pages were in the core, a 2006 estimate placed two thirds of all web pages in the GSCC [11].

The first investigation of the web of this size was conducted by Broder, [12], where 200 million pages and 1.5 billion links were examined via the altavista web crawl in 1999. Though the www has grown considerably in the mean time, the first findings remain of interest: Especially the repeated occurrence of power-law distributions is fascinating, as

[3] `www.google.com`

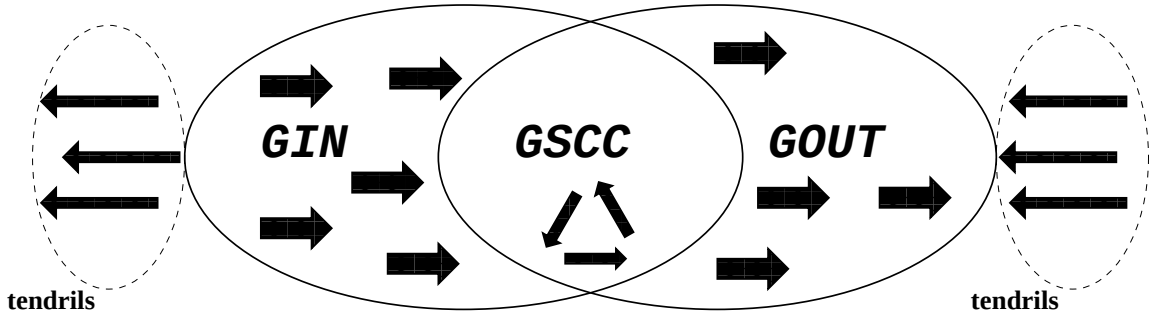


Figure 2.3: GSCC

well as the results concerning the giant component.

In fact, it was found that the in-degree of the www has a power-law with exponent around two, i.e. $p_k^{\text{in}} \sim \frac{1}{k^{2.1}}$, a number already reported in earlier, smaller searches. The out-degree distribution was also a power-law with exponent 2.72. Interestingly, the distribution of the sizes of the weakly connected components also exhibited a power law, with exponent 2.5.

Treating the www as an undirected graph, 91% (186 million) of all nodes were in the giant (weakly connected) component. This component was shown to be surprisingly robust and well connected: if all edges to pages of degree greater than 5 had been removed, the graph would have still contained a weakly connected component of 59 million vertices. The average undirected distance was shown to be 6.83.

The distribution of the sizes of strongly connected components also exhibits a power law. Broder's study [12] also found the (directed) diameter of the www to be at least 28, counting only pairs of vertices for which there actually *exists* a directed path connecting them.

2.4 . . . and Beyond

A myriad of applications can be described by complex networks. In this last section, I will name a few more, and briefly describe some.

One fairly old idea that has been revived lately is the investigation of citation networks, an overview given in [17]. Papers citing other papers form a so-called *citation graph*. In this model, papers are the nodes and citations are the arcs. A new paper links to older papers, and older papers do not change, so they cannot form any new links. When preprints are disregarded, this directed network is acyclic, even though the underlying undirected graph may have cycles. Also note that most empirical surveys of citation graphs measure the current state of the graph, and not how it evolves over time.

In fact, one of the first studies to report a power-law degree distribution $P(k)$ was about citation networks by Redner [42]. He investigated data from the *Institute for Scientific Information* (ISI database) including 783,339 papers and 6,717,198 citations, and data from *Physical Review D* (PRD) of 24,296 papers with 351,872 citations.

Redner tried to fit the data with a stretched exponential,

$$P(k) \sim \exp\left(\left(-\frac{k}{x_0}\right)^{\beta}\right),$$

but this did not explain for the (few, widely scattered) very big values of degree k. For ISI only 64 out of over 700,000 paper are cited more than 1000 times, 282 are cited more than 500 times, while over 300,000 papers are uncited. Redner looked for a function that was less smooth, and showed that the data fitted well to $P(k) \sim k^{-\alpha}$, with α close to 3.

It was later proposed to fit the data with $(k + \text{const})^{-\alpha}$, where $\alpha = 2.9$ for the ISI net, and $\alpha = 2.6$ for the PRD data, and even later it was stated that the data was still not large enough to be sure that the distribution is really fat-tailed. Different possibilities to fit the data exist, although there is evidence for preferential attachment in the citing process. This would imply a scale-free distribution.

Another science where graph theory has been implied to model "real-world" networks is biology: From directed food webs (arcs indicating who eats whom) to neural networks (There are 100 billion neurons in the human brain, the largest network mentioned in [17].), from metabolic reactions to protein networks (such as *protein-protein-interactions* (PPIs)). Some criteria of complex networks (small world effect, etc.) are always present.

Other graphs the theory of complex networks is analyzing are the *Word Web of human language*, various communication networks (*mail networks*, the *telephone call graph*, ...) as well as *power grid networks*, *energy landscape networks* and many, many more. For a very vivid description of why the latter two are important, see the opening chapter of [46].

3 Methods

In this chapter, a few of the methods, definitions and notations used throughout this diploma thesis are presented. It is divided into three sections:

- Methods used by physicists that a mathematician might not have seen on the core course curriculum,
- a brief refreshment of methods used in probability theory, as well as some theorems that might not be familiar, and
- notations and abbreviations used.

3.1 Methods from Statistical Physics

3.1.1 The Mean Field Method and other Continuum Approaches

The mean field method seems to stem from statistical physics; statistical mechanics, to be exact. Finding an exact definition of the mean field method in general (as opposed to some application of mean field theory to some particular example) has proved difficult.

The basic idea behind is to view a discrete process as a continuous one by considering the same model several times, and then taking the mean of the outcome. This comes from the fact that, when particles are being considered, the mean over time is more relevant than the exact number of particles passing through the investigated area. The same thing is possible when quantities of considerable size are being treated. Quoting from [17]

> ..., master equations become very simple. At first sight, this must work for large degrees, but mathematicians know that such limiting is an extremely dangerous operation. Sometimes it works, sometimes not, and while using the continuum approximation, you have to check your work all the time. However, for simple growing networks this approximation usually yields exact results for most useful quantities or produces unimportant deviations.

This is the reason why, as noted in subsection 1.1.1, there are many heuristic results and rather few rigorous ones — they are easier to find. Most (but not all) have been proved true when rigorous treatment was possible.

3.1.2 The Master Equation

When dealing with Markov chains (see subsection 3.2.3), an important aspect is the *master equation*. It gives the rate of change of the probability $P(x,t)$ due to transitions into the

state x from all other states and due to transitions out of a state x into all other states [43], i.e. for a system with n possible states

$$\frac{\partial P(x,t)}{\partial t} = \sum_{i=1}^{n} \left((P(i,t)p_{i,x}(t) - P(x,t)p_{x,i}(t)) \right),$$

where $p_{i,j}(t)\Delta t$ is the probability of a transition from state i to state j during the time change $t \to t + \Delta t$.

3.2 Methods from Probability Theory

3.2.1 Some Distributions

- We say a random variable X has Poisson distribution, $X \sim \mathrm{Po}(\lambda)$ if

$$\mathbb{P}(X = i) = \frac{\lambda^i}{i!} e^{-\lambda} \qquad \text{for} \quad i = 0, 1, \ldots; \quad \lambda > 0.$$

- We say a random variable X has exponential distribution, $X \sim \mathrm{E}(x, \lambda)$ if for its density function $f(x)$ it holds

$$f(x) = \lambda e^{-\lambda x} \qquad \text{for} \quad x > 0.$$

  For its distribution $F(x)$, it holds $F(x) = 1 - e^{-\lambda x}$.

- For a random variable X with Bernoulli distribution with mean p, we will write $X \sim \mathrm{Be}(p)$.

- For a random variable X with binomial distribution with parameters n and p, we will write $X \sim \mathrm{Bi}(n, p)$. For $\binom{n}{k} p^k (1-p)^{n-k}$ we will use the notation $\mathrm{Bi}(k; n, p)$.

3.2.2 Two familiar Inequalities

Let us just state these inequalities (both from [28]) again as a reminder to the reader:

Theorem 3.1 (Chebyshev's inequality)**.** *For a random variable X, if $\mathrm{Var}(X)$ exists it holds*

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq t\right) \leq \frac{\mathrm{Var}(X)}{t^2}, \qquad t > 0.$$

Theorem 3.2 (Markov's Inequality)**.** *For a random variable $X \geq 0$ almost surely, it holds*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}, \qquad t > 0.$$

For a series of random variables X_n we say X_n *converges in distribution* to Z as $n \to \infty$ written as $X_n \xrightarrow{d} Z$ if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(Z \leq x)$ for every real x that is a continuity point of $\mathbb{P}(Z \leq x)$. Equivalently, for integer-valued random variables, $\mathbb{P}(X_n = k) \to \mathbb{P}(Z = k)$ [28].

3.2.3 Markov Chains

For a sequence of random variables $(x_n)_{n\in\mathbb{N}}$, let $\{x_n\}$ denote the filtration of the sequence up to time n, i.e. the event that $(x_0 \in X_0, x_1 \in X_1, \ldots, x_n \in X_n)$ for some events $X_0, X_1, \ldots, X_n$.

Definition 3.1 (Markov Process [29])**.** *Let $(\Omega, \mathcal{B}, \mu)$ be a filtration space with a denumerable stochastic process $(x_n)_{n\in\mathbb{N}}$ defined from Ω to a denumerable state space S of more than one element. The process is called a* denumerable Markov process *if, for any n,*

$$\mathbb{P}\left(x_{n+1} \in c_{n+1} \mid x_0 \in c_0 \cap \cdots \cap x_{n-1} \in c_{n-1} \cap x_n \in c_n\right) = \mathbb{P}\left(x_{n+1} \in c_{n+1} \mid x_n \in c_n\right)$$

for any states $c_0, \ldots, c_{n+1}$ such that $\mathbb{P}\left(x_0 \in c_0 \cap \cdots \cap x_{n-1} \in c_{n-1} \cap x_n \in c_n\right) > 0$.

For a finite, time homogenous Markov process with n possible states, $[n] := \{1, 2, \ldots n\}$, we call the matrix $\mathcal{P} \in \mathbb{R}^{n\times n}$ the *transition matrix* where p_{ij} is the probability of passing from state i to state j.

Note that in $\mathcal{P}^m$, the element $p_{ij}^{(m)}$ gives the probability of passing from step i to step j in m steps. A state i in a Markov chain is called *ergodic* if, for every state j it is possible to go from i to j, possibly via several steps. This is equivalent to saying that there exists an m so that $p_{ij}^{(m)} > 0$ for every j. A Markov chain that consists of ergodic states is called *irreducible*.

If there exists a subset of states $A \subset [n]$ such that $p_{ij} = 0$ for every pair (i, j) such that $i \in A$, $j \in [n]/A$, this subset is called an *absorbing* subset. Once an absorbing subset is reached, it is no longer possible to leave this subset. These subsets are also called essential classes. If a Markov Chain has an absorbing subset of size 1, it is called *reducible*.

If there are absorbing subsets of a finite Markov process that also has ergodic states, it is always possible to write the transition matrix $\mathcal{P}$ in this form:

$$\mathcal{P} = \left(\begin{array}{c:c} \mathbf{R} & \mathbf{0} \\ \hdashline \mathbf{S}' & \mathbf{Q} \end{array}\right), \tag{3.1}$$

where $\mathbf{R}$ is the square matrice associated with the absorbing states of the process, while $\mathbf{Q}$ is the square matrice associated with the ergodic states. It can be shown that, with probability tending to 1 a process that is not irreducible will end in an absorbing state.

We consider a Markov chain consisting of both ergodic and absorbing states. We denote the set of ergodic states by I. Let us denote the random variable $\mathcal{Z}_{ij}$ as the number of visits to state $j \in I$ starting from $i \in I$, $(\mathcal{Z}_{ii} \geq 1)$. Then

$$\mathcal{Z}_i = \sum_{j\in I} \mathcal{Z}_{ij}, \qquad i \in I$$

is the time to absortion of the chain starting from $i \in I$. We denote $\mathbb{E}\mathcal{Z}_{ij}$ with m_{ij} and $\mathbb{E}\mathcal{Z}_i$ with m_i. Let $\mathbf{M}$ be the matrix whose entries consist of m_{ij} and $\mathbf{m}$ the vector of the m_is. Then following theorem holds:

Theorem 3.3 ([44]). *Under the assumptions stated above,*

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}, \tag{3.2}$$

$$\mathbf{m} = \mathbf{M}\mathbf{e} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{e}, \tag{3.3}$$

where $\mathbf{Q}$ *is the submatrix of the transition matrix* $\mathcal{P}$ *as defined in (3.1),* $\mathbf{I}$ *is the identity matrix, and* $\mathbf{e}$ *is a vector consisting of ones.*

3.2.4 Martingales

Definition 3.2 (Martingale [28]). *For a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and an increasing sequence of sub-*σ*-fields* $\mathcal{F}_0 = \{\varnothing, \Omega\} \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = \mathcal{F}$*, a sequence of random variables* $X_0, X_1, \ldots, X_n$ *(with finite expectations) is called a* martingale *if for each* $k = 0, \ldots, n-1$*,* $\mathbb{E}(X_{k+1} \mid \mathcal{F}_k) = X_k$*.*

Often, Ω is a finite space and $\mathcal{F}$ the family of all subsets; $\mathcal{F}_k$ corresponds to a partition $\mathcal{P}_k$ of Ω, with finer partitions for larger k. Note that it is also possible to consider sequences of sub-σ-fields of the form $(\mathcal{F}_n)_{n\in\mathbb{N}}$ with corresponding sequences of random variables of the form $(X_n)_{n\in\mathbb{N}}$.

Theorem 3.4 (Azuma-Hoeffding inequality [28]). *If* $(X_k)_0^n$ *is a martingale with* $X_n = X$ *and* $X_0 = \mathbb{E}X$*, and there exist constants* $c_k > 0$ *such that*

$$|X_k - X_{k-1}| \leq c_k$$

for each $k \leq n$*, then, for every* $t > 0$*,*

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right),$$

$$\mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right).$$

3.3 Notations and Abbreviations

Throughout this diploma thesis, the following standard notation is used to describe asymptotics, see i.e. [28]. For a sequence of two numbers, $a_n \in \mathbb{R}$ and $b_n > 0$, depending on a parameter $n \to \infty$, we write

- $a_n = O(b_n)$ as $n \to \infty$ if there are constants C and n_0 such that $|a_n| \leq Cb_n$ for $n \geq n_0$, i.e., the sequence a_n/b_n is bounded for $n \geq n_0$.
- $a_n = \Omega(b_n)$ as $n \to \infty$ if there exist constants $c > 0$ and n_0 such that $a_n \geq cb_n$ for $n \geq n_0$.
- $a_n = \Theta(b_n)$ as $n \to \infty$ if there exit constants $C, c \geq 0$ and n_0 such that $cb_n \leq a_n \leq Cb_n$, i.e. a_n and b_n are of the *same order of magnitude.* We may also use the notation $a_n \propto b_n$.
- $a_n \sim b_n$ if $a_n/b_n \to 1$.

- $a_n = o(b_n)$ as $n \to \infty$ if $a_n/b_n \to 0$.
- $a_n \ll b_n$ or $b_n \gg a_n$ if $a_n \geq 0$ and $a_n = o(b_n)$.

Both the abbreviations a.a.s., meaning *asymptotically almost surely*, as well as whp, meaning *with high probability* are used to describe that a property $\mathcal{E}_n$ of a random structure that depends on n holds with $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$. U.a.r. stands for uniformly at random.

For two events $\mathcal{E}_1$ and $\mathcal{E}_2$, we say that $\mathcal{E}_1 \subset \mathcal{E}_2$ if $\mathcal{E}_1 \Rightarrow \mathcal{E}_2$. We use the notation $(n)_k := \frac{n!}{(n-k)!}$, as well as the notation $[n] := \{1, 2, \ldots, n\}$.

4 The "classical" Random Graph Model by Erdős and Rényi

4.1 The Model

The initial model of random graphs goes back to Erdős and Rényi, see e.g. [22]. There are two essentially equivalent approaches to see random graphs [28], denoted by $G_{n,N}$ and $G_{n,p}$. In both models, we have n labeled vertices, $v_1, v_2, \ldots, v_n$. Both models are random; they differ in the way the edges are chosen.

In $G_{n,N}$, N edges are chosen at random from the $\binom{n}{2}$ possible edges, where each edge is chosen with equal probability. Out of the $\binom{\binom{n}{2}}{N} =: C_{n,N}$ possible resulting graphs, each one appears with equal probability, i.e. each resulting Graph $G_{n,N}$ appears with probability $\frac{1}{C_{n,N}}$. This model is called the *uniform random graph* model.

Contrarely, in the model $G_{n,p}$ each of the $\binom{n}{2}$ possible edges appears with probability $0 \leq p \leq 1$. Thus, for $p = 0$ we will have a graph on n vertices containing no edges, i.e. $G_{n,0}$, while for $p = 1$ we will have $G_{n,\binom{n}{2}}$, the complete graph K_n. A graph thus obtained is also called *binomial random graph.*

To be correct, we should denote by $\mathbb{G}_{n,N}$ and $\mathbb{G}_{n,p}$ the set of all graphs obtained by the uniform and binomial random graph models, respectively. However, by abuse of notation, when it is clear with which we are dealing, will refer to both the set of all random graphs as well as a certain graph picked from the set with $G_{n,N}$ and $G_{n,p}$.

Definition 4.1. *For a given graph G we define the corresponding random graph as a Erdős-Rényi random graph $G_{n,N}$ with $n = |G|$ and $N = |E(G)|$.*

Another approach concerning $G_{n,N}$ is the following: Out of the $\binom{n}{2}$ possible edges, we choose at first only one. Then, out of the remaining $\binom{n}{2} - 1$ edges, we chose another one at random, and so forth until we chose our N^{th} edge out of $\binom{n}{2} - N + 1$. In this model, the interest lies in increasing N. This can be seen as a random graph process, which we will denote by $(\mathbb{G}(n, N))_N$. Letting $N = N(n)$ and $n \to \infty$, investigations of "typical behavior" are made, depending on $N(n)$. This can mean that if $\lim_{n\to\infty} \mathbb{P}_{n,N}\{G_{n,N} \text{ has property } \mathcal{Q}\} = 1$, "almost all" Graphs in $G_{n,N}$ have this characteristic $\mathcal{Q}$.

Several properties of random graphs have been thouroughly examined, questions such as: How large must N be so that almost all graphs in $G_{n,N}$ have a cycle of order k? How many graphs do not have any tree of order l? If we let $A_{n,N}$ denote the number of graphs of $G_{n,N}$ having a certain property $\mathcal{Q}$, obviously $\mathbb{P}_{n,N}(A) = \frac{A_{n,N}}{C_{n,N}}$ is the probability that any chosen graph of $G_{n,N}$ will have $\mathcal{Q}$.

Definition 4.2. *We define a property $\mathcal{Q}$ on graphs $G = G(V, E)$, $|V| = n$, formally in the following way: $\mathcal{Q} \subseteq 2^{\binom{n}{2}}$, where $2^{\binom{n}{2}} := \{G : G \subseteq K_n \wedge |V(G)| = n\}$*

Intuitively, $G_{n,p}$ and $G_{n,N}$ are equivalent; $G_{n,N}$ can be seen as a random graph $\{G_{n,p} : |E| = N\}$. Asymptotically, this is in fact so, as the following theorems show [28]. In the rest of this chapter, we will use either $G_{n,p}$ or $G_{n,N}$, depending on which is more convenient. Note that, even though the notation is the same, it will always be clear from context which model is being treated.

Proposition 4.1. *Let $\mathcal{Q}$ be a random property of subgraphs of the complete graph K_n a graph G may or may not have, $p = p(n) \in [0,1]$ and $0 \leq a \leq 1$. If for every sequence $N = N(n)$ such that $N = \binom{n}{2} p + O\left(\sqrt{\binom{n}{2} pq}\right)$, where $q = 1-p$, it holds that $\mathbb{P}(G_{n,N} \in \mathcal{Q}) \to a$ as $n \to \infty$, then also $\mathbb{P}(G_{n,p} \in \mathcal{Q}) \to a$ as $n \to \infty$.*

For a proof of this theorem, see [28].

For the other direction no equivalence can be found in such generality. A counterexample is the property of a graph containing exactly N edges. However, with the following definition and lemma, a result can be obtained.

Definition 4.3. *A family of subgraphs $\mathcal{Q} \subseteq 2^{K_n}$ is called increasing if $A \subseteq B$ and $A \in \mathcal{Q}$ imply that $B \in \mathcal{Q}$. Vice-versa, a family of subgraphs is called decreasing if its complement is increasing. An increasing or decreasing family is called monotone.*

Lemma 4.1. *Let $\mathcal{Q}$ be an increasing property of subgraphs of K_n, $0 \leq p_1 \leq p_2 \leq 1$ and $0 \leq N_1 \leq N_2 \leq \binom{n}{2}$. Then*

$$\mathbb{P}(G_{n,p_1} \in \mathcal{Q}) \leq \mathbb{P}(G_{n,p_2} \in \mathcal{Q})$$

and

$$\mathbb{P}(G_{n,N_1} \in \mathcal{Q}) \leq \mathbb{P}(G_{n,N_2} \in \mathcal{Q}).$$

Proof. For this proof we first apply the so-called two-round exposure technique, which applies to the binomial model, viewing the random graph process $G_{N,p}$ as the union of two independent random graph processes G_{n,p_1} and G_{n,p_2}, where the edges of each model are taken, and double edges are replaced by one edge, for $p = p_1 + p_2 - p_1 p_2$. We set $p_0 = (p_2 - p_1)/(1 - p_1)$. Now, G_{n,p_2} can be viewed as a union of two independent random graph processes, G_{n,p_0} and G_{n,p_1}. Thus, $G_{n,p_1} \subseteq G_{n,p_2}$, and with $\mathcal{Q}$ increasing, the first inequality follows, as the event of $G_{n,p_1} \in \mathcal{Q}$ implies $G_{n,p_2} \in \mathcal{Q}$.

To prove this lemma for the uniform model, it suffices to construct a random graph process $\{\mathbb{G}(n,N)\}_N$. $G_{n,N}$ is now the N^{th} process in order, and obviously $G_{n,N_1} \subseteq G_{n,N_2}$. With the same arguments as in the first part of the proof, the second inequality is shown. □

Proposition 4.2. *Let $\mathcal{Q}$ be a monotone property of subgraphs of $\mathcal{G}_n$, $0 \leq N \leq \binom{n}{2}$, and $0 \leq a \leq 1$. If for every sequence $p = p(n) \in [0,1]$ such that*

$$p = \frac{N}{\binom{n}{2}} + O\left(\sqrt{\frac{N\left(\binom{n}{2} - N\right)}{\binom{n}{2}^3}}\right)$$

it holds that $\mathbb{P}(G_{n,p}) \to a$, then $\mathbb{P}(G_{n,N}) \to a$.

For a proof of this proposition, see [28].

4.2 Threshold Functions

For many graph properties $\mathcal{E}$ there exist so-called *threshold functions* [22, 28] which we shall denote by $A(n)$, $A(n) \to \infty$ for $n \to \infty$ such that

$$\lim_{n\to\infty} \mathbb{P}_{n,N(n)}(\mathcal{E}) = \begin{cases} 0 & \text{if } \lim\limits_{n\to\infty} \frac{N(n)}{A(n)} = 0 \\ 1 & \text{if } \lim\limits_{n\to\infty} \frac{N(n)}{A(n)} = +\infty \end{cases}$$

It can be shown (see [28, page 20]) that every monotone property has a threshold.

The next theorem is a good example of threshold functions; several properties are special cases of following general case. Out of historical interest, the proof is from the original paper by Erdős and Rényi [22]. The same result can be shown much quicker using modern probabilistic tools [28].

Definition 4.4. *A graph is said to be* balanced *if it has no subgraph of strictly larger average degree. This means a graph $G(V, E)$ with $|V| = k$ and $|E| = l$ is balanced if for every subgraph with k' vertices and l' edges it holds that $l' \leq k'l/k$.*

Theorem 4.1. *For k, $l \in \mathbb{N}$, $k \geq 2$, $k - 1 \leq l \leq \binom{k}{2}$ let*

$$\mathcal{B}_{k,l} = \left\{ B^1_{k,l}, \ldots, B^R_{k,l} : \forall i \neq j : B^i_{k,l} \ncong B^j_{k,l} \right\}, \qquad 1 \leq R \leq \binom{\binom{k}{2}}{l},$$

be a set of balanced graphs consisting each of k vertices and l edges. Then the threshold function for the property that any random Graph G from $G_{n,N}$ should contain at least one subgraph isomorphic to some element of $\mathcal{B}_{k,l}$ is $n^{2-k/l}$.

Proof. Let $\mathbb{P}_{n,N}(\mathcal{B}_{k,l})$ denote the probability that a random graph of $G_{n,N}$ contains at least one subgraph isomorphic to one element of the class $\mathcal{B}_{k,l}$. There are $\binom{n}{k}$ possibilities of selecting k vertices from which we form a graph in $\mathcal{B}_{k,l}$ (which can be done in R possible ways). The remaining $N - l$ edges can be selected from the $\binom{n}{2} - l$ other possible edges. (Note that we are counting some graphs more than once.) Thus,

$$\mathbb{P}_{n,N}(\mathcal{B}_{k,l}) \leq \binom{n}{k} R \frac{\binom{\binom{n}{2}-l}{N-l}}{\binom{\binom{n}{2}}{N}} = O\left(\frac{N^l}{n^{2l-k}}\right).$$

Assuming that $N = o(n^{2-k/l})$, then $\mathbb{P}_{n,N}(\mathcal{B}_{k,l}) = o(1)$, so there are no subgraphs isomorphic to a subgraph in $\mathcal{B}_{k,l}$ almost sure.

The other direction is a bit longer:

Let $\mathcal{B}^{(n)}_{k,l} := \{S \subseteq K_n : \exists B^i_{k,l} \in \mathcal{B}_{k,l} : S \cong B^i_{k,l}\}$. For any $S \in \mathcal{B}^{(n)}_{k,l}$ we define

$$\mathbf{1}(S) := \begin{cases} 0 & \text{if } S \subseteq G_{n,N} \\ 1 & \text{if } S \nsubseteq G_{n,N} \end{cases}$$

Then it follows

$$\left(\sum_{S \in \mathcal{B}^{(n)}_{k,l}} \mathbf{1}(S) \right) = \binom{n}{k} R$$

$$\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)=\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbb{E}\mathbf{1}(S)=\binom{n}{k}R\frac{\binom{\binom{n}{2}-l}{N-l}}{\binom{\binom{n}{2}}{N}}\sim\frac{R}{k!}\frac{(2N)^l}{n^{2l-k}}.$$

If $S_1\in\mathcal{B}_{k,l}^{(n)}$, $S_2\in\mathcal{B}_{k,l}^{(n)}$, $E(S_1)\cap E(S_2)=\emptyset$, then

$$\mathbb{E}\left(\mathbf{1}(S_1)\mathbf{1}(S_2)\right)=\frac{\binom{\binom{n}{2}-2l}{N-2l}}{\binom{\binom{n}{2}}{N}}.$$

If $|V(S_1)\cap V(S_2)|=s$ and $|E(S_1)\cap E(S_2)|=r$, $1\leq r\leq l-1$, i.e. S_1 and S_2 have s mutual vertices and r mutual edges, then

$$\mathbb{E}\left(\mathbf{1}(S_1)\mathbf{1}(S_2)\right)=\frac{\binom{\binom{n}{2}-2l+r}{N-2l+r}}{\binom{\binom{n}{2}}{N}}=O\left(\frac{N^{2l-r}}{n^{4l-2r}}\right).$$

As $S_1\cap S_2\subseteq S_i$, $i=1,2$, i.e. a subgraph, and by our supposition every S is balanced, it follows $\frac{r}{s}\leq\frac{l}{k}$, so $s\geq\frac{rk}{l}$. At most we then have

$$R^2\sum_{j=\frac{rk}{l}}^{k}\binom{n}{k}\binom{k}{j}\binom{n-k}{k-j}=O\left(n^{2k-\frac{rk}{l}}\right)$$

pairs of such subgraphs. We then obtain

$$\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)^2=$$

$$=\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)+\frac{n!R^2}{(k!)^2(n-2k)!}\frac{\binom{\binom{n}{2}-2l}{N-2l}}{\binom{\binom{n}{2}}{N}}+O\left(\left(\frac{N^l}{n^{2l-k}}\right)^2\sum_{r=1}^{l}\left(\frac{n^{2-k/l}}{N}\right)^r\right).\quad(4.1)$$

From the fact that, for $C>0$, $M,N,x\in\mathbb{N}$,

$$f(x)=\log C\binom{M-x}{N-x}=\log C\binom{M-x}{M-N}$$

is a concave function, i.e. $f(x)+f(y)<2f\left(\frac{x+y}{2}\right)$, it follows that

$$\frac{n!}{(k!)^2(n-2k)!}\frac{\binom{\binom{n}{2}-2l}{N-2l}}{\binom{\binom{n}{2}}{N}}\leq\binom{n}{k}^2\frac{\binom{\binom{n}{2}-l}{N-l}^2}{\binom{\binom{n}{2}}{N}^2}.$$

Note that through this estimate, we have taken into account those possible subgraphs which do not have any common edges but do have possibly one or more common vertices.

We set $\frac{N}{n^{2-k/l}}=:\omega$, according to our assumption $\omega\to+\infty$.

Our next goal is to find an approximation for

$$\mathrm{Var}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)=\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)^2-\left(\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)\right)^2$$

We know that

$$\left(\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)\right)^2\sim O\left(\left(\frac{R}{k!}\right)^2\frac{(2N)^{2l}}{n^{2(2l-k)}}\right)=O\left(\frac{R2^l}{k!}\right)^2\omega^{2l}$$

Also,with (4.1)

$$\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)^2=$$

$$=O\left(\left(\frac{N}{n^{2-\frac{k}{l}}}\right)^l\right)+R^2\,O\left(\binom{n}{k}^2\frac{\binom{\binom{n}{2}-l}{N-l}^2}{\binom{\binom{n}{2}}{N}^2}\right)+O\left(\left(\frac{N}{n^{2-\frac{k}{l}}}\right)^{2l}\sum_{r=1}^{l}\left(\frac{n^{2-\frac{k}{l}}}{N}\right)^r\right)\simeq$$

$$\simeq\frac{R^2}{k!^2}n^{2k}n^{-2l}2^lN^l+O(\omega^l)+O\left((\omega)^{2l-1}\right).$$

Thus,

$$\mathrm{Var}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)=O\left(\frac{1}{\omega}\mathbb{E}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)^2\right)$$

With Theorem 3.1, it follows

$$\mathbb{P}_{n,N}\left(\left|\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)-\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbb{E}\mathbf{1}(S)\right)\right|\geq\frac{1}{2}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbb{E}\mathbf{1}(S)\right)\right)=O\left(\frac{1}{\omega}\right)$$

so

$$\mathbb{P}_{n,N}\left(\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbf{1}(S)\right)\leq\frac{1}{2}\left(\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbb{E}\mathbf{1}(S)\right)\right)=O\left(\frac{1}{\omega}\right).$$

From $\omega\to\infty$, $\sum_{S\in\mathcal{B}_{k,l}^{(n)}}\mathbb{E}\mathbf{1}(S)\to\infty$, it follows more than that $G_{n,N}$ contains at least one subgraph isomorphic to an element $B_{k,l}^i\in\mathcal{B}_{k,l}^{(n)}$ with probability tending to 1: There will be $O(\omega^l)$ many of these isomorphic subgraphs, their number tending to ∞. □

This result has been proved for non-balanced subgraphs as well [7, page 85].
Note that this theorem yields the following interesting results:

- The threshold function for the appearance trees of order k is $n^{\frac{k-2}{k-1}}$.
- For $N\gg n$, a random graph will have a cycle of order k, $k\in\mathbb{N}$ asymptotically almost sure.
- Complete subgraphs of order $k\geq 3$ will start appearing for $N\gg n^{2\left(1-\frac{1}{k-1}\right)}$.

4.3 The Giant Component

An intriguing question is the forming of the so-called giant component. That is, as $n \to \infty$, how strongly must $N \to \infty$ so that a proportion of the vertices of $G_{n,N}$ is in the largest connected component of the graph, i.e. the size of the largest component is $\Theta(n)$. Equivalently, one can investigate how large $p(n)$ must be so that $G_{n,p}$ forms a giant component.

In 1959, Erdős and Rényi [21] showed that, for $N = N(c) = n \log n + cn$, the probability that a graph consists of one large connected component (that contains $n - k$ vertices) and k isolated vertices tends to one, where $k = k(n, c)$. Using this result, they showed that the probability of of $G_{n,N(c)}$ being completely connected tends to $e^{-e^{-2c}}$ for $n \to \infty$:

Theorem 4.2. *Considering $G_{n,N}$, where it holds that*

$$N = N_c = \frac{1}{2} n \log n + cn,$$

let $\mathbb{P}_c(n, N)$ of $G_{n,N}$ be the probability of $G_{n,N}$ being completely connected. It then holds that

$$\lim_{n \to \infty} \mathbb{P}_c(n, N) = e^{-e^{-2c}}.$$

Equivalently, this means that for $p \gg \frac{\log n}{n} + \frac{2c}{n}$, $G_{n,p}$ will be connected with the same probability.

A year later, they published a groundbreaking paper [22], that investigated threshold functions for various properties of random graphs. The results included connectivity characteristics when $N(n) \sim cn$ which were surprising.

We shall prove that the threshold function for a giant component is at $N(n) = \frac{n}{2}$, which is equivalent to $np = 1$ for $G_{n,p}$. A curious aspect here is not only the threshold function, but also the phase transition from one state into the next. Erdős and Rényi suggested a "double jump" in the size of the largest component, stating it would change from $O(\log n)$ to $\Theta(n^{\frac{2}{3}})$ and then to $\Theta(n)$. In [28, page 111] some evidence can be found against this statement. For a more detailed description of the phase transition, see [27].

Prior to proving the above statement, we shall need following theorem (Chernoff's inequality) [28, page 26]. We will also need some implications of Chebysheff's inequality. Primarily, we want to show when a random variable will be reasonably close to its mean with high probability. Needing something stronger than Chebysheff's inequality, we shall use this basic idea which follows from Markov's inequality (theorem 3.2). For $u \geq 0$, $t \geq 0$,

$$\mathbb{P}(X \geq \mathbb{E}X + t) = \mathbb{P}(e^{uX} \geq e^{u(\mathbb{E}X + t)}) \leq e^{-u(\mathbb{E}X + t)} \mathbb{E}e^{uX}, \tag{4.2}$$

and vice versa, for $u \leq 0$

$$\mathbb{P}(X \leq \mathbb{E}X - t) = \mathbb{P}(e^{uX} \leq e^{u(\mathbb{E}X - t)}) \geq e^{-u(\mathbb{E}X - t)} \mathbb{E}e^{uX}. \tag{4.3}$$

Assuming X is a sum of independent random variables, $X = \sum_{i=1}^{n} X_i$, with (4.2) we get, for $u \geq 0$

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq e^{-u(\mathbb{E}X + t)} \prod_{i=1}^{n} \mathbb{E}e^{uX_i}. \tag{4.4}$$

We are especially interested in the case where the X_is are indicator functions, thus $X_i \sim Be(p_i)$, with $p_i = \mathbb{P}(X_i = 1) = \mathbb{E}(X_i)$. We will denote $\lambda := \mathbb{E}(X)$. (4.4) now simplifies to

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq e^{-u(\lambda+t)}(1 - p - pe^u)^u \qquad \text{where} \quad \lambda = np. \tag{4.5}$$

Differentiating and setting to zero, we see that the minimum of (4.5) is reached for $e^u = \frac{(\lambda+t)(1-p)}{p(n-\lambda-t)}$ for $n > \lambda + t$. For (4.4), we get

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t} \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} \qquad \text{for} \quad 0 \leq t \leq n - \lambda. \tag{4.6}$$

Theorem 4.3 (Chernoff's inequality)**.** *Let $X \sim Bi(n, p)$ and $\lambda = np$. We define $\varphi(x) := (1 + x)\log(1 + x) - x$, $x \geq -1$, ($\varphi(x) = \infty$ for $x < -1$)*

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right)\right) \leq \exp\left(-\frac{t^2}{2(\lambda + \frac{t}{3})}\right), \qquad t \geq 0, \tag{4.7}$$

$$\mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left(-\lambda\varphi\left(\frac{-t}{\lambda}\right)\right) \leq \exp\left(-\frac{t^2}{2\lambda}\right), \qquad t \geq 0 \tag{4.8}$$

Proof. (4.6) can be written as

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(\frac{-t}{n-\lambda}\right)\right), \quad 0 \leq t \leq n - \lambda.$$

Similarly, by using exactly the same arguments as above for (4.3),

$$\mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right) - (n-\lambda)\varphi\left(\frac{t}{n-\lambda}\right)\right), \quad 0 \leq t \leq \lambda.$$

For all x, it holds that $\varphi(x) \geq 0$, so (4.7) and (4.8) follow directly. Note that we are only interested in the non-trivial cases where $0 \leq t \leq n - \lambda$ or $0 \leq t \leq \lambda$.

As $\varphi(0) = 0$ and $\varphi'(x) = \log(1 + x) \leq x$, it follows $\varphi(x) \geq x^2/2$ for $-1 \leq x \leq 0$, so (4.8) is shown.

(4.7) can be seen similarly: $\varphi(0) = \varphi'(0) = 0$,

$$\varphi''(x) = \frac{1}{x} \geq \frac{1}{(1 + \frac{x}{3})^3} = \left(\frac{x^2}{2(1 + \frac{x}{3})}\right)''$$

so $\varphi(x) \geq x^2/(2(1 + x/3))$. The inequality then follows. □

Lemma 4.2. *Let X_n be a random variable, $\mathbb{E}X_n \to \infty$ and $(\mathbb{E}X_n)^2 \sim \mathbb{E}(X_n^2)$. It follows that $X_n > 0$ a.a.s. and that $\frac{X_n}{\mathbb{E}X_n} \to 1$. This is equal to $\mathbb{P}(\frac{X_n}{\mathbb{E}X_n} \notin (1-\varepsilon, 1+\varepsilon)) \to 0$, $\forall \varepsilon > 0$.*

Proof. With (3.1), we have $\forall \varepsilon > 0$:

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon\mathbb{E}X_n) \leq \frac{\mathbb{V}X_n}{\varepsilon^2(\mathbb{E}X_n)^2} = o(1).$$

□

To continue, we shall need a basic understanding of *branching processes* [26].

We define the following process: At time $t = 0$ Z_0 has Y_0 children, where $Y_0 = Z_1$ is a non-negative random variable in $\mathbb{N}_0$.

At $t = 1$, each of the Y_0 children has $Y_i \geq 0$, $i = 1, \ldots, Z_1$, children, where all the Y_is are random variables with the same distribution. Then, $Z_2 = \sum_{i=1}^{Z_1} Y_i$, and more generally $Z_{k+1} = \sum_{i=1}^{Z_k} Y_i$. This process is called a Galton-Watson process. As soon as there is no more offspring, the entire process dies out, i.e. $Z_n = 0$ implies $Z_{n+l} = 0$, $l > 0$. This is called extinction. Note that $Z_0, Z_1, \ldots$ form a Markov chain. We assume that the distribution of Y_i does not vary over time.

Let $\mathbb{P}(Z_1 = k) = \mathbb{P}(Y_i = k) = p_k$ denote the probability that Z_1 equals k, which is the probability that an object existing at time t has k children at time $t+1$ for $k = 0, 1, 2, \ldots$, $\sum p_k = 1$. We shall need the probability generating function of Y_i, $f(x) = \sum_{k=0}^{\infty} p_k x^k$ for $|x| \leq 1$, and denote its iterates by

$$f_0(x) = x, \qquad f_1(x) = f(x), \qquad \text{and} \qquad f_{n+1}(x) = f\left(f_n(x)\right), \quad n = 1, 2, \ldots \tag{4.9}$$

Note that it follows immediately that $f_{n+1}(x) = f_n\left(f(x)\right)$. Also, $\mathbb{E}Z_1 = f'(x)|_{x=1} = \sum k p_k$. We shall assume two things:

- $\forall k : p_k \neq 1$ and $p_0 + p_1 < 1$. This means that $f(x)$ is strictly convex on the unit interval.
- $\mathbb{E}Z_1 = \sum_{k=0}^{\infty} k p_k$ is finite (and so $f'(x)$ is finite as well).

According to [26], this basic result was already discovered by Watson in 1874:

Theorem 4.4. *The generating function of Z_n is the n^{th} interate $f_n(x)$.*

Proof. Let $f_{(n)}(x)$ be the generating function of Z_n, $n = 0, 1, 2, \ldots$. The conditioned distribution of $(\, Z_{n+1} \mid Z_n = k \,)$ is $(f(x))^k$, $k = 0, 1, 2, \ldots$. Thus, the generating function of Z_{n+1} is

$$f_{(n+1)}(x) = \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) \left(f(x)\right)^k = f_{(n)}\left(f(x)\right), \quad n = 0, 1, \ldots$$

Obviously, $f_{(0)}$ and f_0 are equal, so by induction and the fact that $f_{n+1}(x) = f_n\left(f(x)\right)$ it follows that $f_{(n)}(x) = f_n(x)$, $n = 1, 2, \ldots$. □

The original problem concerning branching processes was the question of the probability of extinction.

Definition 4.5 (Extinction)**.** *The event that for the above defined sequence $(Z_n)_{n\in\mathbb{N}}$ and a $j_N > 0$ it holds $Z_n = 0$, $\forall n \geq j_N$, i.e. $Z_n = 0$ for all but a finite number of n. We denote the probability of extinction with ρ.*

Recall that $\mathbb{P}(Z_{n+1} = 0 \mid Z_n = 0) = 1$, thus the event $(Z_n = 0)$ implies $(Z_{n+1} = 0)$. Thus,

$$\rho = \lim_{n\to\infty} \mathbb{P}(Y_n = 0) = \lim_{n\to\infty} f_n(0).$$

Theorem 4.5. *If $\mu = \mathbb{E}Z_1 \leq 1$, the extinction probability $\rho = 1$. If $\mu > 1$, then ρ is the unique non-negative solution, $\rho < 1$, of the equation*

$$x = f(x). \tag{4.10}$$

Proof. By induction, we see that $f_n(0) < 1$, $n = 0, 1, \ldots$.

Also, combining the fact that $(Z_n = 0)$ implies $(Z_{n+1} = 0)$ and Theorem 4.4, we see that $0 = f_0(0) \leq f_1(0) \leq f_2(0) \leq \cdots \leq \rho = \lim f_n(0)$. With (4.9) we know $f_{n+1}(0) = f(f_n(0))$, and $\lim f_n(0) = \lim f_{n+1}(0) = \rho$, it follows $\rho = f(\rho)$, and thus also $0 \leq \rho \leq 1$.

If $\mu \leq 1$, then $1 \geq f'(1) \geq f'(x)$, $0 \leq x \leq 1$. With help of the mean value theorem, we get $f'(\zeta)(1-x) = f(1) - f(x)$, so $c(1-x) = 1 - f(x)$ for a constant $c < 1$ and $0 \leq x < 1$, thus $f(x) > x$ for $0 \leq x < 1$, so it follows that $\rho = 1$.

On the other hand, if $\mu > 1$, then $f(x) < x$ for $x = 1 - \varepsilon$, $\varepsilon > 0$ sufficiently small. However, $f(0) \geq 0$. It follows that there exists at least one solution for (4.10) in the half-open interval $[0, 1)$. With Rolle's theorem, we know that if there were two solutions, s_0 and t_0 with $0 \leq s_0 < t_0 < 1$, there would exist ξ and η, $s_0 < \xi < t_0 < \eta < 1$ with $f'(\xi) = f'(\eta) = 1$, which is impossible because $f(x)$ is strictly convex.

Also, $\lim f_n(0)$ cannot be 1: $(f_n(0))_{n \geq 0}$ is a nondecreasing sequence, while it holds $f_{n+1}(0) = f(f_n(0)) < f_n(0)$ if $f_n(0)$ is close to (but less than) 1. It follows that ρ is the the only solution of (4.10). □

Before applying this theorem to find the structure of the giant component, let us see these two examples (from [28]).

Example 4.1. Let $X \sim \mathrm{Po}(c)$. Then the probability generating function is

$$f_X(x) = \sum_{i=0}^{\infty} \frac{c^i x^i}{i!} e^{-c} = \exp(c(x-1)).$$

Now, if $c > 1$, with (4.10) and setting $y = 1 - x$ we obtain the probability of extinction $\rho = 1 - \beta(c)$, where $\beta = \beta(c) \in (0, 1)$ is the determined by the equation

$$\beta + e^{-\beta c} = 1 \tag{4.11}$$

Example 4.2. Let $Y_n \sim \mathrm{Bi}(n, p)$, $np \to c > 1$ as $n \to \infty$. The probability generating function of Y_n is

$$f_{Y_n}(x) = \sum_{i=0}^{n} \binom{n}{i} x^i p^i (1-p)^{n-i} = (1 - p + xp)^n,$$

and for every real number x we have

$$\lim_{n \to \infty} f_{Y_n}(x) = \exp(c(x-1)) = f_X(x), \tag{4.12}$$

The probability generating function of Y_n tends pointwise to the probability generating function of $X \sim \mathrm{Po}(c)$. This means that for $n \to \infty$, the probability of extinction $\rho(n, c)$ of a branching process defined by Y_n converges to $1 - \beta(c)$ where β is defined by (4.11).

With these preliminaries, we now define the "almost-" branching process we need [28]. We approach the structure of $G_{n,p}$ in the following way: In the first step of this process, we pick any vertex v in $G_{n,p}$, find all its neighbors $v_1, \dots, v_r$ and then we mark v as saturated.

In the second step, we choose v_1 and mark it as saturated after finding all its neighbors $v_{11}, \dots, v_{1s}$ in $V(G_{n,p}) \setminus \{v, v_1, \dots, v_r\}$. We continue this process until there are no unsaturated vertices left in this component.

If we follow a breadth-first approach during this process — i.e. the vertices closer to v are saturated first — then this process resembles very strongly the branching process. Note that the difference here is that, while we had defined $Z_i = \sum_{j=1}^{Z_{i-1}} Y_j$, i.e. as a sum of a random number of random variables, here at each step we only add *one* more random variable (namely, the number of new neighbors). This new random variable $X_i = X_i(n, m, p)$ has binomial distribution $\mathrm{Bi}(n - m, p)$, where m denotes the number of elements of the component already found. This differs from the classical branching process, where the distribution of the offspring does not change as the process advances. However, for n large and m small, $\mathrm{Bi}(n - m, p) \sim \mathrm{Bi}(n, p)$. The vertex v being contained in a small component can be seen as equivalent to the process dying out. Likewise, the probability that the v is contained in a very large component is equivalent to the process going on for a very long time.

Theorem 4.6. *For $np = c$, $c > 0$ constant:*

i) If $c < 1$, then a.a.s. the largest component of $G_{n,p}$ has at most $\frac{3}{(1-c)^2} \log n$ vertices.

ii) Let $c > 1$ and let $\beta = \beta(c) \in (0, 1)$ be defined as (4.11). Then $G_{n,p}$ contains a giant component of $(1 - o(1))\beta n$ vertices. Also, a.a.s. the size of the second largest component of $G_{n,p}$ is at most $\frac{16c}{(c-1)^2} \log n$.

Proof. We first prove the case $c < 1$. The probability that a given vertex v belongs to a component of size equal or greater than $k = k(n)$ is surely less than the probability that the sum of $k = k(n)$ random variables X_i (as defined above) is at least $k - 1$, because, with the process defined above, this probability is exactly the same as the process "surviving" at least $k - 1$ steps (not counting the vertice we start from).

We can define independent random variables X_i^+ such that $X_i^+ \sim \mathrm{Bi}(n, p)$. Observe that X_i is bounded from above by X_i^+. Also note that $\sum_{i=1}^{k} X_i^+ \sim \mathrm{Bi}(kn, p)$. For large n, the probability that $G_{n,p}$ contains a component of size at least $k \geq 3 \log n/(1 - c)^2$ is bounded from above by

$$n\mathbb{P}\left(\sum_{i=1}^{k} X_i^+ \geq k - 1\right) = n\mathbb{P}\left(\sum_{i=1}^{k} X_i^+ \geq ck + (1 - c)k - 1\right) \leq$$
$$\leq n \exp\left(-\frac{((1 - c)k - 1)^2}{2(ck + (1 - c)k/3)}\right) \leq n \exp\left(-\frac{(1 - c)^2}{2}k\right) = o(1).$$

The first inequality follows through (4.7), the last through some basic transformations and the fact that $k \geq 3^{\log}/_{n}/(1 - c)^2$.

For the case $c > 1$, we set $k_- = \frac{16c}{(c-1)^2} \log n$ and $k_+ = n^{2/3}$. We will start by showing that for every k, $k_- \leq k \leq k_+$, and all vertices v in $G_{n,p}$, *a.a.s.*, our adapted branching process either dies out after fewer than k_- steps, or there are still at least $(c-1)k/2$ vertices which

have been found but not yet saturated. From this, it follows that there is no component in $G_{n,p}$ with k vertices for $k_- \leq k \leq k_+$.

Note that to check if the process starting at v produces at least $(c-1)k/2$ unsaturated vertices after k steps, we need to identify at most $k + (c-1)k/2 = (c+1)k/2$ vertices of the component wherein v lies, namely, the k saturated vertices (one for each step), and the other unsaturated ones.

As in the previous case, we can bound X_i, $1 \leq i \leq k$, this time from below by independent variables $X_i^- \sim \mathrm{Bi}(n - \frac{c+1}{2}k_+, p)$. Also, the probability that the process generates less than $(c-1)k/2$ unsaturated vertices after k steps or even dies out after k steps is smaller than the probability that

$$\sum_{i=1}^{k} X_i^- \leq k - 1 + \frac{(c-1)k}{2}.$$

Thus the probability that the above described situation happens for any vertex v in $G_{n,p}$ and some k, $k_- \leq k \leq k_+$ is, for n large, bounded:

$$n \sum_{k=k_-}^{k_+} \mathbb{P}\left(\sum_{i=1}^{k} X_i^- \leq k - 1 + \frac{(c-1)k}{2}\right) \leq n \sum_{k=k_-}^{k_+} \exp\left(-\frac{(c-1)^2k^2}{9ck}\right) \leq$$

$$\leq nk_+ \exp\left(-\frac{(c-1)^2}{9c}k_-\right) = o(1).$$

The first inequality follows with (4.8) and some basic transformations. Also, $\mathbb{E}\sum_1^k X_i = k\mathbb{E}X_i$ for X_i is distributed binomially with the same parameters.

Let us define

$$K(v) := \{v\} \cup \big\{u \in V(G) : \exists v_1, \ldots, v_k \in V : \{v, v_1\}, \{v_1, v_2\}, \ldots, \{v_k, u\} \in E(G)\big\},$$

i.e. the component of v. We now consider two vertices, v_1 and v_2 such that $|K(v_i)| \geq k_+$ for $i = 1, 2$. We are looking for the probability of these two vertices not being in the same component, i.e. $\mathbb{P}\big(K(v_1) \neq K(v_2)\big)$.

To find this probability, we again use the process as defined above for k_+ steps, starting at v_1. From above, we now know that there exist still at least $(c-1)k_+/2$ unsaturated vertices, all belonging to $K(v_1)$. We then run a similar process starting from v_2. Either we join v_2 to some already found vertice of $K(v_1)$ during the first k_+ steps, or we again have $(c-1)k_+/2$ vertices that still need to be saturated. The probability of there being no edges between the two sets of unsaturated edges is bounded from above by

$$(1-p)^{[(c-1)k_+/2]^2} \leq \exp(-(c-1)^2 n^{1/3} c/4) = o(\frac{1}{n^2})$$

so the probability that v_1 and v_2 lie in the same component tends to one, for any two vertices v_1, v_2 whose component size is greater than k_+. This means that there are two "classes" of vertices, those who belong to components of size smaller than k_- and those who belong to the "giant component" of size greater than k_+.

To complete the proof, we need to know how many of the vertices are in the giant component, or, equivalently, how many are not? Let us estimate the latter number, which

we shall denote by Y. These vertices will henceforth be called "small" vertices. We already know that our process, for small vertices, is bounded from below by the process distributed with $\mathrm{Bi}(n-k_-,p)$. Thus, the probability $\rho(n,p)$ that a vertex v is small (which equals the probability of extinction of the process) is bounded from above by the probability of extinction $\rho_+ = \rho_+(n,p)$ of the process distributed with $\mathrm{Bi}(n-k_-,p)$.

Following the same line of argumentation, $\rho(n,p)$ is bounded from below by $\rho_- + o(1)$, where $\rho_- = \rho_-(n,p)$ is the probability of extinction for the branching process with distribution $\mathrm{Bi}(n,p)$, the term $o(1)$ bounding the probability that the branching process dies later than k_- steps. From example 4.2 we know that for $np \to c$ for $n \to \infty$, both ρ_- and ρ_+ converge to $1-\beta$ as defined in (4.11). Thus, the expectation of the number of small vertices $\mathbb{E}Y = (1-\beta+o(1))n$. We now only need to show that $Y \sim \mathbb{E}Y$.

To do this, we consider $\mathbb{E}Y(Y-1)$, that is, what is the expectation of picking two small vertices. Clearly, the expectation of picking the first vertice is $n\rho(n,p)$. For the second vertex, we then have two possibilities: either both vertices are in the same component, thus there are less than k_- possibilities of picking one of these vertices, or it is in one of the remaining small components. We pick one of these vertices with expectation $n\rho(n - O(k_-),p)$.

Now, $\mathbb{P}\big(\rho(n-O(k_-),p\big)$ is the same as the probability of picking a small vertex in $G_{n-O(k_-),p}$; this implies

$$\rho(n-O(k_-),p) = \frac{\Big|\big\{v \in G_{n-O(k_-),p} : v \text{ small}\big\}\Big|}{n-O(k_-)} \le \frac{\Big|\big\{v \in G_{n,p} : v \text{ small}\big\}\Big|}{n-O(k_-)} = \\ = \frac{\Big|\big\{v \in G_{n,p} : v \text{ small}\big\}\Big|}{n}(1+o(1)) = \rho(n,p)(1+o(1)).$$

We gather

$$\mathbb{E}(Y(Y-1)) \le n\rho(n,p)(k_- + n\rho(n-O(k_-),p)) \le \\ \le n\rho(n,p)(k_- + n\rho(n,p))(1+o(1)) = \\ = (n\rho(n,p))^2(1+o(1)) = (1+o(1))(\mathbb{E}Y)^2.$$

Thus, $\mathbb{E}Y^2 \le (1+o(1))(\mathbb{E}Y)^2 + \mathbb{E}Y = (1+o(1))(\mathbb{E}Y)^2$. From Chebysheff's inequality (Theorem 3.1) and Lemma 4.2 it follows that $G_{n,p}$ contains $(1-\beta+o(1))n$ small vertices, thus the theorem is proved. □

Speaking of the largest component of a network, remember that one often refers to the "giant component." The other components of networks are mostly much smaller than the largest one. Theorem 4.6 is important concerning complex networks because it states one of the few properties of random graphs that complex networks typically possess as well. It remains the basis of many models for "real" networks. The phase transition also stays important in many other models [38, page 199].

4.4 The Clustering Coefficient of Classical Random Graphs

As described in subsection 1.3.3, an index for transitivity is the so-called clustering coefficient, designated by

$$C = \frac{3 \times \text{ number of triangles}}{\text{number connected pairs of vertices}}.$$

We will proceed with the clustering coefficient of a classical random graph, or rather its expected value. Note that for these calculations, we assume that p is small enough, i.e. $p = p(n) \ll 1/n$.

Now we can proceed with calculating the expected number of triangles in a graph $G(V, E) \in G_{n,p}$. Let τ denote the set of all subgraphs of G isomorphic to K_3, $\tau = \{H \subset G, H \cong K_3\}$, and by X_T we denote the the number of these subgraphs, $X_T = |\tau|$. Thus,

$$X_T = \sum_{1 \le i < j < k \le n} \mathbf{1}\big((i,j,k) \in \tau\big) \quad \text{and} \quad \mathbb{E}X_T = \sum_{1 \le i < j < k \le n} \mathbb{P}\big((i,j,k) \in \tau\big) = \binom{n}{3} p^3.$$

To show that X_T is in fact very close to $\mathbb{E}X_T$, we will apply Lemma 4.2:

$$\begin{aligned}
\mathbb{E}X_T^2 &= \mathbb{E}\left(\sum_{1 \le i_1 \le i_2 \le i_3 \le n} \sum_{1 \le j_1 \le j_2 \le j_3 \le n} \mathbf{1}\big((i_1,i_2,i_3) \in \tau\big)\mathbf{1}\big((j_1,j_2,j_3) \in \tau\big) \right) = \\
&= \mathbb{E}\left(\sum_{1 \le i_1 \le i_2 \le i_3 \le n} \sum_{1 \le j_1 \le j_2 \le j_3 \le n} \mathbf{1}\big((i_1,i_2,i_3) \in \tau, (j_1,j_2,j_3) \in \tau\big) \right) = \\
&= \left(\sum_{1 \le i_1 \le i_2 \le i_3 \le n} \sum_{1 \le j_1 \le j_2 \le j_3 \le n} \mathbb{P}\big((i_1,i_2,i_3) \in \tau, (j_1,j_2,j_3) \in \tau\big) \right).
\end{aligned}$$

To evaluate this sum, let us define $M := \{i_1, i_2, i_3\} \cap \{j_1, j_2, j_3\}$. We also define $p_{\mathbf{i,j}} := \mathbb{P}\big((i_1,i_2,i_3) \in \tau \wedge (j_1,j_2,j_3) \in \tau\big)$. We consider three different cases:

i) $|M| = 3$: There are $\binom{n}{3}$ different possibilities of picking these three vertices, the probability of there being a triangle is $p_{\mathbf{i,j}} = p^3$.

ii) $|M| = 2$: There are $12\binom{n}{4}$ different possibilities of picking these vertices and then arranging the vertices as triangles, the probability of there being edges for these triangles $p_{\mathbf{i,j}} = p^4$.

iii) $|M| \le 1$: $p_{\mathbf{i,j}} = p^6$, and we pick the remaining possiblilites of building couples, i.e. $\binom{n}{3}^2 - \binom{n}{3} - 12\binom{n}{4}$.

Thus,

$$\begin{aligned}
\mathbb{E}X_T^2 &= \binom{n}{3} p^3 + 12\binom{n}{4} p^5 + \left(\binom{n}{3}^2 - \binom{n}{3} - 12\binom{n}{4} \right) p^6 = \\
&= (\mathbb{E}X_T)^2 + \binom{n}{3} p^3(1-p^3) + 12\binom{n}{4} p^5(1-p) = \\
&= (\mathbb{E}X_T)^2 + O(n^4 p^5).
\end{aligned}$$

Note that from the proof of Theorem 4.1, we know that as soon as N (or, equivalently, p) is greater than the threshold function, there are infinitely many subgraphs of a certain

type (in this case, triangles) that will appear in $G_{n,p}$. In this case, it has to hold $p \gg \frac{2}{n}$. For p large enough, it holds $\mathbb{E}X_T \to \infty \Leftrightarrow np \to \infty$.

With the previous estimates, it follows that $\mathbb{E}(X_T^2) \sim (\mathbb{E}X_T)^2$ and with Lemma 4.2, $\frac{X_T}{\mathbb{E}X_T} \to 1$, thus $\mathbb{E}X_T = \binom{n}{3} p^3$.

The expected number of connected pairs of edges is calculated in a similar manner. Let ρ denote the set of all subgraphs of G isomorphic to $\tilde{H}(V,E)$ where $V = \{v_1, v_2, v_3\}$ and $E = \{(v_1, v_2), (v_1, v_3)\}$, and thus $\rho = \{H \subset G, H \cong \tilde{H}\}$. By X_R let us denote the the number of these subgraphs, $X_R = |\rho|$.

For a set of three vertices, $i_1 < i_2 < i_3$, we define a random variable $\mathcal{Z}_{i_1 i_2 i_3}$ as follows:

$$\mathcal{Z}_{i_1 i_2 i_3} := \mathbf{1}(i_1) + \mathbf{1}(i_2) + \mathbf{1}(i_3)$$

where $\mathbf{1}(i_k) = 1$, $k = 1, 2, 3$, if there is a pair of edges centered at i_k, with the edges connecting vertices from $\{i_1, i_2, i_3\}$. Thus,

$$X_R = \sum_{1 \le i_1 < i_2 < i_3 \le n} \mathcal{Z}_{i_1 i_2 i_3} \qquad \text{and}$$
$$\mathbb{E}X_R = \sum_{1 \le i_1 < i_2 < i_3 \le n} \mathbb{E}\mathcal{Z}_{i_1 i_2 i_3} = \binom{n}{3} 3p^2.$$

Arguing as with triangles,

$$\mathbb{E}X_R^2 = \mathbb{E}\left(\sum_{1 \le i_1 < i_2 < i_3 \le n} \sum_{1 \le j_1 < j_2 < j_3 \le n} \mathcal{Z}_{i_1 i_2 i_3} \mathcal{Z}_{j_1 j_2 j_3} \right) =$$
$$= \sum_{1 \le i_1 < i_2 < i_3 \le n} \sum_{1 \le j_1 < j_2 < j_3 \le n} \mathbb{E}\big((\mathbf{1}(i_1) + \mathbf{1}(i_2) + \mathbf{1}(i_3))(\mathbf{1}(j_1) + \mathbf{1}(j_2) + \mathbf{1}(j_3)) \big).$$

Similar to the above case, we define $M := \{i_1, i_2, i_3\} \cap \{j_1, j_2, j_3\}$ and we also define L as those edges present connecting both vertice of i_1, i_2, i_3 and vertices of j_1, j_2, j_3. We define $p_{\mathbf{i,j}} := \mathbb{E}\left(\mathcal{Z}_{i_1 i_2 i_3} \mathcal{Z}_{j_1 j_2 j_3}\right)$. We proceed as with triangles:

i) $|M| = 3$: To start with, there are $\binom{n}{3}$ ways of picking these three vertices, and $p_{\mathbf{i,j}} = 9p^4$.

ii) $|M| = 2$: We can choose the vertices in $12\binom{n}{4}$ different ways. There are two cases:
   a) $|L| = 1$: We distinguish how many edges there are in this construction in total; for three this gives us $p_{\mathbf{i,j}} = 4p^3(1-p)^2$, when there are four edges we have $12p^4(1-p)$ and five egdes present renders $9p^5$.
   b) $|L| = 0$: Here, $p_{\mathbf{i,j}} = p^4(1-p)$

   In sum, for $|M| = 2$, we then have $12\binom{n}{4} p^3(4 + 5p)$.

iii) $|M| \le 1$: As with triangles above, we shall subtract; $p_{\mathbf{i,j}} = 9p^4$, and there are $\binom{n}{3}^2 - \binom{n}{3} - 12\binom{n}{4}$ ways left of picking the vertices.

Thus,

$$\mathbb{E}X_R^2 = \left(\binom{n}{3}^2 - \binom{n}{3} - 12\binom{n}{4} \right) 9p^4 + \binom{n}{3} 9p^2 + 12\binom{n}{4} p^4(1-p) = O\left((\mathbb{E}X_R)^2 \right).$$

Like above, for n large enough, it then follows that $\mathbb{E}X_R^2 \sim (\mathbb{E}X_R)^2$ and with Lemma 4.2, it follows $\frac{X_R}{\mathbb{E}X_R} \to 1$

We can then proceed to evaluate the clustering coefficient $C(G_{n,p})$ in expectation,

$$\mathbb{E}C(G_{n,p}) = \frac{3\mathbb{E}X_T}{\mathbb{E}X_R} = \frac{3\binom{n}{3}p^3}{\binom{n}{3}p^2} = 3p.$$

$p(n)$ normally tends to 0 as $n \to \infty$. Depending on $p = p(n)$, this can and usually will be quite small. In regard to real networks, far too small. Most "real" networks have clustering coefficients above 0.01. The network of company directors even has a clustering coefficient of 0.59 [38, page 182]. In fact, when Watts and Strogatz presented their "small-world" model [45], one of its key features was to combine a high clustering coefficient with a small degree of seperation. The low clustering coeffiecient is one of the reasons why random graphs are a poor model for complex networks.

4.5 The Degree Sequence of Random Graphs

What else makes the classical Erdős-Rényi model a rather unsatisfactory abstraction for real world networks? As we shall see now, if we denote by X_k the number of vertices with degree k, X_k has asymptotically Poisson distribution with mean $\lambda_k = n\mathrm{Bi}(k; n-1, p)$, i.e. $n\binom{n-1}{k}p^k(1-p)^{n-1-k}$ [7].

It will be enough to consider the case where $p \geq \varepsilon n^{-3/2}$ because of the following considerations. Let $Y_k = \sum_{i \geq k} X_i$, that is, Y_k is the number of vertices of degree at least k. If $p = o(n^{-3/2})$, then

$$\mathbb{E}Y_2 = \sum_{j=2}^{n-1} \mathbb{E}X_j \leq n \sum_{j=2}^{n-1} \binom{n-1}{j} p^j \leq \sum_{j=2}^{\infty} \frac{(pn)^j}{j!} = o(1).$$

This means that a.a.e. $G_{n,p}$ consists of independent edges. In this case, if $pn^2 \to \infty$, then there will be $2M$ vertices of degree 1, where $M = |E(G_{n,p})| \sim pn^2/2$, and all other vertices will have degree 0. As this case is not very interesting, we set p fairly enough in following theorem.

Theorem 4.7. *Let $\varepsilon > 0$ be fixed, $\varepsilon n^{-3/2} \leq p = p(n) \leq 1 - \varepsilon n^{-3/2}$, let $k = k(n) \in \mathbb{N}$ and set $\lambda_k = \lambda_k(n) = nBi(k; n-1, p) = n\binom{n-1}{k}p^k(1-p)^{n-1-k}$. Then the following assertions hold:*

i) If $\lim \lambda_k(n) = 0$, then $\mathbb{P}(X_k = 0) = 1$.

ii) If $\lim \lambda_k(n) = \infty$, then $\lim \mathbb{P}(X_k \geq t) = 1$ for every fixed t.

iii) If $0 < \underline{\lim}\lambda_k(n) \leq \overline{\lim}\lambda_k(n) \leq \infty$, then X_k has asymptotically Poisson distribution with mean λ_k:

$$\mathbb{P}(X_k = r) \sim e^{-\lambda_k}\frac{\lambda_k^r}{r!}$$

for every fixed r.

Prior to proving this theorem, we shall need some preliminary results.

Lemma 4.3. *Let $X \in \mathbb{N}$, be a random variable. It follows that $\mathbb{P}(X > 0) \leq \mathbb{E}X$.*

Proof. $\mathbb{E}X = \sum_{k\geq 0} k\mathbb{P}(X = k) \geq \sum_{k\geq 1} \mathbb{P}(X = k) = \mathbb{P}(X > 0)$ □

Theorem 4.8. *Let $S_{n,p}$ be a random variable, $S_{n,p} \sim Bi(n, p)$. Let $u > 1$ and $1 \leq m = \lceil upn \rceil \leq n - 1$. Then*

$$\mathbb{P}(S_{n,p} \geq upn) = \mathbb{P}(S_{n,p} \geq m) < \frac{u}{u-1}\binom{n}{k}p^k(1-p)^{n-k} \leq$$
$$\leq \frac{1}{\sqrt{2\pi}}\frac{u}{u-1}\left(\frac{n}{m(n-m)}\right)^{1/2} u^{-upm}\left(\frac{1-p}{1-up}\right)^{(1-up)n}.$$

For a proof of this theorem, see for example [7].

Theorem 4.9 (See [7])**.** *Let $\lambda = \lambda(n)$ be a non-negative bounded function on $\mathbb{N}$. With $\mathbb{E}_r$ we denote the r^{th} factorial moment of a function. Suppose the non-negative integer valued random variables $X_1, X_2, \ldots$ are such that*

$$\lim_{n\to\infty}(\mathbb{E}_r X_n - \lambda^r) = 0, \qquad r = 0, 1, \ldots$$

Then

$$d(X_n, Po(\lambda)) \to 0.$$

Proof of Theorem 4.7. Without loss of generality, we assume $p \leq 1/2$. Note that $\mathbb{E}X_k = \lambda_k(n)$, which, with Lemma 4.3, yields

$$\mathbb{P}(X_k \geq 1) \leq \mathbb{E}X_k = \lambda_k(n).$$

The first assertion then follows.

We now only consider $\underline{\lim}\,\lambda_k(n) > 0$. We will conclude the remaining proof by showing that for every fixed $r \geq 1$, the r^{th} factorial moment $\mathbb{E}_r X_k$ of X_k is asymptotic to $\lambda_k(n)^r$.

$\mathbb{E}_r X_k$ is the expected number of ordered r-tuples of vertices $x_1, x_2, \ldots, x_r$ such that each vertex x_i has degree k. Let us commence by considering the probability that r given vertices $x_1, x_2, \ldots, x_r$ all have degree k. Let there be l edges joining the x_is, and let vertex x_i be joined to $\tilde{d}_i \leq k$ vertices x_j, $1 \leq j \leq r$. Obviously, $\sum_{i=1}^r \tilde{d}_i = 2l$ and x_i has to be joined to $k - \tilde{d}_i$ vertices outside the set $\{x_1, x_2, \ldots, x_r\}$. The probability of this event is

$$\prod_{i=1}^{r} \mathrm{Bi}(k - \tilde{d}_i; n - r, p). \tag{4.13}$$

Let us first consider the case where $p(n) \leq 1/2$ is bounded away from 0. With Lemma 4.8, $\underline{\lim}\,\lambda_k(n) > 0$ implies that $k(n)$ is about pn, namely, we have $k = pn\alpha_n$ where $\alpha = 1 + o(n)$. It follows

$$\frac{\mathrm{Bi}(k-d; n-r; p)}{\mathrm{Bi}(k; n-1; p)} = \frac{(n-r)_{k-d}(k)_d}{(n-1)_k p^d (1-p)^{r-d-1}} =$$
$$= \frac{(n-r)_{k-r+1}(k)_d p^{-d}(n-k-1)_{r-d-1}(1-p)^{-r+d+1}}{(n-1)_k} \sim$$
$$\sim \frac{(n-r)_{k-r+1} n^d n^{r-d-1}}{(n-1)_k} = \frac{(n-r)!n^{r-1}}{(n-1)!} \sim 1.$$

Thus, with (4.13), it follows that

$$\mathbb{E}X_k \sim (n)_r \mathrm{Bi}(k; n-1, p)^r \sim \lambda_k^r \tag{4.14}$$

Now, let us consider the case where $p = o(1)$. With (4.13) we can bound $\mathbb{E}_r X_k$ the following way:

$$\mathbb{E}_r X_k \leq (n)_r \sum_{l=0}^{R} \binom{R}{l} p^l q^{R-l} \max_{\sum_{i=} 1^r d_i^* = 2l} \prod_{i=1}^{r} \mathrm{Bi}(k - d_i^*; n-r, p),$$

where $R = \binom{r}{2}$ and the maximum is over all sequences $d_1^*, d_2^*, \ldots, d_r^*$ with $\sum_{i=} 1^r d_i^* = 2l$ and $0 \leq d_i^* \leq \min\{r-1, k\}$. Note that $k(n) = o(n)$. Thus, for $0 \leq d \leq \min\{r-1, k\}$ and n sufficiently large,

$$\frac{\mathrm{Bi}(k-d; n-r, p)}{\mathrm{Bi}(k; n-r, p)} \leq \frac{(k)_d}{(n-r-k)_d} p^{-d} \leq 2 \left(\frac{k}{pn}\right)^d,$$

because $(n-r-k)_d > (n-r-k-d+1)^d = n^d(1-o(1))^d > \frac{n^d}{2}$. We then obtain

$$\mathbb{E}X_k \leq n^r \mathrm{Bi}(k; n-r, p)^r \left(1 + \sum_{l=1}^{R} \binom{R}{l} p^l 2^r \left(\frac{k}{pn}\right)^{2l}\right) \leq$$

$$\leq n^r \mathrm{Bi}(k; n-r, p)^r \left(1 + 2^{r^2} \sum_{l=1}^{R} \left(\frac{k^2}{pn^2}\right)^l\right). \tag{4.15}$$

By assumption, $\underline{\lim}\, \lambda_k(n) > 0$, which implies that $k^2 = o(pn^2)$: If this were not the case, there would be some $\eta > 0$ such that $k \geq np^{1/2}$ for n arbitrarily large. This would imply

$$\underline{\lim}\, \lambda_k(n) \leq \underline{\lim}\, n \binom{n}{k} p^k \leq \underline{\lim}\, n \left(\frac{en}{k}\right)^k p^k =$$

$$= \underline{\lim}\, n \left(\frac{epn}{k}\right)^k \leq \underline{\lim}\, n \left(\frac{ep^{1/2}}{\eta}\right)^{\eta p^{1/2} n} \leq$$

$$\leq \underline{\lim}\, n \left(\left(\frac{en^{-3/4}}{\eta}\right)^{n^{-3/4}}\right)^{n\eta} = \underline{\lim}\, n\, \frac{en^{3/4\eta n^{1/4}}}{\eta} = 0,$$

which is contrary to our assumptions. The last inequalities follow from the fact that for p close to zero, $(ap)^{bp}$, $a, b > 0$ is strictly decreasing, thus for our purpose reaches it's maximum at $p = n^{-3/4}$. As $k^2 = o(pn^2)$, (4.15) implies that

$$\mathbb{E}_r X_k \leq n^r \mathrm{Bi}(k; n-r, p)^r \,(1 + o(1)) = \lambda_k^r \,(1 + o(1)).$$

Vice versa, it holds for independent r-tuples of vertices of degree k:

$$\mathbb{E}_r X_k \geq (1-p)^R (n)_r \mathrm{Bi}(k; n-r, p)^r = \lambda_k^r \,(1 + o(1)).$$

Note that $q \to 1$, so

$$\mathbb{E}_r X_k \sim \lambda_k^r, \tag{4.16}$$

so the r^{th} factorial moment of X_k is asymptotic to the r^{th} factorial moment of $\text{Po}(\lambda_k)$. For $\lim \lambda_k(n) = \infty$, then $\mathbb{E}_2 X_k \sim \lambda_k^2$ implies that $\mathbb{E}X_k^2 = \lambda_k^2\,(1 + o(1)) = (\mathbb{E}X_k)^2$, so with Lemma 4.2 we have

$$\lim_{n\to\infty} \mathbb{P}(X_k \geq t) = 1$$

for every fixed t.

On the other hand, if $\overline{\lim}\, \lambda_k < \infty$, then with Theorem 4.9, (4.14) and (4.16) we get that asymptotically X_k has Poisson distribution with mean λ_k. □

4.6 The Diameter

The last property we shall investigate concerning random graphs is their diameter. Seeing that the small-world effect is defined by having diameter that is of the size of the diameter of a random graph, we must show how big this diameter actually is, namely $O(\log n)$.

However, before we prove anything, we will need some preliminary thoughts and four lemmas [7, page 229 ff].

Let us denote by $\Gamma_k(x)$ the set of vertices at distance k from a vertex x of $G \in G_{n,p}$, and by $N_k(x)$ the set of vertices within distance k from x, i.e.

$$\Gamma_k(x) = \{y \in G : d(x,y) = k\} \qquad \text{and} \qquad N_k(x) = \bigcup_{i=0}^{k} \Gamma_i(x).$$

This means that $\operatorname{diam} G \leq d$ iff $N_d(x) = V(G)$ for every vertice x and $\operatorname{diam} G \geq d$ iff there is a vertice y such that $N_{d-1}(y) \neq V(G)$.

Intuitively, $G_{n,p}$ should have small diameter: A vertice x will have about pn neighbors; each of these neighbors will have around another pn neighbors, so $\Gamma_2(x)$ would be presumed to be around $(pn)^2$, and so on. Generally speaking, a subset $W \subseteq V$ would be expected to have only slightly less than $|W| \cdot pn$ neighbors. Thus, we would anticipate that $\Gamma_k(x)$ would be be only slightly less than $(pn)^k$.

If we set $d = d(n) = 2k$, and we assume that k is large so that both $|\Gamma_k(x)|$ and $|\Gamma_k(y)|$ are large enough, with very high probability, either $\Gamma_k(x) \cap \Gamma_k(y) \neq \emptyset$ or else there exists an edge $(v_i, v_j) \in G_{n,p}$ connecting some vertice v_i in $\Gamma_k(x)$ with some vertice $v_j \in \Gamma_k(y)$. It thus follows that with large probability, $\operatorname{diam} G_{n,p} \leq 2k + 1 = d + 1$.

Numerous assumptions are necessary to prove this, as well as the use of concrete constants. As we will work with dependent events, we will need many conditional probabilities. We will also need following theorem [7, page 13]:

Theorem 4.10. *Let* $S_{n,p} \sim \text{Bi}(n,p)$, $q = 1 - p$. *Suppose* $0 < p \leq 1/2$, $\varepsilon pqn \geq 12$ *and* $0 < \varepsilon \leq 1/12$. *Then*

$$\mathbb{P}(|S_{n,p} - pn| \geq \varepsilon pn) \leq (\varepsilon^2 pn)^{-1/2} e^{-\varepsilon^2 pn/3}.$$

In the following lemmas, we shall suppose certain things:

- $0 < p = p(n) < 1$,
- $d = d(n) \in \mathbb{N}$, $d \geq 2$,
- $p^d n^{d-1} = \log(n^2/c)$, $c > 0$ and $\frac{pn}{\log n} \to \infty$ as $n \to \infty$.

What we are going to show is that, for the above restrictions, $G_{n,p}$ has diameter d or $d+1$. Note that $\lim_{n\to\infty}\mathbb{P}(G_{n,p} \text{ is connected}) = \lim_{n\to\infty}\mathbb{P}(\operatorname{diam} G_{n,p} < \infty)$. It can be shown that this probability is greater 0 if $p \geq \frac{\log n + c_0}{n}$ for some constant c_0. See Theorem 4.2.

It can also be shown (see [7, page 41]) that if p or even $pn^{1/2-\varepsilon}$ is bounded away from 0, then $\operatorname{diam} G_{n,p} \leq 2$ for almost every $G_{n,p}$. We thus assume that

- $p = o(n^{-1/2+\varepsilon})$ for every $\varepsilon > 0$.

With the above assumptions, it follows that

$$p = n^{1/d-1}\left(\log n^2/c\right)^{1/d}$$

and

$$d = \left(\log n + \log\log n + \log 2 + O\left(\frac{1}{\log n}\right)\right)\frac{1}{\log(pn)}.$$

Because $pn \gg \log n$, $d = O(\log n/\log\log n)$. It also follows that $(pn)^{d-1} = \log(\frac{n^2}{c})\frac{1}{p} = o(n)$, and thus $p(pn)^{d-2} = o(1)$. As we typically let $n \to \infty$, we may assume $n \geq 100$, $pn > 100\log n$, $(pn)^{d-2} < n/10$ and so $p(pn)^{d-2} < 1/10$.

We will now prove two lemmas that show that $\Gamma_k(x)$ is quite likely to be very close to $(pn)^k$.

Lemma 4.4. *Let x be a fixed vertex, let $1 \leq k = k(n) \leq d-1$ and let $K = K(n)$ satisfy*

$$6 \leq K < \frac{1}{12}\left(\frac{pn}{\log n}\right)^{1/2}.$$

Let $\Omega_k \subset \mathbb{G}_{n,p}$ be the set of graphs for which $a = |\Gamma_{k-1}(x)|$ and $b = |N_{k-1}(x)|$ satisfy

$$\frac{1}{2}(pn)^{k-1} \leq a \leq \frac{3}{2}(pn)^{k-1}$$

and

$$b \leq 2(pn)^{k-1}.$$

Set

$$\alpha_k := K\left(\frac{\log n}{(pn)^k}\right)^{1/2}, \qquad \beta_k := p(pn)^{k-1} \qquad \text{and} \qquad \gamma_k := \frac{2}{n}(pn)^{k-1} = \frac{2}{pn}\beta_k.$$

Then

$$\mathbb{P}\left(\left||\Gamma_k(x)| - apn\right| \geq (\alpha_k + \beta_k + \gamma_k)\,apn \mid \Omega_k\right) \leq n^{-K^2/9}.$$

Proof. First of all, note that α_k, β_k and γ_k are all very small, and also that K can become very large. We will show that with high probability, $|\Gamma_k(x)|$ is very close to $pn\big|\Gamma_{k-1}(x)\big|$. The idea of this proof reminds of proving Theorem 4.6.

To begin with, it is necessary to find $\Gamma_{k-1}(x)$ and $N_{k-1}(x)$. We will do this by first testing which vertices are adjacent to x, i.e. finding $\Gamma_1(x)$, then identifying the vertices

adjacent to $\Gamma_1(x)$, $\Gamma_2(x)$, and so on, until we arrive at Γ_{k-2}. Obviously, in the j^{th} step, we are only interested in "new" vertices, i.e. those not already contained in N_{j-1}.

The probability that a vertex $y \notin N_{k-1}(x)$ is joined to vertices in $\Gamma_{k-1}(x)$, conditional on Ω_k, is $p_a = 1-(1-p)^a$. Because $ap \to 0$, expanding with the binomial theorem, it follows easily that

$$ap(1-\frac{ap}{2}) \le p_a \le ap.$$

We can easily see that, conditional on Ω_k, the random variable $|\Gamma_k(x)|$ has binomial distribution with parameters $n_k = n-b$ and p_a. Since $(pn)^{k-1} \le (pn)^{d-2} \le n/10$, we have $b \le 2(pn)^{k-1} < n/5$, and so $4n/5 < n_k \le n$. Also, $ap(n-n_k) \le \gamma_k apn$ and $(ap-p_a) \le \beta_k ap$. Thus, by Theorem 4.10 we get

$$\begin{aligned}\mathbb{P}\left(||\Gamma_k(x)| - apn| \ge (\alpha_k+\beta_k+\gamma_k)\,apn \mid \Omega_k\right) \le \\ \le \mathbb{P}\left(||\Gamma_k(x)| - apn_k| \ge (\alpha_k+\beta_k)\,apn_k \mid \Omega_k\right) \le \\ \le \mathbb{P}\left(||\Gamma_k(x)| - p_a n_k| \ge \alpha_k apn \mid \Omega_k\right) \le \\ \le (\alpha_k^2 p_a n_k)^{-1/2}\exp(-\alpha_k^2 p_a n_k/3) \le \\ \le \exp(-\alpha_k^2 p_a n_k/3) \le \exp(-\alpha_k^2 (pn)^k/9) = n^{-K^2/9}.\end{aligned}$$

We could apply Theorem 4.10 because

$$\begin{aligned}0 < p_a \le pa \le \frac{3}{2}p(pn)^{k-1} \le \frac{3}{2}p(pn)^{d-2} \le \frac{1}{2},\\ 0 < \alpha_k = K\left(\frac{\log n}{(pn)^k}\right)^{1/2} \le K\left(\frac{\log n}{(pn)}\right)^{1/2} < \frac{1}{12}\end{aligned}$$

and

$$\alpha_k p_a(1-p_a)n_k \ge \frac{K}{2}\left(\frac{\log n}{(pn)^k}\right)^{1/2}\cdot\frac{3}{10}(pn)^k \ge 2K > 12.$$

□

Lemma 4.5. *Let $K > 12$ be a constant and define Ω_k, α_k, β_k, γ_k, $k = 1, 2, \ldots, d-1$, as in Lemma 4.4. Set*

$$\delta_k = \exp\left(2\sum_{l=1}^{k}(\alpha_l+\beta_l+\gamma_l)\right) - 1.$$

If n is sufficiently large, then with probability at least $1-n^{-K-2}$ for every vertex x and every natural number k, $1 \le k \le d-1$, we have

$$\left||\Gamma_k(x)| - (pn)^k\right| \le \delta_k (pn)^k.$$

Proof. As already noted in Lemma 4.4, for $n \to \infty$, α_k, β_k, γ_k tend to zero, and so $\delta_{d-1} \to 0$. We thus can assume that $\delta_{d-1} < 1/4$.

Let x be a fixed vertice and denote by Ω_k^* the set of graphs for which

$$\left||\Gamma_l(x)| - (pn)^l\right| \le \delta_l (pn)^l, \qquad 0 \le l \le k.$$

Because we are assuming that $\delta_{d-1} < 1/4$, it follows that also $\delta_{k-1} < 1/4$ for $k = 1, 2, \ldots, d-1$. With the definition of Ω_k and the fact that Ω_k^* can be seen as a type of decreasing "neighborhood" property, it follows that

$$\Omega_k^* \subset \Omega_{k-1}^* \subset \Omega_k.$$

We shall prove by induction that

$$1 - \mathbb{P}(\Omega_k^*) \leq 2kn^{-K^2/9} \tag{4.17}$$

for every k, $0 \leq k \leq d-1$. For $k = 0$, there is nothing to prove. Assume that $1 \leq k < d-1$ and (4.17) holds for smaller values of k. Obviously,

$$1 - \mathbb{P}(\Omega_k^*) = 1 - \mathbb{P}(\Omega_{k-1}^*) + \mathbb{P}(\Omega_{k-1}^*)\mathbb{P}\left(\left||\Gamma_k(x)| - (pn)^k\right| \geq \delta_k(pn)^k \mid \Omega_{k-1}^*\right).$$

For any graph in Ω_{k-1}^*, $a = |\Gamma_{k-1}(x)|$ satisfies $|(pn)^{k-1} - a| \leq \delta_{k-1}(pn)^{k-1}$, which holds true after multiplying with pn. Thus,

$$\begin{aligned}
\mathbb{P}\left(\left||\Gamma_k(x)| - (pn)^k\right| \geq \delta_k(pn)^k \mid \Omega_{k-1}^*\right) \leq \\
\leq \mathbb{P}(\Omega_{k-1}^*)^{-1}\mathbb{P}\left(\left||\Gamma_k(x)| - apn\right| \geq (\delta_k - \delta_{k-1})(pn)^k \mid \Omega_k\right) \leq \\
\leq (1 - 2(k-1)n^{-K^2/9})^{-1}\mathbb{P}\left(\left||\Gamma_k(x)| - apn\right| \geq 2(\alpha_k + \beta_k + \gamma_k)(pn)^k \mid \Omega_k\right) \leq \\
\leq \mathbb{P}\left(\left||\Gamma_k(x)| - apn\right| \geq 2(\alpha_k + \beta_k + \gamma_k)apn \mid \Omega_k\right) \leq \\
\leq 2n^{-K^2/9}.
\end{aligned}$$

The first inequality follows with conditional probabilities, the second because $dn^{K^2/9} < 1$ and the last because of Lemma 4.4. We now have

$$1 - \mathbb{P}(\Omega_k^*) \leq 2(k-1)n^{-K^2/9} + 2n^{-K^2/9} = 2kn^{-K^2/9},$$

so (4.17) is shown. The statement of Lemma 4.5 now follows quickly. □

The next two lemmas, which we shall state without proof, show that with high probability, two vertices are far from each other.

For two distinct vertices x and y, and $k \in \mathbb{N}$, let us define

$$\begin{aligned}
\Gamma_k^*(x,y) := \Big\{ & z \in \Gamma_k(x) \cap \Gamma_k(y) : \\
& \Gamma(z) \cap \{\Gamma_{k-1}(x) \setminus \gamma_{k-1}(y)\} \neq \emptyset \quad \wedge \quad \Gamma(z) \cap \{\Gamma_{k-1}(y) \setminus \gamma_{k-1}(x)\} \neq \emptyset \Big\}.
\end{aligned}$$

This effectively means that $|\Gamma_k^*(x,y)|$ is the number of paths from x to y that have exactly $2k-1$ edges.

Denote by $\Pi_k \subset \mathbb{G}(n,p)$ the set of graphs of which $|\Gamma_{k-1}(x)| \leq 2(pn)^{k-1}$ and $|\Gamma_{k-1}(y)| \leq 2(pn)^{k-1}$. Assume $K > e^7$ constant, and for $1 \leq k \leq d/2$ define $c_k = c_k(n,p,K)$ by

$$c_k 4p^{2k}n^{2k-1} = (K+4)\log n,$$

and set

$$m_k = m_k(n,p,K) := \frac{2(K+4)\log n}{\log c_k}.$$

Finally, for $d/2 < k \leq d$, set $m_k := m_k(n,p) = 2p^{2k}n^{2k-1}$.

Lemma 4.6. *If n is sufficiently large, then for every k, $1 \leq k \leq d-1$, we have*

$$\mathbb{P}\left(|\Gamma_k^*(x,y)| \geq m_k \mid \Pi_k\right) \leq n^{-K-4}.$$

□

Lemma 4.7. *If n is sufficiently large, then, with probability at least $1-n^{-K}$, the following assertions hold:*

i) For every vertice x,

$$|N_{d-2}(x)| < 2(pn)^{d-2} \qquad \text{and} \qquad \left|\Gamma_{d-1}(x) - (pn)^{d-1}\right| \leq \delta_{d-1}(pn)^{d-1},$$

where δ_{d-1} is the number defined in Lemma 4.5.

ii) For distinct vertices x and y,

$$|N_{d-1}(x) \cap N_{d-1}(y)| \leq 8p^{2d-2}n^{2d-3}$$

and

$$|\Gamma\big(N_{d-1}(x) \cap N_{d-1}(y)\big)| \leq 16p^{2d-1}n^{2d-2}.$$

□

Finally, we can present the main theorem of the section [7, page 233].

Theorem 4.11. *Let c be a positive constant, $d = d(n) \geq 2$ a natural number, and define $p = p(n,c,d)$, $0 < p < 1$, by*

$$p^d n^{d-1} = \log(n^2/c).$$

Suppose that $pn/(\log n)^3 \to \infty$. Then in $\mathbb{G}(n,p)$ we have

$$\lim_{n\to\infty} \mathbb{P}(\textit{diam } G = d) = e^{-c/2} \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}(\textit{diam } G = d+1) = 1 - e^{-c/2}$$

Proof. Theorem 4.11 states that almost surely, $G_{n,p}$ has diameter either d or $d+1$. We will prove this by considering the number of pairs of vertices "far" from each other.

When do we call two vertices "far" apart? We say y is *remote from* x if $y \notin N_d(x)$ (and, equivalently, $x \notin N_d(y)$). We then call (x,y) a *remote pair*.

Let $X = X(G)$ be the number of remote pairs of vertices in G. We shall show that the distribution of X tends to the Poisson distribution with parameter $c/2$. Assuming $\mathbb{P}(\operatorname{diam} G_{n,p} \leq d-1) + \mathbb{P}(\operatorname{diam} G_{n,p} \geq d+2) = o(1)$, this is enough to imply the assertion of this theorem.

Let us start by proving that almost no $G_{n,p}$ contains two remote pairs sharing a vertice. Let x, y and z be fixed distinct vertices. By Lemma 4.5, we know that with probability $1-n^{-K-2}$, $|\Gamma_k(x)| = (pn)^k(1+o(1))$, so, by summing up over these different "neighborhoods" (each disjoint) and estimating somewhat crudely, we get

$$\mathbb{P}\left(|N_{d-1}(x)| < \frac{5}{6}(pn)^{d-1}\right) < n^{-4}$$

provided n is sufficiently large: $n \geq n(p,d)$.

The probability that y is joined to no vertice in a set $W \subset V(G) \setminus \{y\}$, $|W| \geq \frac{5}{6}(pn)^{d-1}$, is

$$(1-p)^{|W|} \leq \exp\left(-\frac{5}{6}p(pn)^{d-1}\right) = \exp\left(-\frac{5}{6}\log(n^2/c)\right) = c^{5/6}n^{-5/3}.$$

It now follows that

$$\begin{aligned}\mathbb{P}\Big(x \text{ is remote from both } y \text{ and } z\Big) \leq \\ \leq \mathbb{P}\left(|N_{d-1}(x)| < \frac{5}{6}(pn)^{d-1}\right) + \mathbb{P}\left(\{y,z\} \cap N_d(x) = \emptyset \mid |N_{d-1}(x)| > \frac{5}{6}(pn)^{d-1}\right) \leq \\ \leq \quad n^{-4} + c^{5/3}n^{-10/3} \quad < n^{-3-1/4}.\end{aligned}$$

Therefore, the probability that $G_{n,p}$ contains two remote pairs sharing a vertice is at most

$$n\binom{n-1}{2}n^{-3-1/4} < n^{-1/4}.$$

With similar arguments, one can show that the probability that $G_{n,p}$ contains three, four, ... remote pairs sharing a vertice is $o(1)$. Thus, the r^{th} factorial moment of X is within $o(1)$ of the expected number of ordered r-tuples of disjoint remote pairs. This implies that

$$\mathbb{E}_r(X) = (n)_r 2^{-r} F_r(1+o(1)) + o(1), \tag{4.18}$$

where F_r is the probability that a fixed r-tuple $\tau := (x_1, \ldots, x_r)$ of vertices consists of vertices remote from other vertices, i.e. for every $x_i \in \tau$ there exists y_i so that (x_i, y_i) is a remote pair. The factor 2^{-r} is there as not to double count pairs of remote vertices.

We will now define some sets which we shall need in consequence for conditional probabilities. For $1 \leq i \leq r$ write

$$\begin{gathered}A_i = \Gamma_{d-1}(x_i) - \bigcup_{j \neq i} N_{d-1}(x_j), \\ T = \bigcap_{i \neq j}\left(N_{d-1}(x_i) \cap N_{d-1}(x_j)\right) = \bigcap_i N_{d-1}(x_i), \\ S = V(G) - \bigcup_{j=1}^{r} N_{d-1}(x_j), \\ S' = S - \Gamma(T), \\ a_i = |A_i|, \quad s = |S|, \quad s' = |S'| \quad \text{and} \quad t = |T|.\end{gathered}$$

Let $K = \max\{r+2, e^7\}$. Then, using Lemma 4.7, for sufficiently large n, with probability at least $1 - n^{-K}$, we have

$$\begin{aligned}|a_i - (pn)^{d-1}| \leq \delta_{d-a}(pn)^{d-1} + 8rp^{2d-2}n^{2d-3} = \\ = (pn)^{d-1}\left(\delta_{d-a} + 8r\big(\log(n^2/c)\big)/(pn)\right) =: \delta(pn)^{d-1}\end{aligned} \tag{4.19}$$

and

$$n \geq s \geq s' \geq n - 8r^2p^{2d-1}n^{2d-2} =: (1-\varepsilon)n. \tag{4.20}$$

To understand the inequalities in (4.20), observe that $|V(G) - \bigcup_{j=1}^{r} N_{d-1}(x_j)| \sim n - r(pn)^{d-1}$. Because $(pn)^{d-1} \sim 2^{\frac{\log n}{p}} \ll n$, $|S| \sim n$. The remaining subtraction then follows with Lemma 4.7 2.

Note that the defined functions δ and ε satisfy

$$\delta \log n \to 0 \qquad \text{and} \qquad \varepsilon \to 0.$$

To show this for (4.19), note that $pn/(\log n)^3 \to \infty$, so $\log n \log(n^2/c)/(pn) \to 0$; the second part of the estimate is similar. For n large,

$$\delta_{d-1} \leq 3 \sum_{l=1}^{d-1} (\alpha_l + \beta_l + \gamma_l) \leq 4(\alpha_1 + \beta_{d-1} + \gamma_{d-1}) =$$

$$= 4 \left[K \left(\frac{\log n}{pn} \right)^{1/2} + p^{d-1} n^{d-2} + 2p^{d-2} n^{d-3} \right] \leq$$

$$\leq 4 \left[K \left(\frac{\log n}{pn} \right)^{1/2} + \frac{3 \log n}{pn} + \frac{6 \log n}{(pn)^2} \right].$$

Likewise, $\varepsilon \to 0$, because

$$p^{2d-1} n^{2d-3} = \frac{(p^d n^{d-1})^2}{pn} < \frac{(3 \log n)^2}{pn} \to 0.$$

Let us assume that A_i, S, and S' are fixed (satisfying (4.19) and (4.20)), and let us denote the probability conditional on this particular choice of these sets by $\tilde{\mathbb{P}}(.)$. In order to estimate F_r, we shall estimate the conditional probability

$$Q_r := \tilde{\mathbb{P}}(\forall x_i \in \tau : \exists y_i \text{ so that } (x_i, y_i) \text{ is a remote pair.}).$$

Set

$$R_r = \tilde{\mathbb{P}}(\forall i, 1 \leq i \leq r : \exists y_i \in S \text{ not joined to } A_i) \tag{4.21}$$

and

$$R'_r = \tilde{\mathbb{P}}(\forall i, 1 \leq i \leq r : \exists y_i \in S' \text{ not joined to } A_i). \tag{4.22}$$

$R'_r \leq Q_r \leq R_r$. Why? Clearly, the property described in (4.22) implies that τ consists of remote vertices. Conversely, x_i being a remote vertice means that there is a y not in $N_d(x_i)$; this means there is a vertex in $V(G) \setminus N_{d-1}(x_i)$ not joined to $\Gamma_{d-1}(x_i)$. Note that $|S|$, $|S'|$ and $N_d(x_i)$ are basically the same size — namely, about n. Because the probability that x_i and x_j, $i \neq j$, share a remote vertex tends to zero, it follows that $\mathbb{P}(y \in \cap_{j \neq i} N_d(x_j) \mid y \notin N_d(x_i)) \to 1$. However, $\mathbb{P}(y \in \cap_{j \neq i} N_{d-1}(x_j) \mid y \notin N_d(x_i)) \leq \mathbb{P}(y \notin S') \to 0$. It follows that $\mathbb{P}(y \in \cap_{j \neq i} \Gamma_d(x_j) \mid y \notin N_d(x_i)) \to 1$. As $\mathbb{P}(y \in \cap_{j \neq i} \Gamma_d(x_j) \mid y \notin N_d(x_i)) \leq \mathbb{P}(y \in S \mid y \notin N_d(x_i))$, so the second part of the inequality follows.

It holds

$$R_r = \prod_{i=1}^{r} \left[1 - (1 - (1-p)^{a_i})^s \right],$$

and R'_r is given in a similar way. Let us estimate R_r from above:

$$(1-p)^{a_i} \le e^{-pa_i} \le e^{-p^d n^{d-1}(1-\delta)} = \left(\frac{c}{n^2}\right)(1+o(1)).$$

Thus, $R_r \le (c/n)^r(1+o(1))$. Similarly, let us estimate R'_r from below:

$$(1-p)^{a_i} \ge e^{-pa_i(1+p)} \ge e^{p^d n^{d-1}(1-p)(1+\delta)} = \left(\frac{c}{n^2}\right)(1+o(1)),$$

$$(1-(1-p)^{a_i})^{s'} \le 1-(\frac{s'c}{n^2})(1+o(1)) = 1-(\frac{c}{n})+o\left(\frac{1}{n}\right).$$

It follows that $Q_r = (c/n)^r(1+o(1))$.

Remember that Q_r is the probability of τ consisting of remote vertices under the condition that A_i, S and S' are given. These sets satisfy (4.19) and (4.20) with probability $1-n^{-K}$. Note that for conditional probabilities $\mathbb{P}(A|B)$, these two simple relations hold:

$$\mathbb{P}(B)\mathbb{P}(A \mid B) \le \mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A \mid B) + \mathbb{P}(B^c)\mathbb{P}(A \mid B^c) \le \mathbb{P}(B)\mathbb{P}(A \mid B) + \mathbb{P}(B^c),$$

which, in our case, transfers to

$$(1-n^{-K})Q_r \le F_r \le (1-n^{-K})Q_r + n^{-K},$$

so $F_r = (c/n)^r(1+o(1))$, which, with (4.18), implies that

$$\mathbb{E}_r(X) = n^r 2^{-r}\left(\frac{c}{n}\right)^r(1+o(1)) + o(1) = \left(\frac{c}{2}\right)^r + o(1). \tag{4.23}$$

Again, we use Theorem 4.9 and conclude that $X \xrightarrow{d} \mathrm{Po}_{c/2}$. In particular, we have

$$\mathbb{P}(\operatorname{diam} G_{n,p} \le d) = \mathbb{P}(X=0) \to e^{-c/2}. \tag{4.24}$$

Now our proof is not hard to complete: If $d=2$, then it follows that

$$\mathbb{P}(\operatorname{diam} G_{n,p} \le 1) = \mathbb{P}(G_{n,p} = K_n) = p^{\binom{n}{2}} \to 0.$$

Note that the property of having diameter at least k is monotone, so by Lemma 4.1 if $0 < p_1 < p_2 < 1$, then $\mathbb{P}(\operatorname{diam} G_{n,p_1} \le k) \le \mathbb{P}(\operatorname{diam} G_{n,p_2} \le k)$. Define $c_1 = c_1(n)$ by

$$p^{d-1}n^{d-2} = \log(n^2/c_1),$$

which is equivalent to $\frac{1}{pn}\log\frac{n^2}{c} = \log\frac{n^2}{c_1}$. As this tends to 0, $c_1 \sim n^2 \to \infty$, so (4.24) implies that

$$\mathbb{P}(\operatorname{diam} G_{n,p} \le d-1) \to 0. \tag{4.25}$$

Equivalently, define $c_2 = c_2(n)$ by

$$p^{d+1}n^d = \log(n^2/c_2),$$

by similar thoughts we obtain $c_2 \to 0$, so (4.24) yields

$$\mathbb{P}(\operatorname{diam} G_{n,p} \le d+1) \to 1. \tag{4.26}$$

Combining (4.24), (4.25) and (4.26), the theorem is proved. □

5 Small-World Networks

5.1 The Basic Idea

More than ten years have passed since Steven Strogatz and Duncan Watts introduced their model of the small-world effect in their ground-breaking paper [45]. This paper is especially important to the theory of complex networks because it was one of the first that modelled a large-scale network by a random graph defined by simple rules [10]. They were looking for a network that would have the following two attributes:

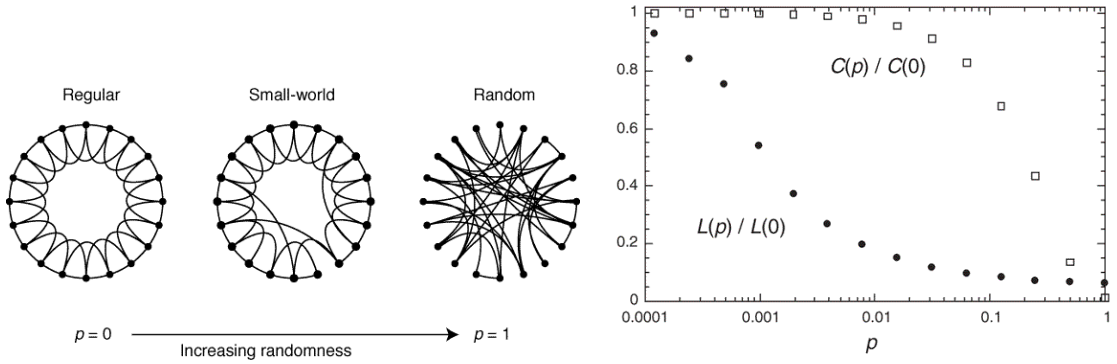
i) It should have a high clustering coefficient, just as graphs on a regular lattice expose where each vertice is connected to it's k nearest neighbors.

ii) It should also have a small diameter, or at least a small mean distance — small in this sense being comparable to the corresponding random graph.

In principal, what they wanted was a network they could fine-tune between a regular lattice and a random graph. They claimed that many "real-world" networks from biology, technology and the social sciences were exactly this: Neither completely organized, nor completely random, they were "somewhere in between".

The original small-world network was defined as follows: n nodes are arranged on a one-dimensional, periodic lattice (ring lattice). There are k (undirected) edges per vertice. Initially, each vertice is thus connected to it's k left and it's k right neighbors. We start from a $2k$ regular graph, so there are nk edges.

With probability p each edge is then rewired, meaning that each edge is deleted with probability p, and then afterwards the number of deleted edges is again added to the graph at random [10]. See Figure 5.1(a) for a visual explanation.



(a) The small-world model

(b) Numerical Results for C and $\bar{l}$

Figure 5.1: Figures explaining the small-world model [45]

For $p = 0$, the model is completely regular. The clustering coefficient is $C = \frac{3(k-1)}{2(2k-1)}$, and the mean distance between two vertices is $\bar{l} = \frac{n}{2k}$. For $p = 1$, the model is a random graph $G_{n,kn}$ with approximate mean distance $\tilde{l} = O(\ln(n))$ and $C \sim 0$. We set $n \gg k \gg \ln(n) \gg 1$, so that the random graph will be connected. (Compare with theorem 4.2.)

It was surprising in the numerical results that followed that for very low values of p, the clustering coefficient did not change much while the average distance between randomly chosen pairs of vertices (also called the *average path length* in the theory of complex networks) decreased dramatically. This means that only few short-cuts are needed to make a regular lattice "small-world." See figure 5.1(b).

Watts and Strogatz tested the model on several "real" networks (the graph of film actors, the graph of the power grid, and the neural network of *Caenorhabditis Elegans*, a very much studied worm) to find that the fit was quite good.

5.2 First Analytical Results

5.2.1 A Toy Small-World Network

Shortly after the initial paper [45], scientists tried to find analytical solutions to the proposed model. This proved more difficult than one might think; it was necessary to simplify the model.

One of the first who found analytical solutions to simplified models were Dorogovtsev and Mendes [16]. They consider a so-called *toy small-world network* which is described as follows:

In the first version of this model, n nodes are arranged on a circle, and a node is connected to its right neighbor by a (directed) arc of unit length:

$$(v_i, v_{i+1 \bmod n}) \in E \quad \forall i = 0 \dots n-1.$$

In the center of the circle, there is another vertex, the *hub* of the graph. Each vertex is connected to this hub with probability p, with undirected connections of length $1/2$. This means that there exist paths of length 1 between all vertices that have a link with the hub. See figure 5.2(a). Note that a path from v_i to v_j will never include more than one short-cut, and that, for $p = 1$, $\bar{l} = 1$. This differs considerably from the initial model. To simplify notation, we will now only consider cases that do not pass from node $n-1$ to node 0.

To obtain $\bar{l}$, we first consider

$$\mathbb{P}(l, k) := \mathbb{P}\{\text{the distance between } v_i \text{ and } v_j \text{ is } l \text{ if } |i-j| = k\},$$

which we can derive as follows: For $l < k$, there need to be two "half" short-cuts separating vertices u_s, u_t that lie $k-l$ apart, $|s-t| = k-l$. There are l possibilities of picking these two vertices, and the probability of these two vertices both having a shortcut is p^2. u_s and u_t must also be the vertices that give the smallest path length, so they are the vertices whose connections to the hub "save" the most space. This means that for all other vertices "outside" of the shortcut, it holds that they are not connected to the hub, which holds

with probability $(1-p)^{l-1}$. Remember that $\sum_{l=1}^{k} \mathbb{P}(k,l) = 1$. Thus,

$$\mathbb{P}(l < k, k) = lp^2(1-p)^{l-1}, \tag{5.1}$$

$$\mathbb{P}(l = k, k) = 1 - p^2 \sum_{i=1}^{k-1} i(1-p)^{i-1}. \tag{5.2}$$

Note that $\mathbb{P}(l < k, k)$ does not depend on k.

Looking for the distribution of shortest path lengths, we define $\mathbb{P}(l)$ to be the probability that a path between any given pair of vertices has lenght l. Note that

$$\mathbb{P}(l) = \frac{1}{n-1} \sum_{k=1}^{n-1} \mathbb{P}(l,k) = \frac{1}{n-1} \sum_{k=l}^{n-1} \mathbb{P}(l,k),$$

which, with (5.1) and (5.2) gives

$$\mathbb{P}(l) = \frac{1}{n-1} \left(1 - (l-1)p - l(n-1-l)p^2\right) (1-p)^{l-1}. \tag{5.3}$$

Note that for $p \to 0$, $\mathbb{P}(l) \to \frac{1}{n-1}$ for any $l \leq n-1$, and if $p \to 1$ then $\mathbb{P}(l) \to \delta_{l,1}$, just as we would expect.

Next, we want to calculate the average shortest distance $\bar{l}$, defined by

$$\bar{l} := \sum_{l=1}^{n-1} l\, \mathbb{P}(l).$$

With (5.3), this gives

$$\bar{l} = \frac{1}{n-1} \left(\frac{2-p}{p} n - \frac{3}{p^2} + \frac{2}{p} + \frac{(1-p)^n}{p} \left(n - 2 + \frac{3}{p} \right) \right).$$

Again, for this term, it holds $\bar{l}(p \to 0) \to n/2$ and $\bar{l}(p \to 1) \to 1$.

Setting $\rho := pn$ and $z := l/n$, we define

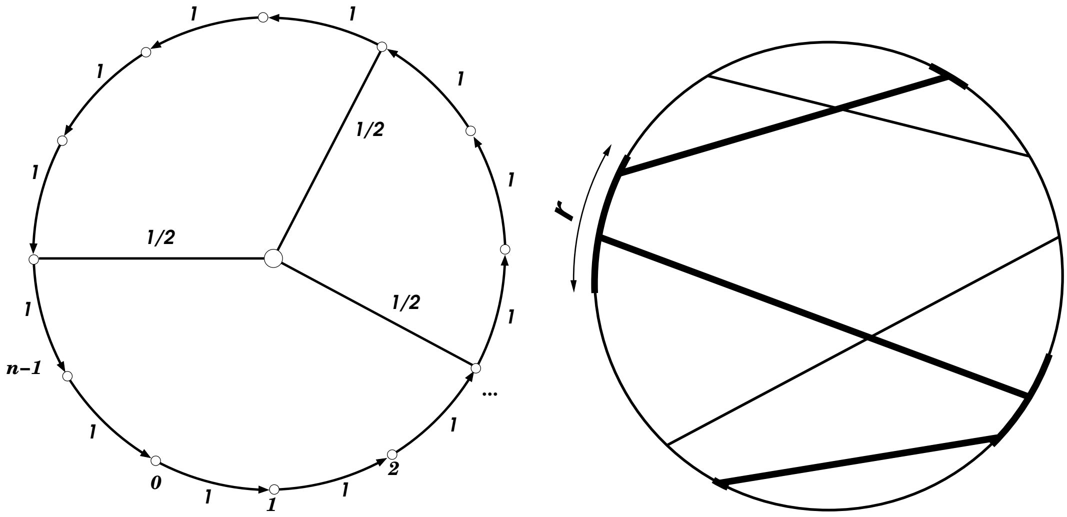
$$n\mathbb{P}(l) := Q(z,\rho) = \left(1 + \rho z + \rho^2 z(1-z)\right) e^{-\rho z}$$

for $0 \leq z \leq 1$. This is the scaling function; it's average value is

$$\frac{\bar{l}}{n} := \bar{z} = \frac{1}{\rho^2} \left(2\rho - 3 + (\rho+3)e^{-\rho}\right).$$

For $\rho \to 0$ as $n \to \infty$, $\bar{z} \to 1/2 - \rho^2/24$, and for $\rho \gg 1$, $\bar{z} \to 2/\rho$. The plots of these functions shows that the results are close to previous simulations. For example, the plot of $\bar{l}/n$ resembles results from [45].

The expressions for $\mathbb{P}(l)$ and the like are more tedious in second version of the model, the undirected case. Their derivation is also more complicated because the possibilities of initially going in what would intuitively be the "wrong" direction have to be considered. It is worth noting, however, that the scaling function of the undirected model $\tilde{Q}(z,\rho) = 2Q(2z,\rho)$. Thus both models qualitatively give the same results. Also consider that $\bar{l}(p = 0) = 1/2$, while in the directed model, for the mean distance it holds $\bar{l}_{\text{undir}}(p = 0) = 1/4$.



(a) The toy small-world on n nodes. (b) Mean Field approach to the small world model.

Figure 5.2: Figures of the two small world models presented in this section.

5.2.2 The Mean-Field Solution

Newman, Moore and Watts published a paper that followed shortly after the toy small-world which treated the mean-field solution of the small-world network model [35]. (Compare to section 3.1.1.) They slightly modify the original model by Watts and Strogatz, but much less than in the toy small-world. In this version of the model, we again have a circle lattice and each vertex is connected to its k nearest neighbors on both sides with undirected edges. We have nk edges. For each edge of the lattice a random edge is added with probability p.

In this mean-field approximation, the distribution of path lengths is calculated. Quantities over realizations of the model are given by their averages. Newman *et al.* first consider the continuum version of the model, seeing the underlying lattice of the system as continuous. In this version, shortcuts have length zero. The assumption here is that, if the density of shortcuts is low, the discrete and continuous models are equivalent, and thus the solution of this system will be a solution of the small-world model with general k. See figure 5.2(b).

We now consider a neighborhood ρ of radius r centered around a randomly chosen vertice v_i:

$$\rho(r) := \{v_j : \text{ There exists a path of length } r \text{ or less between } v_i \text{ and } v_j\}$$

We let $m(r)$ be the number (on average!) of nodes *not* belonging to ρ, and $s(r)$ be the number of gaps around the lattice among which those $m(r)$ nodes are divided. This is the same as the number of clusters of sites in $\rho(r)$. Note that, for the continuum model, $m(r) \in \mathbb{R}$, $s(r) \in \mathbb{R}$. We define the rescaled variables

$$\mu(r) = \frac{m(r)}{n}, \qquad \nu(r) = \frac{s(r)}{n},$$

and begin looking for differential equations that our system will satisfy. This is possible because we are assuming the system to be continuous.

When r increases, $m(r)$ decreases by the number of borders of growing clusters (of which there are $2s$) times the number of connections on the lattice of each vertex k. So

$$\frac{dm}{dr} = -2kr \qquad \text{or} \qquad \frac{d\mu}{dr} = -2k\nu,$$

which holds true for all n and p.

The number of gaps changes for two reasons:

i) For every shortcut found that leads outside of ρ, there is a new gap. We define $\xi = 1/(kp)$, so n/ξ is the average number of shortcuts in the graph, and the density of the ends of shortcuts is $2/\xi$. For growing r, shortcuts are found at a rate $4ks/\xi$. The probability that a new found shortcut falls into a gap around the ring is m/n. This means that clusters are created at a rate of $4kms/\xi n$.

ii) The number of clusters grows smaller when two clusters merge. This happens if, when r changes to $r + \Delta r$, there are gaps smaller than $2k\delta r$. Distributing m sites on s gaps, we see that the size of the gaps has the same probability distribution as the distribution of the smallest of $s-1$ uniformly distributed random numbers x between 0 and m, namely

$$p(x) = \frac{s-1}{m}\left(1 - \frac{x}{m}\right)^{s-2}.$$

It follows that the probability of one particular gap being smaller than $2k\Delta r$ is $1 - (1 - 2k\Delta r/m)^{s-1}$; for small Δr, this tends to $2k(s-1)\Delta r/m$. This is valid for s gaps.

Finally, we get differential equations for s and ν:

$$\frac{ds}{dr} = \frac{4kms}{\xi n} - \frac{2ks(s-1)}{m}$$

and

$$\frac{d\nu}{dr} = \frac{4k\mu\nu}{\xi} - \frac{2k\nu(\nu - \frac{1}{n})}{\mu}.$$

Newman *et al.* state that these equations will only be exact when the average values $\mu(r)$ and $\nu(r)$ are quite accurate, meaning that the real distribution is very close to the center, which will be the case when there are either almost no shortcuts, i.e. $n \ll \xi$ or when the density of shortcuts is very high, $n \gg \xi$.

By some transformations, the solutions to μ and ν can be found, where $\mu = \mu(r, \xi, n)$. From this, the average vertex separation $\bar{l}$ can be derived, We find that

$$\bar{l} = \frac{n}{k} h(\frac{n}{\xi}),$$

where $h(x)$ is a universal scaling function which is given by

$$h(x) = \frac{1}{2\sqrt{x^2 + 2x}} \tanh^{-1} \frac{1}{\sqrt{x^2 + 2x}}. \tag{5.4}$$

Numerical simulations (with $k = 1$) show a good fit for the average path length for large and small values of n/ξ, but there are some problems when $\xi \sim n$. Note that, apart from the factor of $1/4$, $h(x)$ is the fraction by which the average path length on a small-world graph is reduced if the graph has x short cuts.

5.3 Some Rigorous Approaches

As the interest in complex networks increased, several models — among them the small-world model — have been "discovered" and adapted by mathematicians. In this section, I will describe two of these models, and give quick résumés of their proofs.

5.3.1 A Markov Chain Small-World Model

Catral, Neumann and Xu published a paper describing the matrix analysis of a Markov chain small-world model [14]. The principle idea behind this model is lovely, finding average (and maximum) mean pathlength by considering mean first passage times. However, the proof is very technical, and the argumentation of the results might be a little disappointing.

The model starts from n nodes placed on a ring lattice where every node is connected to its nearest neighbors by a directed edge. Starting from any vertice v_j on the lattice at time t, the initial random walk is as one would expect: with probability $p = 1/2$, the position at $t+1$ is v_{j+1} and, with equal probability, it is v_{j-1}. This process is called *completely local*, because there are no direct transitions between non-neighboring vertices. It can be represented by an ergodic Markov chain with transition matrix $\mathbf{P}_0$:

$$\mathbf{P}_0 = \begin{pmatrix} 0 & 1/2 & 0 & \dots & 0 & 1/2 \\ 1/2 & 0 & 1/2 & \ddots & & 0 \\ 0 & 1/2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1/2 & 0 \\ 0 & & \ddots & 1/2 & 0 & 1/2 \\ 1/2 & 0 & \dots & 0 & 1/2 & 0 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

This initial process is now modified; from vertex v_j, we introduce the transition probability ε. With probability $0 \leq \varepsilon \leq 1/(n-2)$, the next step away from v_j is v_i, $\forall i \neq j \pm 1$. These random jumps are the shortcuts in this model. With probability $\delta = [1-(n-1)\varepsilon]/2$, the random walk proceeds to one of v_j's neighbors. Thus, the transition matrix $\mathbf{P}_\varepsilon$ is

$$\mathbf{P}_\varepsilon = \begin{pmatrix} \varepsilon & \delta & \varepsilon & \dots & \varepsilon & \delta \\ \delta & \varepsilon & \delta & \ddots & & \varepsilon \\ \varepsilon & \delta & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \delta & \varepsilon \\ \varepsilon & & \ddots & \delta & \varepsilon & \delta \\ \delta & \varepsilon & \dots & \varepsilon & \delta & \varepsilon \end{pmatrix} \in \mathbb{R}^{n\times n}. \tag{5.5}$$

Note that for $\varepsilon = \frac{1}{n}$, the network is completely random, each node is an equal likely target. This is also called the *completely global* case. For $\varepsilon = 0$, the network is again completely local.

The properties we are interested in, in connection with ε, are the maximum and average mean first passage time of $\mathbf{P}_\varepsilon$. Catral *et al.* consider what they define as the reduction ratio, i.e. by which factor does the mean (or maximum) passage time change when with probability ε short-cuts are introduced to the system?

It is sufficient to consider the mean first passage times from nodes $2, \ldots, n$ to node 1. We redefine our process to be able to use Theorem 3.3: We consider the modified matrix $\tilde{\mathbf{P}}_\varepsilon$, where $p_{1j} = \delta_{1j}$, where δ_{1j} is the Kronecker delta: A reducible Markov chain is formed, where the only absorbing state is 1. Counting the mean passage to to vertice 1 is now equal to counting the time until absorption, and these times are given by the vector

$$z_\varepsilon = (\mathbf{I} - \hat{\mathbf{P}}_\varepsilon)^{-1}\,\mathbf{e} \;\in \mathbb{R}^{n-1}, \tag{5.6}$$

where $\hat{\mathbf{P}}_\varepsilon$ is the submatrix obtained from $\mathbf{P}_\varepsilon$ (or $\tilde{\mathbf{P}}$) by deleting it's first row and column. $\mathbf{e}$ denotes a column vector of ones. The average mean first passage time is then defined as

$$\bar{z}_\varepsilon = \frac{1}{n-1}\sum_{i=1}^{n-1}(z_\varepsilon)_i.$$

For both the "extreme" cases, $\varepsilon = 0$ and $\varepsilon = 1/n$, we have

$$(z_0)_i = i(n-i) \qquad \text{and} \qquad (z_{1/n})_i = n,$$

for $i = 1, \ldots, n-1$. Thus, the maximum and average mean first passage times are given by

$$\max_{1\le i\le n-1}(z_0)_i = k(n-k) \qquad \text{and} \qquad \max_{1\le i\le n-1}(z_{1/n})_i = n, \tag{5.7}$$

where $k = \lfloor (n+1)/2 \rfloor$ and

$$\bar{z}_0 = \frac{n(n+1)}{6} \qquad \text{and} \qquad \bar{z}_{1/n} = n.$$

The *reduction ratios* of the maximum mean first passage time f and the average mean first passage time g are defined as follows:

$$f := f(\varepsilon) := \frac{\max_{1\le i\le n-1}(z_\varepsilon)_i}{\max_{1\le i\le n-1}(z_0)_i} \qquad \text{and} \qquad g := g(\varepsilon) := \frac{\bar{z}_\varepsilon}{\bar{z}_0}. \tag{5.8}$$

This means f and g quantify the changes on the average and maximum mean first passage in comparison to the completely local variant — Catral *et al.* say f and g measure the *degree of separation*. Note that $f|_{\varepsilon=1/n} = n/\lfloor k(n-k)\rfloor$ and $g|_{\varepsilon=1/n} = 6/(n+1)$; thus, the degree of separation is reduced from 1 to $O(1/n)$.

Let us define

$$r := \frac{1-n\varepsilon}{1+\sqrt{1-(1-n\varepsilon)^2}}. \tag{5.9}$$

Note that $-1 \le r \le 1$ and for $0 \le \varepsilon \le 1/(n-2)$, r is well defined. Using r, we can now state some of the theorems of this paper.

Theorem 5.1. *Consider the ergodic Markov chain of the ring network on n vertices whose transition matrix $\mathbf{P}_\varepsilon$ is given by (5.5). $z_\varepsilon \in \mathbb{R}^{n-1}$ is the vector of mean first passage times given in (5.6) and let r be as given in (5.9). Suppose that $0 \le \varepsilon \le 1/(n-2)$. Then*

$$(z_\varepsilon)_i = \frac{n(1+r^2)(1-r^i)(1-r^{n-1})}{(1-r^2)(1-r^n)}, \quad i = 1, \ldots, n-1.$$

Theorem 5.1 is then used to obtain the following two theorems:

Theorem 5.2. *Under the assumptions of Theorem 5.1, if we further assume that* $0 < \varepsilon \leq 1/n$*, then*

$$\max_{1\leq i\leq n-1}(z_\varepsilon)_i = \frac{n(1+r^2)(1-r^k)(1-r^{n-k})}{(1-r^2)(1-r^n)},$$

where $k = \lfloor (n+1)/2 \rfloor$*. If we further assume that* $1/n \leq \varepsilon \leq 1/(n-2)$*, then*

$$\max_{1\leq i\leq n-1}(z_\varepsilon)_i = \frac{n(1+r^2)(1-r^{n-1})}{(1-r)(1-r^n)}.$$

Theorem 5.3. *Under the assumptions of Theorem 5.1, the average mean first passage time as defined in (5.7) is given by*

$$\bar{z}_\varepsilon = \frac{n(1+r^2)(n(1-r)(1+r^n)-(1+r)(1-r^n))}{(n-1)(1-r)(1-r^2)(1-r^n)}.$$

From these theorems, it then follows:

Theorem 5.4. *Under the assumptions of Theorem 5.1, the ratios* f *as in (5.8) is given by*

$$f = \begin{cases} \frac{n(1+r^2)(1-r^k)(1-r^{n-k})}{k(n-k)(1-r^2)(1-r^n)}, & \text{if } 0 < \varepsilon \leq 1/n; \\ \frac{n(1+r^2)(1-r^{n-1})}{k(n-k)(1+r)(1-r^n)}, & \text{if } 1/n \leq \varepsilon \leq 1/(n-2). \end{cases} \tag{5.10}$$

where $k = \lfloor (n+1)/2 \rfloor$*.*

Theorem 5.5. *Under the assumptions of Theorem 5.1, the ratio* g *as in (5.8) is given by*

$$g = \frac{6}{n^2-1}\left(\frac{n(1+r)^2(1+r^n)}{(1-r^2)(1-r^n)} - \frac{1+r^2}{(1-r)^2}\right). \tag{5.11}$$

Let us denote by ε_c the smallest value of ε where such that $f|_{\varepsilon=\varepsilon_c} \leq 1/2$. The last theorem we shall state is the following:

Theorem 5.6. *Under the assumptions of Theorem 5.1, let* f *be the reduction ratio in the maximum mean first passage time, which is given in (5.10). Then*

$$\varepsilon_c = \frac{32}{n^3} + O(n^{-4}).$$

Interestingly, Catral *et al.* compared the reduction ratio g to $h(W)$ as defined in (5.4). Remember that $h(W)$ was defined as a function of the expected number of shortcuts (W) used in the model defined in subsection 5.2.2. Here, the expected number of shortcuts in the random walk is $\varepsilon(n-2)\bar{z}_\varepsilon$. Plotting $h(x)$ against f and g, it can be seen that the fit is quite good. Considering this, and that Catral *et al.* [35] stated that $\bar{l} \approx (n/4)h(np)$ when $k = 1$, and the fact that $\max_{1\leq i\leq n-1}(z_0)_i \approx n^2/4$ for large n, it follows that

$$\frac{n}{4}h(W) \approx \frac{n}{4}f \approx \frac{1}{n}\max_{1\leq i\leq n-1}(z_\varepsilon)_i,$$

so we can observe that both f and g seem to scale like $h(W)$ as described in [35]. This, and some numerical simulations on the clustering coefficient, appear to be the main results of

this paper. Though the idea of modelling the small-world with a Markov chain random walk is very appealing, the main result is not very relevant to the theory of complex networks.

The stated theorems can be proved in order of their appearance via matrix analysis. In fact, the proofs are all quite technical: $\mathbb{I} - \hat{\mathbb{P}}_\varepsilon$ is calculated, which renders proof of Theorem 5.1. After some technical calculations (considering monotonicity and $\min r$ and the like) the other theorems follow easily in the same order that they have been stated.

5.3.2 Spatial Random Graphs

Another paper on the subject considering so-called *spatial or geometric random graphs* was published by Ganesh and Xue [24]. It holds only two theorems, both on connectivity of the described graph process, one also bounds the diameter. Here, I will describe both possible variants of their model and state both theorems. I will then give accounts of the proof of the first theorem. The proof of Theorem 5.8 follows with the Stein-Chen Method, see for example [6].

In both models we consider a sequence of undirected random graphs G_n on $n \in \mathbb{N}$ vertices. The concept of geometric random graphs is to associate the nodes with coordinates in a Euclidean space, the probability of an edge between two nodes in this random graph is then some function of the distance between these nodes. The node locations are also the result of some random process. In our case, node positions are modelled by a stochastic point process, e.g. independent and identically distributed uniformly on a square. We connect nearest neighbors up to a certain number to be defined, where a nearest neighbor is defined to have the smallest distance in the Euclidean space. Also, two vertices can be joined by randomly added shortcuts.

In *Model A*, each of the n nodes is connected to its m_n nearest neighbors. Note that this relation need not be symmetric. A shortcut is present between each pair of nodes with probability p_n, independently of all other edges. Multiple edges are replaced by simple ones.

Note that the spatial proximity in this model is of no importance for the proof. Nodes could be connected to any other number m_n of nodes, and the proof would still hold.

In *Model B*, n nodes are located u.a.r. on the torus obtained by identifying the opposite sides of the square of area n centred at the origin with each other, i.e. the interval $[-\sqrt{n}/2, \sqrt{n}/2]$ is "folded." Each node is connected to all nodes within a circle of radius r_n centered at the node itself. Additional shortcuts are present between each pair of nodes with probability p_n, independent of all other edges.

The main difference between Model A and Model B is that in Model A, the degree of a vertice is bounded from below by m_n, while in Model B, there exists no such bound. This improves connectivity of Model A a great deal.

Let us denote by C_n the event that G_n is connected, and by D_n the diameter of G_n. If G_n is not connected, then $D_n = \infty$.

Theorem 5.7. *Suppose that the sequences m_n and p_n are such that*

$$\frac{m_n}{n} \to 0 \quad \text{as } n \to \infty, \qquad \text{and}$$
$$(m_n + 1) n p_n \geq 2(1+\delta) \log \frac{n}{m_n + 1} \tag{5.12}$$

for some $\delta > 0$ and all sufficiently large n. Then, for the random graph described in Model A, with parameters m_n and p_n, we have

$$\lim_{n\to\infty} \mathbb{P}(C_n) = 1, \qquad \lim_{n\to\infty} \mathbb{P}\left(D_n \le 7\left(\log \frac{n}{m_n+1} + 1\right)\right) = 1. \tag{5.13}$$

Conversely, if

$$\frac{m_n}{n} \to 0 \quad \text{as } n \to \infty, \quad \text{and}$$
$$(m_n+1)np_n < (1-\delta)\left(\frac{m_n+1}{m_n+2}\right)^2 \log \frac{n}{m_n+1} \tag{5.14}$$

for some $\delta > 0$ and infinitely many n, then there is a sequence of node locations such that $\lim_{n\to\infty} \mathbb{P}(C_n) = 0$.

Note that a node will have m_n nearest neighbors, and on average $p_n(n-1)$ other connecting edges via shortcuts. This means that the conditions of Theorem 5.7 state that, to have a connected graph, the product of these quantities should be about $\log n$.

Theorem 5.8. *Supppose that the sequences r_n and p_n are such that*

$$\pi r_n^2 + np_n = \log n + c_n \quad \text{and} \quad \lim_{n\to\infty} c_n = c. \tag{5.15}$$

Then, the number of isolated vertices in the random graph generated by Model B with parameters r_n and p_n converges in distribution to a Poisson random variable with mean e^{-c}. Moreover, if $\lim_{n\to\infty} c_n = -\infty$, then the random graph generated by Model B is disconnected whp.

Note that a vertice has mean degree $\pi r_n^2 + np_n$. Theorem 5.8 states that if the mean degree is much smaller than $\log n$, the graph will be disconnected, as is also the case for "classical" random graphs (see Theorem 4.2).

Outline of proof of Theorem 5.7: Because of the way Model A is constructed, note that every vertex belongs to a component of size at least $m_n + 1$. If the graph is disconnected, there must thus be an isolated component of at least this size. The idea of the proof is to see these components — here, we shall call them clusters — as new entities that must be connected for the graph as a whole to be connected. On these entities, we can use the theory of the Erdős-Rényi random graph models.

Let G be a graph fulfilling the assumptions in (5.12). We group the vertices of the graph G into disjoint "discs" A_k, $k = 1,2,\dots,K_n$; each A_k has between $m_n + 1$ and $2m_n + 1$ vertices, and for all A_k, it holds that the diameter is at most 6. The latter property of these groups is possible because they are formed by taking a cluster and possibly adding vertices from other clusters that intersect. In order to show that G is connected, it suffices to show that these clusters are connected via shortcuts.

We first replace clusters A_k larger than $m_n + 1$ with clusters $\tilde{A}_k$ of size $m_n + 1$, and then we replace these clusters $\tilde{A}_k$ with one node k. An edge is put between nodes k_1 and k_2 if there is at least one shortcut between A_{k_1} and A_{k_2}. We call this graph $\tilde{G}$. If $\tilde{G}$ is connected, then G is connected as well. $\tilde{G}$, however, turns out to be a classical Erdős-Rényi

graph with edge probability $\tilde{p}_n = 1 - (1 - p_n)^{(m_n+1)^2}$. With a few transformations, it then follows that

$$\tilde{p}_n \geq (1 + \delta') \frac{\log K_n}{K_n}$$

for any $0 < \delta' < \delta$ and n sufficiently large. Note that $K_n = O(n)$. By the results of Erdős and Rényi (Theorem 4.2), $\tilde{G}$ is connected, so G is connected as well. The second claim follows from [7][Theorem 10.17].

Vice versa, for the other part of the theorem, we consider a sequence n_k $k \in \mathbb{N}$, that satisfies (5.14) along the sequence n_k, m_{n_k}, p_{n_k} for some $\delta > 0$. First, it is shown that it is possible to partition the nodes into sets of size either $\tilde{m}_{n_k} + 1$ or $\tilde{m}_{n_k} + 2$ where $\tilde{m}_{n_k}$ satisfies $(\tilde{m}_{n_k} + 1)(m_{n_k} + 1) \to 1$ as $k \to \infty$.

Let us now consider a deterministic sequence of node configurations consisting of clusters $A_1, \ldots, A_{q_k}$. The sizes of the clusters are either $\tilde{m}_{n_k} + 1$ or $\tilde{m}_{n_k} + 2$, where $\tilde{m}_{n_k}$ is defined as above. Within the clusters, the nodes are within Euclidean distance ε_{n_k}, and any two vertices in distinct clusters are more than ε_{n_k} apart, $\varepsilon_{n_k} > 0$. Obviously, the edges connecting different clusters can only come from shortcuts.

The idea is, again, to replace clusters by single nodes, and leave the shortcuts as only edges of the modified graph. As the clusters are of different size, the probability of shortcuts between clusters is neither identical nor independent. This problem is solved by adding a pseudo-node to clusters of only $m_{n_k} + 1$ vertices that again has shortcut probability p_{n_k}. We can then consider the graph $\tilde{G}$ where each (modified or not) cluster A_i is substituted by a single node i; there is an edge between two nodes i and j only if there was a shortcut between (modified) A_i and A_j. The initial graph G is connected only if $\tilde{G}$ is connected as well. Again, $\tilde{G}$ is a classical random graph. After some calculations, (and bounding p_{n_k} by $\tilde{p}_{n_k}$) it can be seen that

$$q_k \tilde{p}_{n_k} \leq (1 - \delta') \log q_k$$

for any $0 < \delta' < \delta$ and all k sufficiently large. Note that $q_k \to \infty$ as $k \to \infty$ because $m_n/n \to 0$. It thus follows (again with Theorem 4.2) that $\tilde{G}$ is disconnected whp, so G is disconnected as well. □

6 Models with Preferential Attachment

6.1 The Preferential Attachment Model of Barabási and Albert

Another ground-breaking paper in network theory was published in 1999, when what would later be coined the "Barabási-Albert Model" was introduced [1]. It was one of the first models that took into account both network growth as well as preferential attachment. On the one hand, this means dealing with a network where the number of vertices, n, is not taken as given, but it is changing with time: We are considering a random graph process, differing from the "classical" random graph process because not only does the number of edges, but also the number of vertices changes with time.

On the other hand, preferential attachment means that the probability with which vertices coming into the system connect to old vertices is not uniform. If a vertex has many incident edges already (i.e. high degree), then the probability is higher of a new vertex connecting to this node than to a node with less incident edges. There are many catchy slogans that explain this phenomena, such as "popularity is attractive" or "the rich get richer."

Intuitively, one of the reasons seems to be that highly connected vertices are simply more *visible* than vertices that are not as connected, which implies that new vertices in the system tend to "find" them easier and thus rather attach to them. This seems like a valid explanation of how new sites link to older sites when thinking of web pages, for example. Redner noted in his paper on citation networks that important papers get cited again and again (and are thus found easier in bibliographies, which are often used as guidelines for further reading — this might make the paper in question be cited even more), while papers that are not very relevant are quickly forgotten [42]. The concept of preferential attachment also gives an explanation why there are (mostly few) vertices in every system that are highly connected, while most do not have very many connections.

In the paper that introduced the first preferential attachment model [1], the random graph process was defined as follows: Starting with an initial number m_0 of vertices, at each time step t another vertex is added to the model. This vertex sends m edges to m already existing vertices, the probability that a vertex i is chosen is proportional to the degree of i, i.e.

$$\Pi(i) = \frac{\mathrm{d}(i)}{\sum_j \mathrm{d}(j)}. \tag{6.1}$$

Thus, after t time steps, there are $m_0 + t$ vertices in the system, and mt edges. Numerical simulations published in this paper already showed the stationary scale-free property of the degree distribution in this system, i.e. $P(k) \propto k^{-\gamma_{\text{model}}}$, which seemed to have an exponent $\gamma_{\text{model}} = 2.9 \pm 0.1$. With numerical simulations, [1] also compaired how the degrees of given vertices changes over time, only to show what one would have expected: The "rich-get-richer" phenomenon holds in a system with preferential attachment, meaning that older vertices acquire new edges at the expense of younger ones.

Barabási and Albert state that the scale-free distribution occurring in this system is a result of both the growing nature of the model *and* of preferential attachment. In simulations, they showed that alternative models where either the growing nature of the model or the mechanism of preferential attachment was omitted did not tend to a scale-free degree distribution. The growing model which omitted preferential attachment — the probability of attaching to vertex i was identical, $\Pi(i) = 1/(m_0 + t - 1)$, for all i — lead to a degree distribution $P(k) \propto \exp(-\beta k)$. Conversely, starting a model with n vertices without edges and adding edges by picking a random vertex and connecting it to m of the remaining vertices with probability proportional to their degrees leads to a distribution that was scale-free at the beginning of the process, but turned out not to be stationary. Not having allowed double edges or loops, after about $n^2/2$ steps, the graph was complete.

6.2 First Calculations, Explanations, and Criticism

6.2.1 A Mean-Field Approach

Shortly after the appearance of [1], the first heuristic explanations why the Barabási-Albert model results in a scale-free degree distribution were published. Barabási and Albert themselves, together with Jeong, published a paper giving a mean field approach to the problem [3], as well as mean-field explanations as to why the preferential attachment and growing properties of the model are important. Again, the basic line of argument is to say that the change in vertex degree is continuous, so with (6.1), this change can be written as

$$\frac{\partial \,\mathrm{d}(i)}{\partial t} = A\Pi(i) = A\frac{\mathrm{d}(i)}{\sum_{j=1}^{m_0+t-1}\mathrm{d}(j)},$$

where it is shown that the constant $A = m$, so $\frac{\partial \,\mathrm{d}(i)}{\partial t} = \frac{\mathrm{d}(i)}{2t}$. With only few calculations it then follows that for the probability density it holds

$$P(k) = \frac{2m^2 t}{m_0 + t}\frac{1}{k^3},$$

i.e. $P(k) \sim Ak^{-\gamma}$ where $\gamma = 3$, and the probability does not depend on t. The result for the growing model without preferential attachment is shown in a similar manner. Chapter two of [17] gives a very intuitive explanation of this approach on the stated models: vertices are visualized as buckets and edges as units of water raining into them.

6.2.2 Linear and Non-linear Preferential Attachment

Krapivsky *et al.* examined the degree distribution of variations of the Barabási-Albert model [30] heuristically. In fact, they looked into different functions of preferential attachment: At each time step t, a new vertex is added to their model which attaches to exactly one older vertice. The probability of a vertex of degree k being chosen is proportional to k^β, $\beta \geq 0$. Because only one new edge attaches at each time step, some of the problems mentioned in section 6.2.3 are avoided. However, it is unclear from which graph the process starts from. The line of argumentation is as follows:

Let $N_k(t)$ be the average number of sites with degree k at time t. N_k changes with time as follows:

$$\frac{dN_k}{dt} = \frac{1}{M_\beta}\left((k-1)^\beta N_{k-1} - k^\beta N_k\right) + \delta_{k1},$$

where $M_\beta(t) = \sum_j j^\beta N_j(t)$: With probability $(k-1)^\beta / M_\beta$, a new link comes in from one of the nodes with $k-1$ edges. Equivalently, it is necessary to subtract the possibility of a node with k links becoming connected to the new vertex. The last term δ_{k1}, the Kronecker delta, is caused by the new node with only one edge.

With help of the low-order moments of the degree distribution, Krapivsky *et al.* infer that for $\beta = 1$, (what they call a *linear connection kernel*), $N_k(t) = tn_k$, where

$$n_k = \frac{4}{k(k+1)(k+2)}.$$

For the sublinear connection kernel, a similar relation is derived,

$$n_k = \frac{\mu}{k^\beta} \prod_{j=1}^{k} \left(1 + \frac{\mu}{j^\beta}\right)^{-1},$$

where $\mu = \sum_{k=2}^{\infty} \prod_{j=2}^{k} \left(1 + \frac{\mu}{j^\beta}\right)^{-1}$. Asymptotically, this gives a stretched exponential. It is important to note that for $0.8 \leq \beta \leq 1$, the dependence of n_k on β for $1 \leq k \leq 1000$ is not very strong, so that it is difficult to discriminate between different βs, and even between a stretched exponential and a power law, something that Redner [42] had already noted when analyzing citation networks.

For the superlinear case ($\beta > 1$), it is shown that a single dominant site links to almost every other site. In fact, for $\beta > 2$, there is a nonzero probability that the initial site is connected to every other site on the graph, i.e. that this process forms a star. For general $\beta > 1$, it is shown to hold that

$$N_k = J_k t^{k-(k-1)\beta}$$

with $J_k = \prod_{j=1}^{k-1} j^\beta / \left(1 + j(1-\beta)\right)$ as long as $k - (k-1)\beta > 0$, or $k < \beta/(\beta-1)$. For $k > \beta/(\beta-1)$, all but a negligible number of sites are associated with N_1. Note that in this case the distribution depends on t and thus also on the number of sites in the network, so it is not scale-free.

6.2.3 Some Problems with the Barabási-Albert Model

A topic as interesting as complex networks was bound to attract the attention of mathematicians sooner or later. Bollobás and Riordan gave an overview of which mathematical results were to be encountered in the theory of scale-free random graphs [10]. As the Barabási-Albert model was the first of its kind, and as it initiated so much following thought, it is surprising that it took some time until a criticism of this model was published, noting some important inconsistencies.

The problems begin when we start the graph process (see section 6.1): At time $t = 1$, how should it be chosen to which vertices the m edges attach to when all have degree zero? An idea to avoid this problem would be to start with a small graph G_0 instead of m_0 disconnected vertices; however, Bollobás *et al.* state that in this case the choice of G_0 is relevant in a non-trivial manner. For example, the maximum degree can change considerably depending on the choice of G_0.

Another problem arises for $m \geq 2$ with the rule of preferential attachment itself: At time $t+1$, we add the $t+1^{\text{th}}$ vertex to the system. This vertex must now connect to a set N_{t+1} of m earlier vertices. Thus, because we pick m edges, each vertex is chosen with probability

$$\mathbb{P}(i \in N_{t+1}) = m\frac{\mathrm{d}(i)}{\sum_j \mathrm{d}(j)}. \tag{6.2}$$

The description of the model does not include multiple edges, however, so picking one edge after the other independently is not possible. A full description of the model would give a precise specification of the distribution of N_{t+1}, meaning the probability that $N_{t+1} = S$ for the $\binom{t}{m}$ possible choices of sets S of earlier vertices.

Note that the distribution is not uniquely specified by only giving the probabilities at $i \in N_{t+1}$ for each vertex i: This gives only t marginal probabilities, while the distribution of N_{t+1} has $\binom{t}{m} - 1$ degrees of freedom. Bollobás and Riordan continue by proving a statement to emphasize that there are several models fitting the description by the original model from section 6.1. Briefly, it states that for any integer valued function $f(n)$ that obeys certain properties there is a random graph process $T^{(n)}$ that satisfies (6.2) with $m = 2$ such that with probability 1, $T^{(n)}$ has exactly $f(n)$ triangles for all sufficiently large n.

6.3 An Extension of the Barabási-Albert Model

Shortly after its appearance, Dorogovtsev, Mendes and Samukhin extended the Barabási and Albert Model [18]. In their version of the model, at each time-step t, one new vertex and simultaneously m new *directed* edges are introduced to the graph. For the directed edges, it is of no importance where they come from; it might be from the new vertex, from existing vertices or even from outside the system, we now only consider the incoming degree of a vertice. Let us denote the vertex added to the system at time s by s. We denote its in-degree by k_s^{in}. The probability that one of the m new links points to s is proportional to

$$A_s = A + k_s^{\text{in}}$$

which is called the *attractiveness* of a vertice. Upon their appearance, sites have an initial attractiveness $A \geq 0$; the total attractiveness increases as the site obtains new edges. Note that for the case $A = m$, the model is equivalent to the initial Barabási-Albert model, where the outgoing edges count towards the attractiveness of a site as well. However, in contrast to the former model note that it is possible that a vertex receives more than one new edge at the same time, which is simply not defined in the Barabási-Albert model. Even though the line of argumentation may not be perfectly rigorous, the criticisms from section 6.2.3 do not apply to this model.

Dorogovtsev *et al.* proceed to derive the *master equation*: Let $p(d, s, t)$ be the distribution of the in-degree d of vertex s at time t, i.e. k_s^{in}. Starting the network with one vertex and m incoming links, at time t, there are t sites in the network, and $m(t-1)$ incoming links. Let A_Σ denote the total attractiveness of the network, i.e. $A_\Sigma = (m + A)t = (1 + a)mt$, where $a = A/m$. The probability of one of the new edges connecting to site s is A_s/A_Σ. The probability that site s receives l new links of the incoming m is $\mathcal{P}_s^{(ml)} := \binom{n}{l} (A_s/A_\Sigma)^l (1 - A_s/A_\Sigma)^{m-l}$. Thus, for the master equation with initial condi-

tion $p(k,s,s) = \delta_{sk}$, where δ_{sk} is again the Kronecker delta, it follows

$$\begin{aligned} p(d,s,t+1) &= \sum_{l=0}^{m} \mathcal{P}_s^{(ml)} p(d-1,s,t) = \\ &= \sum_{l=0}^{m} \binom{m}{l} \left(\frac{d-1+am}{(1+a)mt}\right)^l \left(1 - \frac{d-1+am}{(1+a)mt}\right)^{m-l} p(d-1,s,t). \end{aligned} \tag{6.3}$$

The degree distribution of the entire network is given by $P(d,t) = \sum_{u=1}^{t} p(d,u,t)/t$, for which an equation can be found by summing up over equation (6.3). Assuming that $P(d) := \lim_{t\to\infty} P(d,t)$ exists, a stationary connectivity distribution can be found; after some calculations it follows

$$\begin{aligned} P(d) &= (1+a)\frac{\Gamma\left((m+1)a+1\right)}{\Gamma(ma)} \frac{\Gamma(d+ma)}{(d+2+(m+1)a)} \cong \\ &\cong (1+a)\frac{\Gamma\left((m+1)a+1\right)}{\Gamma(ma)} (d+ma)^{-(2+a)}, \end{aligned} \tag{6.4}$$

where the latter relation holds for $ma + q \gg 1$. Thus, the scaling exponent γ of the distribution is

$$\gamma = 2 + a = 2 + \frac{A}{m}.$$

Note that the scaling exponent of the distribution can now be tuned via A.

6.4 Some Rigorous Results on Exact Models

6.4.1 The Diameter and Clustering Coefficient of the LCD Model

The model

After their criticism of the Barabási-Albert Model (see section 6.1), Bollobás and Riordan published a mathematically correct version of a model that builds on the same principle of linear preferential attachment [9], and proved properties of both diameter as well as degree distribution.

As in section 6.1, a random graph process $(G_m^n)_{n\geq 0}$ is described where at each time step n a new vertex $v_n := n$ is introduced to the model. n then connects to m already existing vertices (itself included - loops are allowed) one at a time. This means that double edges are possible. For simplicity, we will start with the description of the process for $m = 1$.

Let us write $\mathrm{d}_G(v)$ for the degree of vertex v in a graph G. The description of the graph process $(G_1^n)_{n\geq 0}$ is as follows: Starting from the empty graph G_1^0, at each time step n one vertex v_n is added to the graph G_1^{n-1} that connects to one other vertex v_i, $1 \leq i \leq n$ where i is chosen at random with

$$\mathbb{P}(i=s) = \begin{cases} \mathrm{d}_{G_1^{n-1}}(v_s)/(2t-1) & 1 \leq s \leq n-1, \\ 1/(2t-1) & s = n. \end{cases} \tag{6.5}$$

Note that, in this definition, the edge going away from vertex n already contributes to the sum of edge degrees, i.e. the "beginning" of the new edge is counted already.

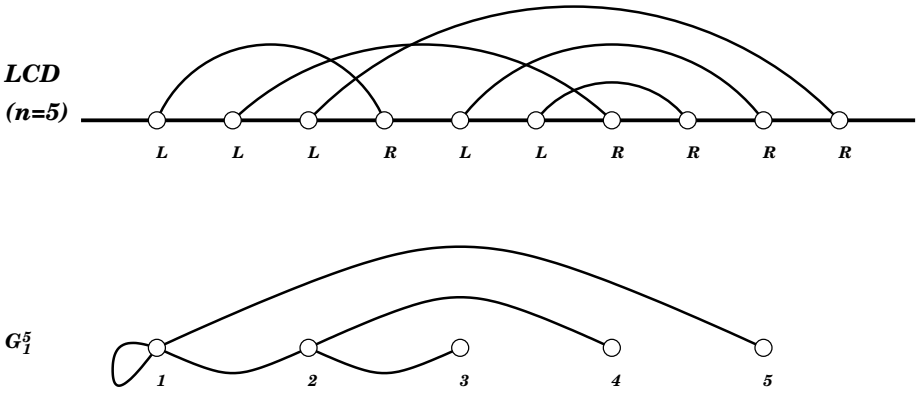


Figure 6.1: A visual explanation of the LCD Model: An n-pairing for $n = 5$, then G_1^5 is formed.

For general m, the process $(G_m^n)_{t\geq 0}$ is defined by running the process (G_1^n) on a sequence $v_1', v_2', \ldots$ of vertices. $(G_m^n)_{t\geq 0}$ is then formed from $(G_1^{mn})_{t\geq 0}$ by associating the vertices $v_1', v_2', \ldots, v_m'$ with vertex v_1, then the vertices $v_{m+1}', v_{m+2}', \ldots, v_{2m}'$ with vertex v_2, and so on, and then adding the edges accordingly. Thus, from (G_1^n), we can identify a process (G_m^n), and thus deduce important properties; for the moment we will consider (G_1^n). Let us denote the probability space of all (G_1^n) as $\mathcal{G}_1^n$. Equivalently, let $\mathcal{G}_m^n$ denote the probability space of all (G_m^n).

An *n-pairing* is defined as a partition of the set $\{1, 2, \ldots, 2n\}$ into pairs. An n-pairing can be visualized as connecting pairs of $2n$ distinct points on the x-axis with each other via semi-circular chords in the upper half plane, which gives it the name of a *linearized chord diagram* (LCD). This is the origin of the name of this model. Identifying a pairing with an LCD, we can speak of chords and their left and right endpoints.

From an LCD L, a graph $\Phi(L)$ can be formed in the following way: Starting from left to right, all left end points of chords up to and including the first right endpoint are identified with vertice 1. The next left endpoints up to and including the next right endpoint are identified with vertice 2, and so on. Chords joining left and right endpoints are now replaced with edges connecting the corresponding vertices. Bollobás *et al.* claim that, if L is chosen uniformly at random from all $(2n)!/(n!2^n)$ existing n-pairings, then $\Phi(L)$ has the same distribution as a random $(G_1^n) \in \mathcal{G}_1^n$. An explanation for this is that, by taking an LCD L' with $n-1$ chords, an LCD L can be formed by placing a new right endpoint to the right of the last point in L' and placing the left end point uniformly at random in one of the $2n-1$ possible places between already existing points. The formation of $\Phi(L)$ then corresponds to adding a vertex to $\Phi(L')$ and adding an edge with probability as defined by (6.5). See figure 6.1 for a visual explanation.

The Diameter

Theorem 6.1. *For a fixed $m \geq 2$ and a positive real number ε, a.e. $G_m^n \in \mathcal{G}_m^n$ is connected and has diameter* $\operatorname{diam}(G_m^n)$ *satisfying*

$$(1-\varepsilon)\log n/\log\log n \leq \operatorname{diam}(G_m^n) \leq (1+\varepsilon)\log n/\log\log n. \tag{6.6}$$

Outline of proof: The proof of the lower bound is quite quick. A graph G_1^N is considered,

with $N = mn$. Remember that the set of vertices is $\{1, 2, \dots N\}$. After some technical lemmas, it is proved that the distance between n and $n-1$ in G_m^n is greater than $L = \log n / \log(3Cm^2 \log n)$, where C is a constant: We consider the sum of all possible paths of length $l \leq L$ from n to $n-1$, and find that this number tends to zero in expectation.

The upper bound is much more complicated, the main probabilistic tool used is Theorem 4.3. Remember, $N = nm$. First, a $2N$-pairing is generated in the following way: Take $2N$ independent samples from the uniform distribution on $[0, 1)$, $x_1, x_2, \dots, x_{2N}$. Under the assumption that these pairs are distinct, we form an LCD as follows: Let x_{2i-1} and x_{2i} form pairs, for all i, then re-label the vertices from left to right in ascending order. This will give a pairing with the correct distribution. However, it is also possible to generate a pairing starting with the right end-points $r_1, r_2, \dots, r_N$; these are not uniformly distributed, but have density function $2x$. They are so-called $M_2(0, 1)$ random variables. Given $r_1, r_2, \dots, r_N$, the left endpoints $l_1, l_2, \dots, l_N$ are independent with l_i uniformly distributed on $[0, r_i]$.

Taking $N = mn$ independent $M_2(0, 1)$ random variables $r_1, r_2, \dots, r_N$, and sorting them into ascending order, $R_1, R_2, \dots, R_N$, we are only interested in every m^{th} endpoint, which we shall denounce by $W_i := R_{mi}$. Let $w_i := W_i - W_{i-1}$, where $W_0 = 0$. Bollobás *et al.* proceed to prove that, for i large enough, whp W_i is close to $\sqrt{i/n}$, while for certain intervals whp for at least half the vertices i it holds that $w_i \geq \frac{1}{10\sqrt{in}}$.

The right endpoints $R_1, R_2, \dots, R_{mn}$ are now taken as given, then independent random variables L_i, $1 \leq i \leq mn$ are inserted into the model where each L_i is uniformly distributed on $[0, R_i]$. We thus obtain an LCD on the chords $\{L_i, R_i\}$. G_m^n is obtained by taking m edges from each vertex i (corresponding to W_i, $1 \leq i \leq n$) to $l_{i,j}$, where $l_{i,j} = k$ if $W_{k-1} < L_{m(i-1)+j} < W_k$.

It is easier to consider only the W_is, "ignoring" the R_is. This is possible because, if we consider $L_{m(i-1)+j}$ to be random variables on $[0, W_i]$ instead of on $\left[0, R_{m(i-1)+j}\right]$, this will only increase the diameter by increasing the probability of loops at i. We will also only consider the case $m = 2$, again because removing edges will only increase the diameter; thus, vertex i only attaches to two edges, $l_{i,1}$ and $l_{i,2}$. The $l_{i,j}$s are independent with $\mathbb{P}(l_{i,j} = k) = w_k / W_i$. With this simplified process, we then form our graph $G = G(W_1, \dots W_n)$ with which it shall be proved that

$$\mathbb{P}\left(\operatorname{diam}(G) \leq (1+\varepsilon) \log n / \log\log n\right) = 1 - o(1).$$

The next step is to define a vertex i as *useful* if $i \leq n/(\log n)^5$ and $w_i \geq (\log n)/n$, i.e. if it is early enough in the sequence of vertices, and if its probability of being chosen as a neighbor for a new vertex is big enough.

First, a lemma is stated proving that with probability $1 - o(1)$ every vertex v of G is joined by a path of length at most $8 \log\log n$ to a useful vertex. The idea behind this is first to show that the number of vertices reached after k steps is not much smaller than two times the number of vertices reached after $k-1$ steps— much as in the proof of Theorem 4.10 — as long as no useful vertices are used. Then, it is shown that for a non-useful vertex i, the probability that $l_{i,1}$ is useful is at least $(\log n)^{-3}$, meaning that the probability of a non-useful vertex attaching to a useful one is fair. Combining these statements, the lemma then follows.

The next step is to show that for a useful vertex v, $1 \leq v \leq n$, with probability $1 - o(n^{-1})$ it holds that there is a path in G between v and 1 of length at most $(1/2 + \varepsilon) \log n / \log\log n$.

To prove this, we first define a vertex to be *good* if $w_i \geq \frac{1}{10\sqrt{in}}$. Then, a weighting function is introduced, $f_k = \sum_{i\in\Gamma_k} \frac{1}{\sqrt{in}}$. Let $\Gamma_0 := \{v\}$ and let Γ_k be the set of those j that are good and in a certain interval of $[n]/\{\Gamma_0 \cup \Gamma_1 \cup \ldots \Gamma_{k-1}\}$ and reachable from Γ_{k-1} by edges that are chosen in a way that maintains independence of edges. Let $N_k = \Gamma_0 \cup \Gamma_1 \cup \ldots \Gamma_k$.

Considering an interval $I_t = \left[2^t + 1, 2^{t+1}\right]$ for a fixed t in a certain range, it is shown that for

$$\mu_1 := \sum_{i\in\Gamma_k,\ i>2^{t+1}} \frac{\sqrt{2^t}}{80\sqrt{i}},$$

$\mu_1/2$ vertices of I_t are likely to be hit by edges coming out of $\Gamma_k \cap \left[2^{t+1}+1, n\right]$. A similar relationship is proved for edges coming from I_t into $\Gamma_k \cap \left[s+1, 2^t\right]$, where s is defined earlier in the proof. With help of the predefined function and some calculations, the proof is complete. □

Bollobás *et al.* mention that for the case $m = 1$, it has been proved that the diameter is $\Theta(\log n)$. (See [40].)

The Clustering Coefficient

As for the clustering coefficient of G_m^n, we shall see that it is actually relatively low, tending to zero as $n \to \infty$. This is one of the aspects of complex networks that the Barabási-Albert Model did not take into account. The following theorem is from [9].

Theorem 6.2. *For fixed $m \geq 1$, the expected value of the clustering coefficient $C(G_m^n)$ satisfies*

$$\mathbb{E}(C(G_m^n)) \sim \frac{m-1}{8}\frac{(\log n)^2}{n}$$

as $n \to \infty$.

The proof makes use of another theorem stated in [9]. A fixed graph S is defined on $[n]$ such that each edge ij in S is oriented from j to i if $j \geq i$. $V^+(S)$ is defined as the set of vertices in S from which edges leave, while $V^-(S)$ is defined as the set of vertices in S that have incoming edges. We define $\mathrm{d}_S^{\mathrm{in}}(i)$ and $\mathrm{d}_S^{\mathrm{out}}(i)$ as the in- and out-degrees of vertex i, respectively. Let $C_S(t)$ be the number of edges of S "crossing" t, i. e. , the number of edges ij with $i \leq t$ and $j \geq t$. As described in chapter 3, we say S is a subgraph of G_1^n writing $S \subseteq G_1^n$ if for the edge $ij \in E(S)$ it holds that $ij \in E(G_1^n)$, i.e. we are not dealing with isomorphic subgraphs. Let $\mathrm{d}_S^{\mathrm{out}} \leq 1$; it is possible that $S \subset G_1^n$. The following theorem is almost verbatim from [9].

Theorem 6.3. *Let S be a possible subgraph of G_1^n. With the notation above, the probability p_s that $S \subseteq G_1^n$ satisfies*

$$p_s = \prod_{i\in V^-(S)} \mathrm{d}_S^{in}(i)! \prod_{i\in V^+(S)} \frac{1}{2i-1} \prod_{t\notin V^+(S)} \left(1 + \frac{C_S(t)}{2t-1}\right).$$

Furthermore,

$$p_s = \prod_{i\in V^-(S)} \mathrm{d}_S^{in}(i)! \prod_{ij\in E(S)} \frac{1}{2\sqrt{ij}} \exp\left(O\left(\sum_{i\in V(S)} \frac{C_S(i)^2}{i}\right)\right).$$

Outline of proof of Theorem 6.2: As in section 4.4, first the expected number of triangles $T_{G_m^n}$ is counted, then the expected number of pairs of connected vertices $P_{G_m^n}$. Remember that G_m^n can be formed from G_1^{mn} by joining groups of m vertices to one new vertex. Thus, for given a, b, c with $1 \leq a < b < c \leq n$ it holds that abc is a triangle in G_m^n if and only if there are integers

$$\begin{array}{ccc} m(a-1) < i, & m(b-1) < j, & m(c-1) < k, \\ i' \leq ma, & j' \leq mb, & k' \leq mc, \end{array}$$

so that $\{ij', jk', i'k\} \in E(S)$ and $S \subseteq G_1^{mn}$. Assuming $\mathrm{d}_S^{\mathrm{out}}(v) \leq 1$ for all v, from Theorem 6.3 we get

$$p_S = \eta_1 \prod_{x \in V^-(S)} \mathrm{d}_S^{\mathrm{in}}(x)! \prod_{xy \in E(S)} \frac{1}{2\sqrt{xy}} = \eta_2 \prod_{x \in V^-(S)} \mathrm{d}_S^{\mathrm{in}}(x)! \frac{1}{8m^3abc},$$

with η_1 and η_2 bounded and tending to 1 for $a \to \infty$. Considering the number of possible choices for i, i', j, j', k, k', the expected number of triangles with vertices a, b, c in G_m^n follows quickly, namely

$$\eta_3 \frac{m(m-1)(m+1)}{8abc},$$

again with η_3 bounded and tending to 1 as $a \to \infty$. Summing up over a, b, and c with $1 \leq a < b < c \leq n$, the expected number of triangles $T_{G_m^n}$ in G_m^n follows:

$$T_{G_m^n} = (1 + o(1)) \sum_{1 \leq a < b < c \leq n} \frac{m(m-1)(m+1)}{8abc} \sim \frac{m(m-1)(m+1)}{48} (\log n)^3. \tag{6.7}$$

In the next step, the number of pairs of adjacent edges ab, ac must be considered. For the proof, it is necessary to compare the different cases $b, c \leq a$, $b \leq a < c$ (equivalent to $c \leq a < b$) and $a < b, c$. With similar methods as before, it then follows that in expectation, we have

$$P_{G_m^n} \sim \frac{m(m+1)}{2} n \log n. \tag{6.8}$$

Thus, the for clustering coefficient $C(G_m^n)$, with (6.7) and (6.8) it holds $\mathbb{E}(C(G_m^n)) \sim \frac{m-1}{8} \frac{(\log n)^2}{n}$ which completes the proof. □

6.4.2 The Buckley-Osthus Model and the Degree Distribution

In principle, the Buckley-Osthus model is to the model described in section 6.3 what the LCD model is to the Barabási-Albert model: The idea of the model is the same, however, the definition of the model is put into a mathematically correct state, which is based on the LCD model [13]. Remember, at each timestep t a new vertex t is introduced to the model, and m new edges attach to random vertices, with preference given to higher in-degree. The Buckley-Osthus model differs from the model by Dorogovtsev *et al.* because the edges come from the new vertex t. Let us start with the case $m = 1$:

In the beginning, as in subsection 6.4.1 we have G_1^0, the empty graph, or G_1^1, the graph with one vertex and one loop. G_1^n is formed from G_1^{n-1} by adding the vertex n and one

directed edge from n to a vertex $i \in V(G_1^{n-1}) \cup \{n\}$ where i is chosen with probability

$$\mathbb{P}(i = j) = \begin{cases} \frac{\mathrm{d}^{\mathrm{in}}_{n-1}(j)+a}{(a+1)n-1} & \text{if } 1 \le j \le n-1 \\ \frac{a}{(a+1)n-1} & \text{if } j = n. \end{cases} \tag{6.9}$$

In the same way as in the previous section, we can form G_m^n from G_1^{mn}. Note that for the case $a = 1$, the model described here is equivalent to the model described in section 6.3. Thus, the results we shall state are valid for the LCD model as well. By abuse of notation, we shall denote both processes by G_m^n.

Let $N_n^m(d)$ be the number of vertices with in-degree d in G_m^n. The first result states that for $a \in \mathbb{N}$, the degree sequence is scale-free.

Theorem 6.4. *Let $r \ge 1$ and $a \ge 1$ be fixed natural numbers, and let $(G_m^n)_{n\ge 0}$ be the graph process defined above. Let $\varepsilon > 0$ and define*

$$\mathcal{P}_{in}(d) := (a+1)\frac{(am+a)!}{(am-1)!}\frac{(d+am-1)!}{(d+am+a+1)!}$$

Then with probability tending to 1 as $n \to \infty$, we have

$$(1-\varepsilon)\mathcal{P}_{in}(d) \le \frac{N_n^r(d)}{n} \le (1+\varepsilon)\mathcal{P}_{in}(d)$$

for all d so that $0 \le d \le n^{1/100(a+1)}$. In particular, for any d in this range with $d \to \infty$, with probability tending to 1 we have

$$\frac{N_n^r(d)}{n} = (a+1)\frac{(am+a)!}{(am-1)!}d^{-2-a}(1+o(1)) = \Theta(d^{-2-a}).$$

Note that, because $\Gamma(x) = (x-1)!$ for $x \in \mathbb{N}$, this result confirms the heuristic result of Dorogovtsev *et al.* stated in section 6.3. (Compare to (6.4).)

The proof of Theorem 6.4 follows by summation after proving following theorem:

Theorem 6.5. *For any a, $r \in \mathbb{N}$ and any k with $n^{1-1/20(a+1)} \le k \le n - n^{1-1/20(a+1)}$ and $0 < d \le n^{1/20(a+1)}$, the in-degree $\mathrm{d}_n^{in}(k)$ of the k^{th} vertex satisfies*

$$\mathbb{P}(\mathrm{d}_n^{in}(k) = d) = \binom{d+ar-1}{ar-a} p^{ra}(1-p)^d(1+o(1)) + o(n^{-1}),$$

where $p := (k/n)^{1/(a+1)}$.

The idea behind this model is to expand the LCD model. G_n^m is formed via an $(a,1)$–*matching* M_n^a instead of just a pairing, which is defined as the partition of the set $[(a+1)n]$ into n sets with $a+1$ elements each. One element within each set is singled out as the *head* of the set, and all the other a vertices (*tails*) are attached to a chord pointing to the head. We can again visualize this matching on the positive x-axis. Taking M_1^a to be the first $a+1$ natural numbers on the x-axis where uniformly at random one of them is defined as the head, the M_n^a follows inductively by adding a new points at the right end of the existing model, and then uniformly at random choosing one of the $(a+1)n-1$ "gaps" between two of the already existing points to put the tail of this set. We can then label

every point with H or T, depending on if the point is a head or a tail. This will be referred to as an (H,T)-pattern. For a given (H,T)-pattern P and $i = 1, 2, \ldots, n$, let $h(0, i; P)$ be the number of heads which could be connected to the i^{th} block of a tails in P. A given P corresponds to exactly $\prod_{i=1}^{n}(h(0,i;P) - i + 1)$ $(a,1)$-matchings.

To construct a directed graph $\Phi(M_n^a)$ on n vertices from M_n^a, we proceed as in subsection 6.4.1. Advancing from left to right, all points up to and including the a^{th} T are joined together to form vertice v_1, from here until up to and including the $2a^{\text{th}}$ T all points are joined together to form vertice v_2, and so on. The a chords originating at the the a Ts forming each vertex form one edge in $\Phi(M_n^a)$. The in-degree $\mathrm{d}^{\text{in}}(v_i)$ of each vertex is equal to the number of heads encountered in forming it, while it always holds that its out-degree $\mathrm{d}^{\text{out}}(v_i) = 1$. With similar arguments as with the LCD model, it follows that $\Phi(M_n^a)$ has the same distribution as a random $G_1^n \in \mathcal{G}_1^n$. See figure 6.2.

Outline of the proof of Theorem 6.4. Martingales play an important role in this proof. The proof starts by calculating the expected value of $D_{k,n}$, the sum of the degrees of the first k vertices in G_n^1. The probability that $D_{k,n} = D_{k,n-1} + 1$ is

$$p := \frac{D_{k,n-1} + k(a-1)}{(a+1)n - 1}.$$

It follows quickly that

$$\mathbb{E}\left(D_{k,n} \mid D_{k,n-1}\right) = \frac{k(a-1)}{(a+1)n-1} + D_{k,n-1}\frac{(a+1)n}{(a+1)n-1},$$

so

$$\mathbb{E}\left(D_{k,n}\right) = \frac{k(a-1)}{(a+1)n-1} + \mathbb{E}\left(D_{k,n-1}\right)\frac{(a+1)n}{(a+1)n-1}$$

from which, with the boundary condition $\mathbb{E}\left(D_{k,k}\right) = 2k$, an expression for $\mathbb{E}\left(D_{k,n}\right)$ can be obtained. With these thoughts in mind, an estimate of $\mathbb{E}\left(D_{k,n}\right)$ for a certain range of k is given. With help of Lemma 3.4, it follows that D_k is quite close to its mean, namely the event $\mathcal{A}_{k'} := \{|D_l - \mathbb{E}(D_l)| \leq 3\sqrt{n}\log n : l = k, k'\}$ holds with probability tending to 1.

Let us define $h(k,k') := \sum_{j=k+1}^{k'} \mathrm{d}_n^{\text{in}}(j)$, the sum of in-degrees (or heads) over a certain interval $(k,k']$. The expected value of $h(k,k')$ follows easily, as does the proof that $h(k,k')$ should be close to its mean.

We proceed to define intervals in a pattern or matching: For $0 \leq k < \tilde{k} \leq n$, let $(k,\tilde{k}]$ denote the set of points after the $(ak)^{\text{th}}$ tail up to and including the $(a\tilde{k})^{\text{th}}$ tail. $t(k,\tilde{k}) = a(\tilde{k}-k)$ denotes the number of tails in the interval, whereas as above, $h(k,\tilde{k})$

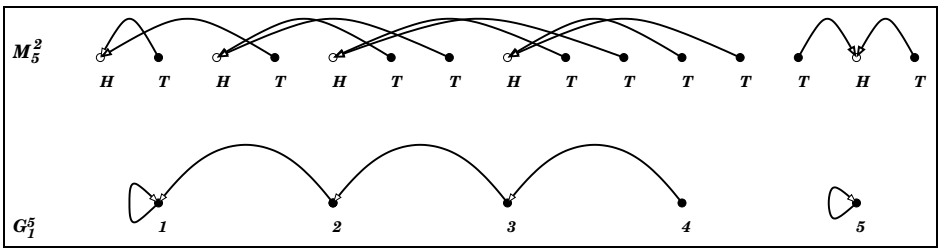


Figure 6.2: A visual explanation of the Buckley-Osthus model for $n = 5$ and $a = 2$.

denotes the number of heads. $h(k,\tilde{k};P)$ is written when a particular (H,T)-pattern P is being considered. Let $D_k(P) = h(0,k;P) + k$ denote the partial degree sum (in- and out-degrees) given a particular (H,T)-pattern P.

Let $R \leq k \leq n - R$ with $R = n^{1-1/20(a+1)}$, and define

$$k^* = k + n^{1/2-1/5(a+1)} = (1+\varepsilon^*)k \qquad \text{and} \qquad k' = k + n^{1/2+1/5(a+1)} = (1+\varepsilon')k.$$

Given an (H,T)-pattern $P \in \mathcal{A}_{k'}$, let $M := h(k,k';P)$, and denote the set of all (H,T)-patterns that agree with P outside $(k,k']$ by Γ_M. Let $(\Gamma_M, \mathbb{P}_{\Gamma_M})$ be the probability space where the elements of Γ_M are chosen uniformly at random.

A second probability space is defined on P and M as follows:
Let $N := t(k,k') + h(k,k';P)$, i.e. the number of all points in the interval $(k,k']$ and let $\Gamma_p = \Gamma_{M/N}$ be the space of all (H,T)-patterns generated by N independent random trials, the probability for a head being $p = M/N$ and the probability for a tail being $1-p$. Let $\mathbb{P}_{\Gamma_p}$ denote the probability measure over this set.

Let $\mathcal{A}_{k^*,p} := \{h_p(k,k^*) \in [(1\pm\delta)(p((1-p))t(k,k^*)]\}$, where $\delta = n^{-1/90(a+1)}$ and the number of heads which occur in Γ_p before $t(k,k^*) = a(k^*-k)$ tails have appeared is denoted by $h_p(k,k^*)$, and $[(1\pm\delta)x]$ is the set of natural numbers $i \in \mathbb{N}$ such that $(1-\delta)x \leq i \leq (1+\delta)x$.

After bounding $\frac{p}{1-p}$, using the Chernoff inequality (Theorem 4.3 it is shown that $\mathcal{A}_{k^*,p}$ holds with probability tending to 1 in $(\Gamma_p, \mathbb{P}_{\Gamma_p})$, from which it follows that $h_p(k,k^*)$ is close to $\mathbb{E}(h(k,k^*))$.

Similarly, for $\mathcal{A}_{k^*,M} := \{h_M(k,k^*) \in [(1\pm\delta)(p((1-p))t(k,k^*)]\}$ it is shown that $\mathbb{P}_{\Gamma_M}(\mathcal{A}_{k^*,M}) \to 1$.

From here, the next step is to extend the probability space Γ_M from (H,T)-patterns on $(k,k']$ to include the corresponding $(a,1)$-matchings. Let $\mathcal{M}'(P)$ denote the set of $(a,1)$-matchings in $\mathcal{M}_n^a$ whose (H,T)-patterns agree with P outside $(k,k']$, and define $\mathbb{P}_{\mathcal{M}'(P)}$ to be the probability on $\mathcal{M}'(P)$ when the elements are chosen uniformly at random. For $\mathcal{A}_{k^*} := \{h(k,k^*) \in [(1\pm\delta)(p((1-p))t(k,k^*)]\}$, it can be shown that $\mathbb{P}_{\mathcal{M}'(P)}\left(\mathcal{A}_{k^*}\right)^C = o(n^{-1})$ a.a.s.

Let $\mathcal{M}^*(P)$ denote the set of all $(a,1)$-matchings in $\mathcal{M}_n^a$ whose (H,T) patterns agree with P outside $(k,k^*]$. As a last step before beginning to actually count vertex degrees, it is shown that one is able to consider patterns instead of the corresponding matchings with only a negligible error.

Finally, from here on, it is possible to calculate $\mathbb{P}_{\mathcal{M}^*(P)}(d_{k+1} = d+1)$ as the number of (H,T)-patterns in $(k,k^*]$ with d Hs in in $(k,k+1]$ and $h(k,k^*) - d$ Hs in $(k+1,k^*]$ divided by the number of (H,T)-patterns in $(k,k^*]$ times an error factor $1+o(1)$. After some combinatorial thoughts, it is necessary to go back to the probability space $(\mathcal{M}_n^a, \mathbb{P})$ of all $(a,1)$-matchings. After a few estimates and calculations, the statement of Theorem 6.4 follows in expectation. With Lemma 3.4 it is shown that the the number of vertices with in-degree k will indeed be close to their mean. □

This shows that the Buckley-Osthus model is indeed scale-free. As the introduction of an initial attractiveness a just adds to the "randomness" of the model, we would expect the diameter of the Buckley-Osthus Model to be somewhere between the diameter of an Erdős-Rényi random graph and the diameter of the LCD model, so fairly low. Unfortunately, no such intuition is possible for the clustering coefficient.

7 More Models

In this finalizing chapter, some other networks that did not quite fit into any of the previous chapters will be presented. The first two are important in the development of the theory of complex networks, they will presumeably also become classics when talking of this subject in few years time. The latter two are very recent papers that I found interesting and thus worth mentioning.

7.1 The Copying Model

Quite an early model on the web graph that imposes a technique similar to preferential attachment is the so-called *copying model* by Kumar *et al.* [32].

Here, the intuition is the following: *Some* new nodes that come to the web are very innovative in their topic and find out where to link to all by themselves — i.e. the first person to make a page for foreign students coming to study in Vienna will think of all the sites of interest "from scratch".

However, *most* people making a web-site are interested in some topic where there already exists a wide variety of ressources. They will then "copy" some of these links, and, as they might have a different approach to the same topic, will add a few which they find important under a different aspect. Coming back to our example, a person making a web page for international students coming to Vienna University of Technology will copy many (but probably not all) links from the initial site for foreign students, and might add a few that are specific for studying at this particular university.

Kumar *et al.* give two models, the *linear growth model* and the *exponential growth model*. In the latter, at each time step t a fraction of the number of vertices existing at $t-1$ is added to the model; the argumentation behind being that new sites won't be "discovered" for quite a while. We will only explain the linear model and give results for it here, as the comparison with other models given is easier.

The description of the model is then as follows: The shape of this model at $t = 0$ is not defined, so at $t = 0$ the model consists of an arbitrary graph G_0.

Then, at time $t > 0$, a new vertex v is added to the graph G_{t-1}. This vertex connects via *directed, labeled* edges to a fixed number of edges d in the following way: First, with uniform probability, a node u is chosen from G_{t-1} to be the "prototype" vertex. We now can define the *copy factor* $0 < \alpha < 1$: For every edge i, $i = 1, 2, \ldots, d$, with probability α, the edge connects v to a vertex chosen uniformly at random in G_{t-1}. With probability $1 - \alpha$, edge i "copies" the destination of the i^{th} edge of vertex u, meaning that it connects to the same vertex.

Note that the larger the topic, the larger the probability that a node from this topic will be chosen as the prototype vertex. Results are then given for $d = 1$. Let $N_{t,k}$ be the number of vertices with in-degree k. In [32], Kumar *et al.* show following

Theorem 7.1. *For $r > 0$, the limit $P_r = \lim_{t\to\infty} \frac{N_{t,r}}{t}$ exits, and satisfies*

$$P_r = P_0 \prod_{i=1}^{r} \frac{1 + \frac{\alpha}{i(1-\alpha)}}{1 + \frac{2}{i(1-\alpha)}}$$

and

$$P_r = \Theta\left(r^{-\frac{2-\alpha}{1-\alpha}}\right).$$

We thus see that the copying model also results in a scale-free degree distribution. What is also shown in [32] is that the number of cliques $K_{i,j}$ — bipartite subgraphs where all the possible edges are present — will be very high, in G_t there will be $\Omega(t\exp(-i))$ cliques $K_{i,i}$. This was one of the aims when constructing this model, and is also not surprising when thinking of the intuition behind it.

7.2 The Cooper-Frieze Model

Cooper and Frieze published an extremely general model of the web graph, that takes into account several concepts of other models we have seen already, i.e. preferential attachment, network growth, as well as the concept of uniform randomness of the Erdős-Rényi Model [15]. They tried to maximize the choice given at each step.

The Cooper-Frieze model is based on the web graph, however, it is undirected. Interestingly, a justification is given for the undirectedness of the model: not only is it easier to analyze, but also, through the way the large search engine Google works, it is possible to find the internet addresses of all sites that have links to a certain webpage. (To do this, simply enter `link:node-url` in Google.)

Cooper and Frieze refer to their model as a web graph, as opposed to a scale-free graph, because they take a so-called configuration model. This means that they do not select vertices directly with a probability proportional to degree, but rather represent each vertex by points and from these select one uniformly, i.e. each vertex j is represented by the $\mathrm{d}(j)$ half-edges attached to it, and from all half edges present in the system, one is picked uniformly at random. Here, this means that preferential attachment is to pick a vertex w with following probability:

$$\mathbb{P}\left(w \text{ is the terminal vertex of a half-edge that has been chosen u.a.r.}\right) = \frac{\mathrm{d}(w)}{2|E|}.$$

This model takes into account many aspects that make other models somewhat unrealistic. We still consider discrete time steps t. There are two possibilities at time t: Either a new vertex is added to the model (and connects to nodes already in the model), or an existing vertex sends out new edges within the model. The number of new edges varies from time to time. There is a certain chance that the nodes a new edge connects to are chosen uniformly at random, and a counter-probability for choosing these nodes with preferential attachment.

To be precise, at time $t = 0$ the graph consists of one vertex v_0 and no edges. At time $t > 0$

- with probability $1 - \alpha$ we add a *new* vertex to the system (procedure *new*), and

- with probability α we add generate edges from an existing, *old* vertex (procedure *old*).

In procedure *new*, we have

- a vector $\mathbf{p} = (p_i, i \geq 1)$ where p_i is the probability that a new node generates i new edges,
- the probability β that the choices of terminal nodes are made uniformly, and
- the probability $1-\beta$ that the choices of terminal nodes are made according to degree.

Conversely, in procedure *old*, we have

- a vector $\mathbf{q} = (q_i, i \geq 1)$ where w_i is the probability that the old node generates i new edges,
- the probability δ that the old node is selected uniformly,
- the probability $1-\delta$ that the old node is selected according to degree,
- the probability γ that the choices of terminal nodes are made uniformly, and
- the probability $1-\gamma$ that the choices of terminal nodes are made according to degree.

A finiteness condition is imposed, meaning that there exist j_0, j_1 such that $p_j = 0$ for $j > j_0$ and $q_j = 0$ for $j > j_1$.

Not surprisingly, to state the results of Cooper and Frieze, it is necessary to introduce a large number of parameters.

Notation 7.1. *Let $D_k(t)$ be the number of vertices of degree k at time t. Let $\mu_p = \sum_{j=1}^{j_0} jp_j$, $\mu_q = \sum_{j=1}^{j_1} jq_j$, and $\theta = 2((1-\alpha)\mu_p + \alpha\mu_q)$. The parameters of the model are transformed as follows:*

$$\begin{aligned}
a &= 1 + \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha} + \frac{\alpha\delta}{1-\alpha},\\
b &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta} + \frac{\alpha(1-\delta)}{\theta},\\
c &= \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha},\\
d &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta},\\
e &= \frac{\alpha\delta}{1-\alpha},\\
f &= \frac{\alpha(1-\delta)}{\theta}.
\end{aligned}$$

Then, recursively, define $d_0 := 0$ and for $k \geq 1$

$$d_k(a+bk) := (1-\alpha)p_k + (c + d(k-1))d_{k-1} + \sum_{j=1}^{k-1}(e + f(k-j))q_j d_{k-j}.$$

This system of equations has a unique solution.

Theorem 7.2. *There exists a constant $M > 0$ such that almost surely for all $t, k \geq 1$*

$$|\bar{D}_k(t) - td_k| \leq Mt^{1/2} \log t.$$

The number of vertices at step t is whp asymptotic to $(1 - \alpha)t$. This means that the proportion of vertices of degree k is whp asymptotic to

$$\bar{d}_k = \frac{d_k}{1 - \alpha}.$$

Theorem 7.3. *There exist some constants $C_1, C_2, C_3, C_4 > 0$ such that*

i) $C_1 k^{-\zeta} \leq d_k \leq C_2 \min\{k^{-1}, k^{-\zeta/j_1}\}$, where $\zeta = (1 + d + f\mu_q)/(d + f)$.

ii) If $j_1 = 1$, then $d_k \sim C_3 k^{-(1+1/(d+f))}$.

iii) If $f = 0$, then $d^k \sim C_4 k^{-(1+1/d)}$.

iv) If the solution conditions given below hold, then

$$d_k = C\left(1 + O\left(\frac{1}{k}\right)\right) k^{-x},$$

where C is constant and

$$x = 1 + \frac{1}{d + f\mu_q}.$$

We say that $\{q_j; j = 1, 2, \ldots, j_1\}$ is periodic *if there exists $m > 1$ such that $q_j = 0$ unless $j \in \{m, 2m, 3m, \ldots\}$.*

Let

$$\phi_1(y) = y^{j_1} - \left(\frac{d + q_1 f}{b} y^{j_1 - 1} + \frac{q_2 f}{b} y^{j_1 - 2} + \cdots + \frac{q_{j_1 - a} f}{b} y + \frac{q_{j_1} f}{b}\right).$$

The solution conditions are:

i) $f > 0$ and either (a) $d + q_1 f > 0$ or (b) $\mathbf{q}$ is not periodic.

ii) The polynomial $\phi_1(y)$ has no repeated roots.

Theorems 7.2 and 7.3 are taken almost ad verbatim from [15]. In words, this means that the expected value of the number of vertices of degree k will be about t times the value d_k defined in the recursion, td_k, and under certain conditions (that are not necessary) the degree sequence will indeed be power law with exponent x.

7.3 Thickened Trees

Another model was ideated by Drmota, Gittenberger and Panholzer only recently [20]. They start with a random graph process called the **PORT-model**: At time step $t = 1$, the graph consists only of one vertex, 1, that will be the root of the tree. Then, at each timestep $t = 2, 3, \ldots$ a new vertex labeled t is added to the model. As in the Barabási-Albert Model, it attaches to one "older" vertex $i < t$, where the vertice it attaches to is chosen randomly. Higher preference is once again given to vertices with higher degree, the

probability of attaching to a vertex with degree $d = k + 1$ (where k is the in-degree) is proportional to d.

In fact, what is being constructed here is a tree with an obvious orientation: Because of the vertex labels, we know which nodes have appeared when (this is the *history* of the process); the younger nodes have edges that are directed to older nodes.

The model can also be interpreted as a planar version of recursive trees: For a vertex with in-degree k, note that there are $d = k + 1$ possibilities of attaching a new vertex to it. This is why the model is called *plane oriented recursive tree* model. It is proved in [31] that the PORT-model is scale-free with degree distribution

$$\lim_{n\to\infty} p_n(d) = \frac{4}{d(d+1)(d+2)},$$

where $p_n(d)$ is the probability of a random vertex of a PORT of size n having degree d.

The idea behind *thickened trees* is to modify the PORT-model so that the global tree-like structure is maintained, while at a local level the graph is highly clustered: For all $k \geq 0$, denote by $\mathcal{T}_k$ a set of labeled graphs, a so-called *substitution set*. For now, the only specification for the graphs in $\mathcal{T}_k$ is that for a graph $G \in \mathcal{T}_k$, $k + 1$ ordered and labeled *half-edges* come out of some vertices.

Let T be a tree generated by the PORT-model, it is modified as follows: Let v be a vertex of in-degree k. v is *cut* from the tree in such a way that v is removed, and so are the halves of the edges connecting to v. Thus, all vertices that had been adjacent to v now are incident to a half-edge. Then a random graph from $\mathcal{T}_k$ is inserted to where v was, and the half-edges are *glued* together respecting a certain order, meaning that the predecessor of v is attached to half-edge 0, then in ascending order the remaining successors are attached to the $1^{\text{st}}, 2^{\text{nd}}, \ldots, k^{\text{th}}$ half-edge. This procedure is followed for every vertex v in T. Finally, the nodes in the new graph $G = G(T)$ are relabeled in a way that is consistent with the labeling of T and the labellings in the $\mathcal{T}_k$s. See figure 7.1 for a visual explanation.

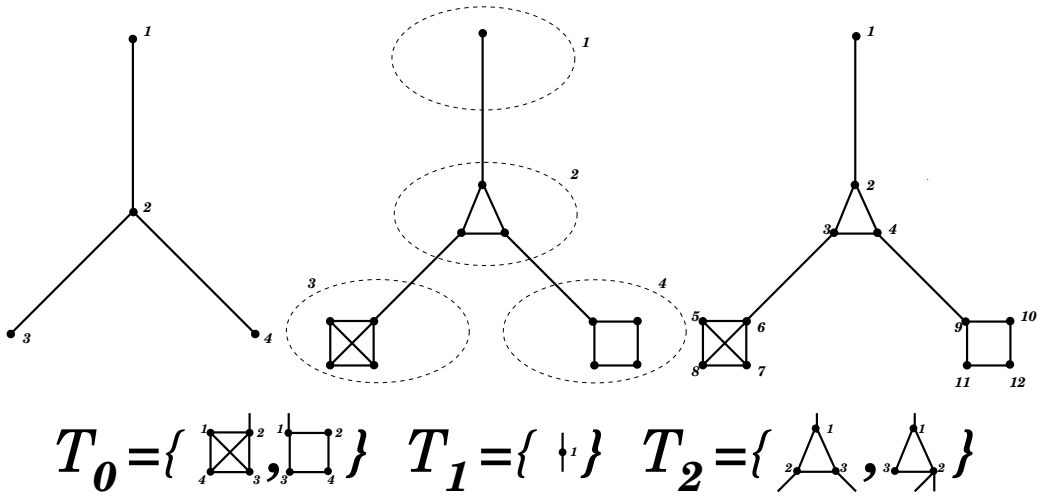


Figure 7.1: The principle of the PORT-Model. Intuitively, the vertices of this model are blown-up — thickened — and then replaced with random vertices from given sets. Then, the vertices are relabeled in an order-preserving way.

Remember that the clustering coefficient of the LCD model (see subsection 6.4.1) was

fairly low. In this model, it can be tuned to be fairly high, depending on how clustered (and how big) the graphs in the $\mathcal{T}_k$s are. It is also possible to treat this model analytically, and the influence of local changes on global structure can be shown. Note that the clusters are not formed by an evolution process, but are taken as given.

Drmota *et al.* use a generating function approach to prove that the degree distribution of thickened trees tends to a scale-free form as follows:

Let $\mathcal{T}_k$ be defined as described above. Let

$$t_k(z) = \sum_{G \in \mathcal{T}_k} \frac{z^{|G|}}{|G|!}$$

be the exponential generating function of these graphs. Let $N_d(G)$ denote the number of nodes in G of degree d (half-edges included), then let

$$t_k^{(d)}(z,u) = \sum_{G \in \mathcal{T}_k} \frac{z^{|G|}}{|G|!} u^{N_d(G)}.$$

Finally, let

$$T_d(z,y,u) = \frac{1}{z} \sum_{k \geq 0} t_k^{(d)}(z,u) y^k.$$

Theorem 7.4. *Let $\mathcal{T}_k$ be substitution sets (as described above) so that the equation*

$$\rho = \int_0^1 \frac{dt}{T_d(\rho, t, 1)}$$

has a unique positive solution in the region of convergence of $T_d(z,y,u)$ and that the $T_d(z,y,u)$ can be represented as

$$T_d(z,y,u) = \frac{C_0(z,y) + C_1(z,y)(1-y)^{r'} y^{d+\alpha}(u-1) + O\left((1-y)^{r'}(u-1)^2\right)}{(1-y)^r}, \qquad (7.1)$$

where r' and $r \in \mathbb{R}$ with $0 < r' \leq r$, $\alpha \in \mathbb{N}$, $C_0(z,y)$ and $C_1(z,y)$ are power series that contain $z = \rho$ and $y = 1$ in their regions of convergence and that satisfy $C_i(\rho, 1) \neq 0$ for $i = 0, 1$, and the $O(\cdot)$-term is uniform in a neighborhood of $z = \rho$ and $y = 1$.

Let $p_n(d)$ denote the probability that a random node in a thickend PORT of size n has degree d. Then the limits

$$\lim_{n \to \infty} p_n(d) =: p(d)$$

exist and we have, as $d \to \infty$,

$$p(d) \sim \frac{C}{d^{r+r'+1}}.$$

This theorem is taken almost ad verbatim from [20]. In [19], it is proved that (7.1) is satisfied in quite general conditions.

7.4 Protean Graphs

Łuczak and Prałat introduced a model for protean[1] graphs in [33]. Basically, the idea of this graph process is as follows: In a graph of n vertices, at each step t a random vertex j dies, its adjacent edges are removed from the graph. However, simultaneously, another vertex is born (which can be interpreted as the same vertex — the number of vertices in the system does not change). This "new" vertex attaches at random to d other vertices in the graph, where higher priority is given to "older" vertices. The new vertex is now the youngest vertex in the system. This aging process can be seen as a permutation of the n vertices, where at step t for the dying vertex j it then holds $\pi_t(j) = n$. For the other vertices i so that $\pi_{t-1}(i) > \pi(j)$ it now holds $\pi_t(i) = \pi_{t-1}(i) - 1$.

Though this model is theoretically very interesting and leads to a power law degree distribution, it could be criticized as a bad model for the web graph because the number n of vertices is fixed, while the web is currently growing very quickly. Thus not surprisingly, this model has a sequel: *Growing protean graphs.* In [41], Prałat and Wormald expand the model from [33] to a growing model.

This graph process $(G_t)_{t\geq 0}$ is dependent on the parameters $1/2 < p \leq 1$, $d \in \mathbb{N}$, and $0 < \eta < 1$. The process starts with G_0 consisting of one vertex, v_1, and no edges. At time t, G_t is formed from G_{t-1} in the following way:

- With probability p a new vertex $v_{|V_{t-1}|+1}$ is added to G_{t-1}. From this vertex, d edges are then added one after the other with v_i being chosen with probability $i^{-\eta}/\sum_{j=1}^{N_{t-1}+2} j^{-\eta}$. Loops and multiple edges are allowed.

- Contrarely, with probability $1-p$, if there is only one vertex left in the graph do nothing; if $|V_{t-1}| > 1$, choose a random vertex v_i, $i \in \{1, 2, \ldots, |V_{t-1}|\}$, delete it and all the edges incident to it. Then relabel the remaining vertices: v_{j+1} is now called v_j for $i \leq j \leq |V_{t-1}| - 1$.

The protean graph G_t is now called $\mathcal{P}_t(p, d, \eta)$.

Note that this graph is not necessarily connected: Intuitively, when several "old," well-connected vertices are removed one after the other, it is possible that graph will be disconnected. After a few thoughts, similar arguments as in [33] can be used to show that the degree distribution follows a power law. A statement on connectivity is also given.

Let $n = n(t) = (2p-1)t$. It is shown in [41] that $|V_t| - 1$ will be concentrated around n, i.e. there will be about n vertices in G_t. Let $Z_k = Z_k(n, p, d, \eta)$ be the number of vertices of degree k in $\mathcal{P}_t(p, d, \eta)$, and $Z_{\geq k} = \sum_{l\geq k} Z_l$. Following theorem is verbatim from [41]:

Theorem 7.5. *Let $1/2 < p \leq 1$, $d = o(t^{(1-\eta)/3})$ and $0 < \eta < 1$, $k = k(n) \geq \log^2 n$, and $d = o(k)$. Then a.a.s.*

$$Z_{\geq k} = n\left(\frac{1-\eta}{(1-p)/p + \eta}\frac{d}{k}\right)^{1/\eta}(1 + o(1)) + O(\log^3 n).$$

[1] *protean*: 1. of or resembling Proteus in having a varied nature or ability to assume different forms, 2. displaying a great diversity or variety: versatile (source: `www.merriam-webster.com`)

7.5 Summary and Outlook

In this diploma thesis, I hope to have given an introduction to what the modelling of complex networks is about. After an introduction where the basic notions of a complex network were defined as well as some basic ideas — the clustering coefficient, power laws, the diameter, etc. — I gave some motivating examples of complex networks in chapter 2. Then, before we could "get to the point," in chapter 3, the basic tools used in this diploma thesis were presented.

Finally, in chapter 4, we started with the models. This chapter was rather lengthy, and held several rigorous proofs about the "classical" random graph model of Erdős and Rényi, such as on threshold functions for balanced subgraphs, the giant component, the clustering coefficient, and the degree sequence of Erdős-Rényi random graphs.

Then, in chapter 5, one of the first models to analyze real world networks was presented, the small-world model. It put its focus on high clustering coefficient and small diameter. A simplified version of this model, including some analytical results, was included, as was a mean-field solution. This chapter also included first rigorous results by mathematicians, the technical Markov chain small world, and an intriguing model of spacial random graphs.

The following chapter 6 dealt with preferential attachment: The Barabási-Albert Model was introduced, and first results via a mean field solution given, as well as reasons why the preference function must be linear and a short criticism of this model. Then, the model was amplified by the model of Dorogovtsev and Mendes, followed by some rigorous results about the diameter on the mathematical extension of the Barabási-Albert model, the LCD model, and further results on the degree sequence of the Buckley-Osthus Model.

We now end with this chapter 7, where we quickly introduced the almost classical copying and Cooper-Frieze models, as well as pointing out two recent models: thickened trees and protean graphs.

This diploma thesis gives only a very short introduction to the the rapidly expanding theory of complex networks; many more models are available. Looking at each chapter, examples of papers that have not been cited here are easy to find, be it small world models built to examine the spread of diseases, or preferential attachment models taking the aging of sites into account. The theory of complex networks is an active area of research where many new models are yet to be born — more models that will surely be as intriguing as the ones I have come across researching for this thesis.

Bibliography

[1] Réka Albert and Albert-László Barabási. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[2] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, 74(1):47–97, 2002.

[3] Reka Albert, Albert-Laszlo Barabasi, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272:173–187, 1999.

[4] Albert László Barabási. Network theory - the emergence of the creative enterprise. *Science*, 308:639–641, 2005.

[5] M. J. Barber, A. Krüeger, T. Krüger, and T. Roediger-Schluga. The network of EU-funded collaborative r&d projects. 2005.

[6] A. D. Barbour and Louis H. Y. Chen. *An introduction to Steins method*. World Scientific, New Jersey, NJ, 2005.

[7] Béla Bollobás. *Random graphs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985.

[8] Béla Bollobás. *Modern graph theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[9] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[10] Béla Bollobás and Oliver M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34. Wiley-VCH, Weinheim, 2003.

[11] Anthony Bonato. *A course on the web graph*, volume 89 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[12] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer Networks*, 33(1-6):309 – 320, 2000.

[13] Pierce G. Buckley and Deryk Osthus. Popularity based random graph models leading to a scale-free degree sequence. *Discrete Math.*, 282(1-3):53–68, 2004.

[14] Minerva Catral, Michael Neumann, and Jianhong Xu. Matrix analysis of a Markov chain small-world model. *Linear Algebra Appl.*, 409:126–146, 2005.

[15] Colin Cooper and Alan Frieze. A general model of web graphs. *Random Structures Algorithms*, 22(3):311–335, 2003.

[16] S. N. Dorogovtsev and J. F. F. Mendes. Exactly solvable small-world network. *Europhys. Lett.*, 50:1–7, 2000.

[17] S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of networks*. Oxford University Press, Oxford, 2003. From biological nets to the Internet and WWW.

[18] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633–6, Nov 2000.

[19] M. Drmota, B. Gittenberger, and R. Kutzelnigg. Combinatorial models for cooperation networks. *To appear*, 2009.

[20] M. Drmota, B. Gittenberger, and A. Panholzer. The degree distribution in thickened trees. Discrete Math. Theor. Comput. Sci. Proc., AG, pages 153–166. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2008.

[21] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[22] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[23] Linton C. Freeman. Some antecedents of social network analysis. *Connections*, 19(1):39–42, 1996.

[24] Ayalvadi Ganesh and Feng Xue. On the connectivity and diameter of small-world networks. *Adv. in Appl. Probab.*, 39(4):853–863, 2007.

[25] John Guare. *Six Degrees of Separation*. Vintage Books, Random House, Inc., New York, 1990.

[26] Theodore E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963.

[27] Svante Janson, Donald E. Knuth, Tomasz Łuczak, and Boris Pittel. The birth of the giant component. *Random Structures Algorithms*, 4(3):231–358, 1993. With an introduction by the editors.

[28] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[29] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp. *Denumerable Markov chains*. Springer-Verlag, New York, second edition, 1976. With a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40.

[30] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85:4629–4632, 2000.

[31] Markus Kuba and Alois Panholzer. On the degree distribution of the nodes in increasing trees. *J. Combin. Theory Ser. A*, 114(4):597–618, 2007.

[32] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. pages 57–65, 2000.

[33] Tomasz Łuczak and Paweł Prałat. Protean graphs. *Internet Math.*, 3(1):21–40, 2006.

[34] Stanley Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[35] C. Moore, M. E. J. Newman, and D. J. Watts. Mean-field solution of the small-world network model. *Physical Review Letters*, 84:3201, 2000.

[36] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64(2):025102, Jul 2001.

[37] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98(2):404–409 (electronic), 2001.

[38] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256 (electronic), 2003.

[39] M. E. J. Newman, S. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64, 2001.

[40] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.

[41] P Prałat and N. Wormald. Growing protean graphs. *Internet Math.*, 4(1):1–16, 2008.

[42] S. Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, 4(2):131–134, 1998.

[43] L. E. Reichl. *A Modern Course in Statistical Physics*. Wiley-VCH Verlag GmbH & Co., Weinheim, 2004.

[44] E. Seneta. *Nonnegative matrices and Markov chains*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.

[45] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[46] Duncan J. Watts. *Six degrees*. W. W. Norton & Co. Inc., New York, 2003. The science of a connected age.